Name - Avinash Gautam
College Roll No. - 20201407
Exam Roll No. - 20020570009
Stream - B.Sc. (Hons.) Computer Science
Subject - Statistical Method (G.E.)
Semester - I
Paper - Statistical Method's Practical Exam

Q.①
Ans :-

Theory & Formula :-

Continuous frequency distribution :- A table showing the distribution of frequencies in different classes is called frequency table. We use struges rule for changing discrete frequency distribution to continuous frequency distribution.

Struges Rule :-

① Find the range :-

$$\text{range} = \text{Max} - \text{Min}$$

② Calculate no. of classes :-

$$k = 1 + 3.22 \log_{10} N$$

N = total observations

③ Calculate Class width :-

$$\text{Class width} = \text{range}/k$$

Bar graph :- It is a simple graphical representation of categorical data. We use rectangular bars ~~which~~ thats height is propotional to the corresponding frequencies. They can plotted vertical or horizontal.

Pie-Chart :- A pie-chart is a circular statistical graph which is divided into slices to illustrate numeric portion. In pie-chart, we calculate the angle ~~of~~ for a particular variable and plot it on the graph.

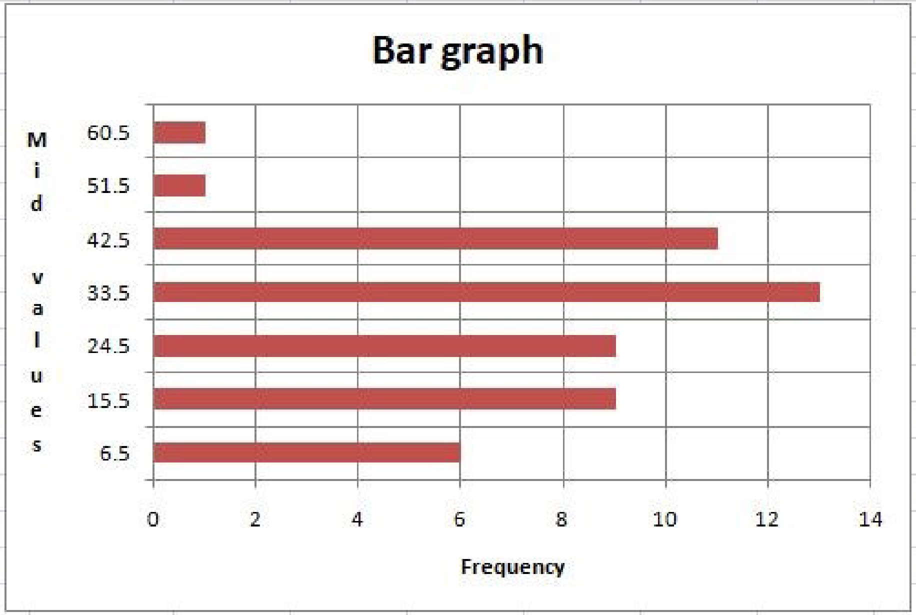
$$\text{angel} = \left(\frac{\text{corresponding freq.}}{\text{Sum of all freq.}}\right) \times 360$$

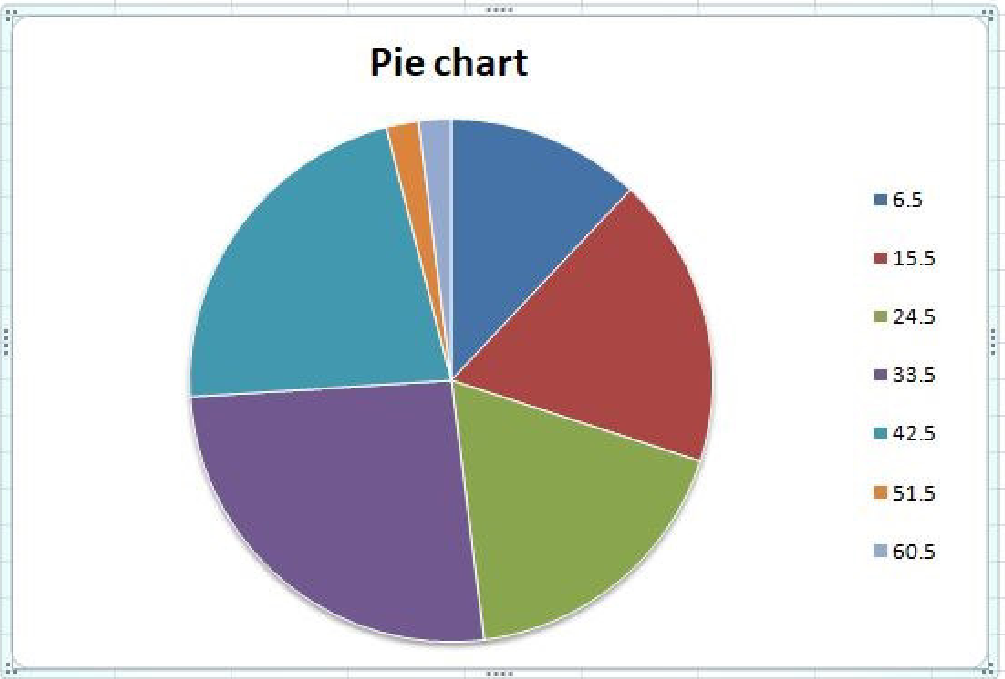
Calculation :-

42	31	19	14	23	28	17	53	22	21
20	30	63	9	30	29	12	21	32	25
18	42	38	44	7	39	6	15	41	45
30	2	44	15	37	5	36	31	46	46
19	24	22	30	30	13	31	30	44	6

N	**50**
max	**63**
min	**2**
range	**61**
k	**6.643978**
class width	**9.181246**

class interval		frequency	x
2	**11**	**6**	**6.5**
11	**20**	**9**	**15.5**
20	**29**	**9**	**24.5**
29	**38**	**13**	**33.5**
38	**47**	**11**	**42.5**
47	**56**	**1**	**51.5**
56	**65**	**1**	**60.5**





Result :-

$N = 50$

$Max = 63$

$Min = 2$

$Range = 61$

$K = 6.64 \approx 7$

$Class\ width = 9.18 \approx 9$

Class interval (29-38) have maximum frequency.

Q.② Ans

Theory & formula :-

Karl pearsen correlation coefficient :- It is a coefficient which is used to compare the relationship between two variables. It is a unit free measure. It measures the intensity or the degree of a linear relationship. It always lie between -1 to +1. We can denote it by r.

$$r = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y} \quad —①$$

where σ_x = Standard Deviation of x

σ_y = Standard Deviation of Y

σ_{xy} = covariance of x & Y

$$\sigma_{xy} = \frac{1}{N} \sum (x-\bar{x})(y-\bar{y})$$

Calculation :-

X	Y	x = X - Mean(X)	y = Y - Mean(Y)	x^2	y^2	x*y
15	10	-20	-12	400	144	240
20	15	-15	-7	225	49	105
25	20	-10	-2	100	4	20
30	20	-5	-2	25	4	10
35	22	0	0	0	0	0
40	25	5	3	25	9	15
45	26	10	4	100	16	40
50	28	15	6	225	36	90
55	32	20	10	400	100	200
				1500	**362**	**720**

Mean(X)	**35**
Mean(Y)	**22**

cov(XY)	**80**
SD(X)	**12.90994**
SD(Y)	**6.342099**

r	**0.977086**

Result :-

$$\sigma_{xy} = 80$$

$$\sigma_x = 12.909$$

$$\sigma_y = 6.342$$

$$r = 0.977086$$

Q. 3
Ans

Theory & Formula :-

Variability :- We can calculate variability by coefficient of variation. It is the ratio of the standard deviation to the mean times 100. We can denote it by CV.
Lower CV means more consistency and less variability.
Higher CV means less consistency and more variability.

$$C.V. = \frac{\sigma}{\bar{x}} \times 100 \quad — ①$$

Standard Deviation

$$\sigma = \sqrt{\frac{1}{N} \sum (x_i - \bar{x})^2} \quad — ②$$

Mean

$$\bar{x} = \frac{\sum x_i}{N} \quad — ③$$

where N = total no. of observations

Calculations :-

For data set I :-

$$\sum x = 49 + 82 + 77 + 54$$
$$\sum x = 262$$
$$N = 4$$

So Mean from eq. ③

$$\bar{x} = \frac{262}{4} = 65.5$$

Now

$$\sum (x - \bar{x})^2 = (49 - 65.5)^2 + (82 - 65.5)^2 + (77 - 65.5)^2 + (54 - 65.5)^2$$

$$\sum (x - \bar{x})^2 = 272.25 + 272.25 + 132.25 + 132.25$$

$$\sum (x - \bar{x})^2 = 809$$

So Standard Deviation from eq. ②

$$\sigma = \sqrt{\frac{809}{4}} = \sqrt{202.25}$$

$$\sigma = 14.22$$

So the ~~covariance~~ CV from eq. ①

$$C.V = \frac{14.22}{65.5} \times 100$$

$$\boxed{C.V = 21.71}$$

For Data set II :-

$$\Sigma x = 159 + 121 + 138 + 152$$

$$\Sigma x = 570$$

$$N = 4$$

Mean from eq. ③

$$\bar{x} = \frac{570}{4} = 142.5$$

Now

$$\Sigma (x - \bar{x})^2 = (159 - 142.5)^2 + (121 - 142.5)^2 + (138 - 142.5)^2 + (152 - 142.5)^2$$

$$\Sigma (x - \bar{x})^2 = 845$$

Standard Deviation from eq. ②

$$\sigma = \sqrt{\frac{845}{4}} = \sqrt{211.25}$$

$$\sigma = 14.534$$

So the C.V from eq. ①

$$C.V. = \frac{14.534}{142.5} \times 100$$

$$\boxed{C.V. = 10.12}$$

Result :-

C.V for data set I = 21.71

C.V for data set II = 10.12

So. C.V for I > C.V for II

Q (4)

Ans

Theory & Formula :-

Arithmetic Mean :- Arithmetic mean of a set of observations is their sum divided by the no. of observations.

$$\bar{x} = \frac{\Sigma f_i x_i}{N} \quad -(1)$$

where f = frequency

x = mid value

$N = \Sigma f$

Median :- Median is a value of a data set which divides the whole data set into two equal parts.

$$M_d = l + \frac{(\frac{N}{2} - cf)}{f} \times h \quad -(2)$$

where l = lower limit of median class

h = magnitude of median class

cf = cumulative frequency of preciding class

f = frequency of M_d class.

Calculation :-

Marks	f	x	cf	fx
0 – 15	3	7.5	3	22.5
15 – 30	30	22.5	33	675
30 – 45	40	37.5	73	1500
45 – 60	42	52.5	115	2205
60 – 75	35	67.5	150	2362.5
75 – 90	20	82.5	170	1650
Total →	170			8415

Mean

$$\bar{x} = \frac{8415}{170} = 49.5$$

$$\boxed{\bar{x} = 49.5}$$

Median

$$\frac{N}{2} = \frac{170}{2} = 85$$

So the M_d Class = (45-60)

$l = 45$
$h = 15$
$cf = 73$
$f = 42$

$$M_d = 45 + \frac{(85-73)}{42} \times 15$$

$$\boxed{M_d = 49.2857}$$

Result :-

Mean $(\bar{x}) = 49.5$

Median $= 49.2857$

We can say that data is very little skewed because Mean > Median. The curve of the data is positively skewed very little so we can consider it as a symmetrical data.