[Q.] Consider set of data which are the high temperature recorded for 50 consecutive days. Summarize this data by creating a frequency distribution of temperature.

50, 45, 49, 50, 43, 49, 50, 44, 45, 49, 44, 44, 44, 51, 51,
~~49, 47, 46, 50, 49, 51, 47, 43, 43, 49, 45, 46, 45, 51, 46~~
55, 36, 43, 53, 58, 53, 39, 58, 29, 33, 43, 47, 30, 53, 55, 50, 37, 36, 54, 43

DATE: __/__/__
PAGE: ____

Name :- Harshit Kumar

Statistics & Probability

Statistics

$$\left\{\begin{aligned}(a+b)^3 &= a^3+b^3+3a^2b+3ab^2\\ (a-b)^3 &= a^3-b^3-3a^2b+3ab^2\end{aligned}\right\} \left\{\begin{aligned}(a+b)^2 &= a^2+2ab+b^2\\ (a-b)^2 &= a^2-2ab+b^2\end{aligned}\right\}$$

Inclusive class interval

class Interval	frequency
25 - 29	1
30 - 34	2
35 - 39	5
40 - 44	10
45 - 49	16
50 - 54	13
55 - 59	3
Total	50

Exclusive class interval

class Interval	frequency
25 - 30	1
30 - 35	2
35 - 40	5
40 - 45	10
45 - 50	16
50 - 55	13
55 - 60	3
Total	50

let x = Temperature

#

x = Temperature	frequency
29	1
30	1
33	1
36	2
37	1
38	1
39	1

43	6
44	4
45	4
46	3
47	5
49	4
50	5
51	4

53	3
54	1
55	2
58	1
Total	50

→ cumulative frequency distribution

C.I	frequency	less than cumulative frequency
25-30	1	1
30-35	2	1+2=3
35-40	5	3+5=8
40-45	10	8+10=18
45-50	16	18+16=34
50-55	13	34+13=47
55-60	3	47+3=50
Total	50	

→ cumulative frequency distribution

C.I	frequency	greater than cumulative frequency
25-30	1	49+1=50
30-35	2	47+2=49
35-40	5	42+5=47
40-45	10	32+10=42
45-50	16	16+16=32
50-55	13	3+13=16
55-60	3	3
Total	50	

DATE: / /
PAGE:

Statistics

Main fn of Statistics

- data → collect the data
- → classification / data manipulation technique.
- → Present the data
- → Analysis
- → conclusion

Statistics is a science that helps us make decisions & draw conclusion in presence of uncertainity & Variablity.

Data:- Data can be defined as systematic record of particular quantity. It is a collection of fact & fig to be used for a specific purpose such as survey or analysis. when arranged in an organised form can be called as information. eg:- collection of income & expenditure of all household in Karnataka state represent the data.

Primary data:-

The data which are collected from unit or individual directly for the purpose of certain study are known as primary data. The survey method are most useful method to collect primary data.

eg:- Data publish by other authorities in original form are considered as primary data.

DATE: / /
PAGE:

secondary data :- The data which had been collected by some individual or agency & Statistically treated to draw certain conclusion. The same data are used & analyzed to extract some other information, are termed as secondary data.

eg :- Trends of population growth, change in sex ratio, mortality rate etc.

Source of secondary data :- different publications of trade & commerce association, summarized data of experiment conduct in engineering laboratories.

Variable :- A variable is any characteristics whose value may change from one object to another in the population. some of eg of variable are :- Age, Height, weight etc.

Types of Variable :-

① Quantitative Variables → (discrete variable) (continuous variable)
② Qualitative Variables

Quantitative Variable :- A variable which can be measured numerically is known as Quantitative Variable eg :- Age, height, weight etc.

Qualitative variables :- A variable which cannot be measured numerically but measured based on Quality is known as Qualitative variable.

eg :- occupation, gender, machine type etc.

⇒ continuous data is convert into discrete data, But discrete data is not converted into continuous data

DATE: / / PAGE:

Raw data :- data which can be collected for a Variable at available time is known as raw data

eg :- let variable is "Marks of the students". The collection of all marks of student in class consitutes a raw data

Marks of Students (n) :

25, 96, 45, 63, 45, 75, 65, 20, 15, 46
35, 20, 34, 45, 63, 40, 55, 05, 82, 89
65, 38, 41, 58, 60, 72, 76, 83, 88, 95

Tabulation of data :- A table is a systematic arrangement of statical data in columns & rows. Rows are horizontal arrangement whereas column are vertical. The purpose of tabulation is to simpling the presentation and to facilate comparisons.

⇒ Advantages of Tabulation :- Tables make it possible for the analyst to present a huge mass of data in detailed manner within minimum space.

DATE: / / PAGE:

Frequency distribution :- A frequency distribution is useful statical tool for originating raw data into some meaningful form. The frequency distributions is two types :-

① Discrete frequency distributions (without class interval)

② continuous frequency distributions (with class interval)

[Q.] The following data represents the weight of 30 participants in a competition :-

52, 74, 40, 60, 82, 100, 75, 83, 53, 80,
95, 50, 109, 79, 105, 48, 55, 62, 95, 60,
69, 77, 70, 49, 88, 96, 80, 99, 108, 56

(i) Prepare a frequency table with class interval 40-50, 50-60, .. etc

(ii) determine the relative frequency

(iii) Draw Histogram

Sol^n :- let x = weight of participants

Max = 109
Min = 40

C-I	Tally marks	frequency	Relative frequency
40-50	III	3	$\frac{3}{30} = 0.10$
50-60	$\cancel{IIII}$	5	$\frac{5}{30} = 0.166 = 0.17$
60-70	IIII	4	$\frac{4}{30} = 0.13$
70-80	IIII	4	$\frac{4}{30} = 0.13$
80-90	$\cancel{IIII}$	5	$\frac{5}{30} = 0.17$
90-100	$\cancel{IIII}$	5	$\frac{5}{30} = 0.17$
100-110	IIII	4	$\frac{4}{30} = 0.13$
		30	

Adding all we always get 1

Arithmetic mean = $\frac{\text{Sum of observation}}{\text{Total no. of observation}}$

5 3 2 1

$\frac{5+3+2+1}{4} = \frac{11}{4} = 2.5$

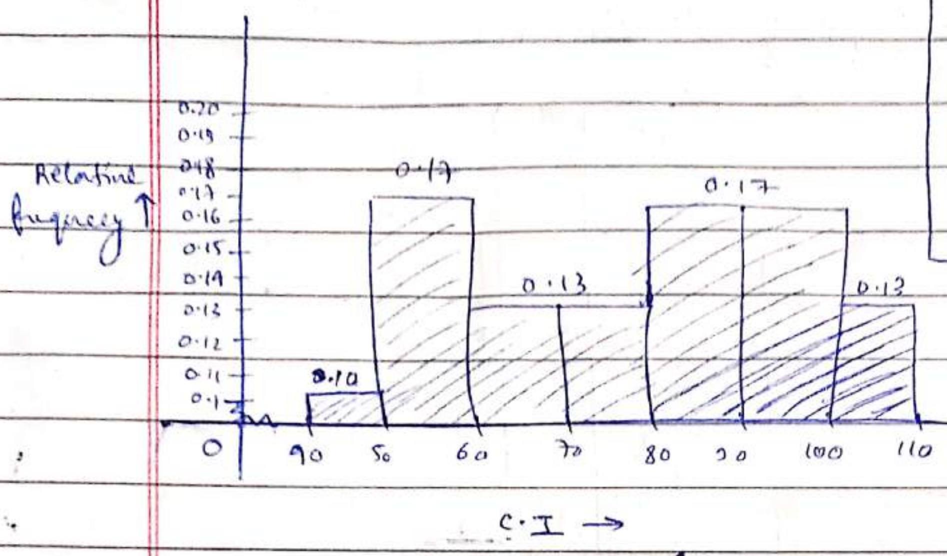
convert inclusive to exclusive

39.5 - 49.5

49.5 - 59.5

59.5 - 69.5

69.5 - 79.5



Relative frequency	New C-I

Measures of Central Tendency

1. Arithmetic mean, mean, Average
2. Median
3. Mode
4. Geometric mean
5. Harmonic mean

→ Various measures of central tendency

1 ⇒ $x_1 \quad x_2 \quad x_3 \quad x_4$

5 3 2 1

$\boxed{\text{Arithmetic Mean}}$

$\frac{5+3+2+1}{4} = \frac{11}{4} = 2.5$

x = Marks of the students

$\bar{x} = \frac{x_1+x_2+\dots+x_n}{n}$

$$\boxed{\bar{x} = \frac{x_1+x_2+\dots+x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

Imp

* case(i) :- without frequency | case(ii) :- Discrete frequency distribution | case(iii) :- continuous frequency distribution

case(i) :- without frequency

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

5, 9, 2, 3, 4, 5, 2, 1, 3, 1, 4

$$\bar{x} = \frac{5+9+2+\dots+4}{11}$$

$= 3.54$

→ diagram are part of case-(i)

case(ii) :- Discrete frequency distribution

x	f	fx
1	2	2
2	2	4
3	2	6
4	2	8
5	2	10
9	1	9

3.54

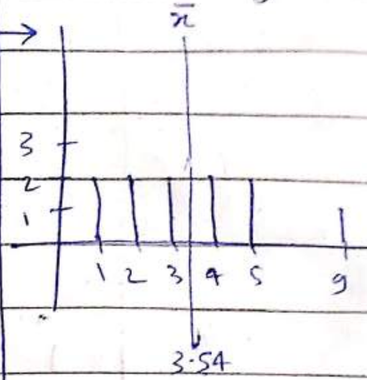
$\sum fx = 39$

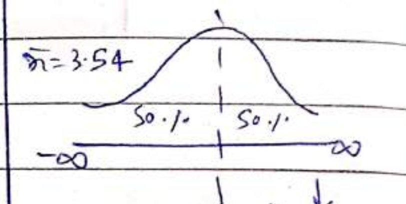
total no. of observation = 11 i.e $\sum f = 11$

1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 9

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{39}{11} = 3.54$$

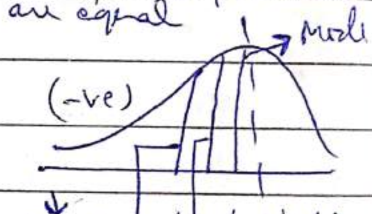
case(iii) :- continuous frequency distribution





$\bar{x}$ ⇒ Symmetric distribution

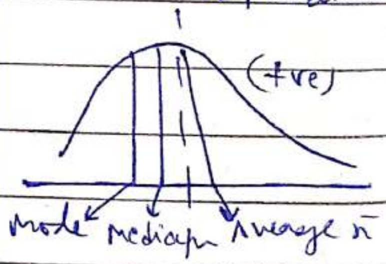
⇒ In symmetric distribution mean, mode, median are equal



Non symmetric distribution

Average $\bar{x}$

Median



Mode Median Average $\bar{x}$

$$\frac{\sum fx}{\sum f} = \frac{5\times3 + 15\times5 + 25\times10 + 35\times7 + 45\times2}{27}$$

$= 24.6296$

case(iii) :- continuous frequency distribution

C.I	frequency	x = mid value of C.I
0 - 10	3	$\frac{0+10}{2} = 5$
10 - 20	5	15
20 - 30	10	25
30 - 40	7	35
40 - 50	2	45

$$\bar{x} = \frac{\sum fx}{\sum f}$$

x = mid value of the C.I

DATE: / /
PAGE:

Arithmetic Mean :- The arithmetic mean of set of observation is defined as the sum of observations divided by number of observations

eg:- calculate the mean of observation :- 3, 5, 1, 6, 2, 7

Sol^n :- $\text{Mean} = \frac{3+5+1+6+2+7}{6} = \frac{24}{6} \Rightarrow 4$

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$ $\boxed{\text{here, } n \text{ is the no. of observations}}$

[Q.2] The following data represents the ages (in years) of 10 policy holders:-
24, 26, 30, 15, 35, 22, 10, 32, 26 & 28

Find average age of policy holder.

Sol^n :- $\text{Average age} = \frac{24+26+\dots+28}{10} = \frac{248}{10} \Rightarrow 24.8$

Arithmetic Mean with Frequency :- let $x_1, x_2 \dots x_n$ be the set of observation of variable x with frequency $f_1, f_2, \dots f_n$ respectively. The arithmetic mean of x is denoted by $\bar{x}$ and is calculated using formula :-

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \dots f_nx_n}{f_1 + f_2 + \dots + f_n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

DATE: / /
PAGE:

(Q) The following frequency table represents the starting salaries in (* 1000 Rupees/annum) of 30 graduating seniors of a liberal arts college :-

Starting salary	25	30	35	40	45	48	52
No. of students	5	8	4	7	3	2	1

calculate average starting salary of the graduates.

Soln :-

Starting salary (x)	No. of students (f)	fx
25	5	$25 \times 5 = 125$
30	8	$30 \times 8 = 240$
35	4	$35 \times 4 = 140$
40	7	$40 \times 7 = 280$
45	3	$45 \times 3 = 135$
48	2	$48 \times 2 = 96$
52	1	$52 \times 1 = 52$
	30	1068

$\bar{x} = \frac{1068}{30} \Rightarrow 35.6$

The avg. salary is Rs. 35.6×1000

$\Rightarrow$ Rs. 35600

Arithmetic mean for Grouped Frequency :- If variable x is represented using class interval with corresponding grouped frequency, then arithmetic mean of x is denoted by $\bar{x}$ & it is calculated using formula :-

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \dots + f_nx_n}{f_1 + f_2 + f_3 + \dots + f_n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

DATE: / /
PAGE:

[8] The life of 80 condensers obtained in a life testing experiment has given below :-

life of condensers (in years)	0 - 2	2 - 4	4 - 6	6 - 8	8 - 10
No. of condensers	15	25	10	18	12

Calculate the avg life of condensers.

Sol^n :-

life of condenser (in years)	No. of condensers (f)	x = mid value of C.I	fx
0 - 2	15	$\frac{0+2}{2} = 1$	$15 \times 1 = 15$
2 - 4	25	$\frac{2+4}{2} = 3$	$25 \times 3 = 75$
4 - 6	10	5	$10 \times 5 = 50$
6 - 8	18	7	$18 \times 7 = 126$
8 - 10	12	9	$12 \times 9 = 108$
Total	80		374

$\bar{x} = \frac{374}{80} = 4.675$ (The avg age of condensers is 4.675 years)

Pooled or combined mean :- if $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots \bar{x}_k$ be mean of k-groups/items with corresponding sample size $n_1, n_2, n_3 \ldots n_k$ respectively. then pooled or combined mean is calculated using formula

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + \cdots + n_k}$$

[9] During a special promotion, a discount chain sold 575, 410 and 520 microwave ovens in three of its store at average price of Rs. 4950, Rs. 5250 & Rs. 5000 respectively. what is the mean price of the ovens sold ?

Sol^n :- $n_1 = 575$, $n_2 = 410$, $n_3 = 520$

$\bar{x}_1 = 4950$, $\bar{x}_2 = 5250$, $\bar{x}_3 = 5000$

$$\bar{\bar{x}} = \frac{(575 \times 4950) + (410 \times 5250) + (520 \times 5000)}{575 + 410 + 520}$$

$= 5049$

⇒ The average price of the microwave ovens in Rs. 5049

DATE: / /
PAGE:

[Q.] In a survey of 5 cement companies, the profit (in Rs. lakh) earned during a year was 15, 20, 10, 35 & 32. Find the average profit of the companies.

Sol:- Average profit of the companies, $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

$$= \frac{15+20+10+35+32}{5} \Rightarrow 22.4$$

[Q.] From the following information on the number of defective components in 1000 boxes :-

No. of defective components	0	1	2	3	4	5	6
No. of Boxes	25	306	402	200	51	10	6

Calculate average no. of defective components for the whole of the production line.

Sol :-

No. of defective components (x)	No. of Boxes (f)	fx
0	25	25 × 0 = 0
1	306	306 × 1 = 306
2	402	402 × 2 = 804
3	200	200 × 3 = 600
4	51	51 × 4 = 204
5	10	10 × 5 = 50
6	6	6 × 6 = 36
	1000	2000

$\bar{x} = \frac{2000}{1000} \Rightarrow 2$

∴ Avg number of defective component for the whole of production line is :- 2

DATE: / /
PAGE:

(Q.) The following distribution gives the pattern of overtime work done by 100 employees of a company.

overtime hours	10-15	15-20	20-25	25-30	30-35	35-40
No. of Employees	11	20	35	20	8	6

calculate the avg overtime work done per employee

Soln:-

overtime hours	No. of Employees (f)	x = mid value of C.I	fx
10-15	11	$\frac{10+15}{2} = 12.5$	$12.5 \times 11 = 137.5$
15-20	20	$\frac{15+20}{2} = 17.5$	$17.5 \times 20 = 350$
20-25	35	$\frac{20+25}{2} = 22.5$	$22.5 \times 35 = 787.5$
25-30	20	$\frac{25+30}{2} = 27.5$	$27.5 \times 20 = 550$
30-35	8	$\frac{30+35}{2} = 32.5$	$32.5 \times 8 = 260$
35-40	6	$\frac{35+40}{2} = 37.5$	$37.5 \times 6 = 225$
	100		2310

$$\bar{x} = \frac{2310}{100} \Rightarrow 23.1$$

∴ avg overtime work done per employee is 23.1

[Q] There are two units of an automobiles company in two different cities employing 760 and 800 persons respectively. The arithmetic means of monthly salaries paid to persons in these two units are Rs. 18750 and Rs. 16,950 respectively. Find the combined arithmetic mean of salaries of the employees in both the units

Soln:-

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1+n_2+n_3}$$

$$= \frac{760 \times 18750 + 800 \times 16950}{760+800} \Rightarrow \frac{14250000 + 13560000}{1560}$$

$$\Rightarrow \frac{27810000}{1500} \simeq 17826.9$$

DATE:
PAGE:

[Q.] The mean yearly salary paid to all employees in a company was Rs. 24,00,000. The mean yearly salaries paid to male & female employees were Rs. 28,00,000 and Rs. 19,00,000 respectively. Find out the percentage of male to female employees in the company.

Soln :- let consider mean salaries of male as 25,00,000 and female as 19,00,000

$$\frac{25,00,000x + 19,00,000y}{x+y} = 24,00,000$$

$$25,00,000x + 19,00,000y = 24,00,000x + 24,00,000y$$

$$100000x = 500000y$$

$$\therefore x = 5y$$

$\therefore$ Percentage of males $= \frac{5y}{6y} \times 100 \Rightarrow 83.33\%$

$\therefore$ Percentage of females $= 100 - 83.33\% = 16.67\%$

$\Rightarrow$ (males : Females = 83.33 : 16.67)

Median :- Median of distribution is the value of the variable which divides into two equal parts. And median is a positional average.

Median without Frequencies :- let $x_1, x_2, x_3, \ldots x_n$ be the set of n observation of variable x. Then median is calculated using following Procedure.

$\boxed{\text{Step1:-}}$ $\Rightarrow$ Arrange the observation in ascending or descending order

$\boxed{\text{Step-2:-}}$ $\Rightarrow$ If "n" is odd, then $\left(\frac{n+1}{2}\right)^{th}$ value is the median.

$\boxed{\text{Step-3:-}}$ $\Rightarrow$ If "n" is even, then median is $\left(\frac{n}{2}\right)^{th}$ & $\left(\frac{n+2}{2}\right)^{th}$ value.

DATE: / /
PAGE:

[Q.1] Actual waiting time (in months) for the first job on the selected sample of nine people having different field of specialization are 1.6, 11.7, 14.0, 3.6, 5.7, 8.9, 6.3, 2.2, 1.9. Find median waiting time.

Soln :- let n = waiting time for the first job in months

ascending order :- 1.6, 1.9, 2.2, 3.6, 5.7, 6.3, 8.9, 11.7, 14.0

Since $n = 9$ is odd. $\left(\frac{9+1}{2}\right) \Rightarrow \left(\frac{10}{2}\right) \Rightarrow 5$.

The 5.7 is the median.

[Q.2] The monthly income (in thousands) of 10 households are 50, 76, 30, 26, 15, 67, 83, 90, 55 & 35. Find the median income of the families.

Soln :- let n = monthly income (in thousands)

Ascending order :- 15, 26, 30, 35, 50, 55, 67, 76, 83, 90

Since, $n = 10$ is even, then $\left(\frac{10}{2}\right) = 5$, the median is average of 50 & 55

$\left(\frac{10+2}{2}\right) \Rightarrow 6$

$$\text{Median} = \frac{50+55}{2} = 52.5$$

Median with frequency :- let $x_1, x_2, \ldots x_n$ be the set of n observations of variable x with frequency $f_1, f_2, \ldots f_n$. The median calculated using following procedure.

$\boxed{\text{Step1:-}}$ ⇒ Arrange the observations in ascending / descending order

$\boxed{\text{Step2:-}}$ ⇒ Find commulative frequecies

$\boxed{\text{Step3:-}}$ ⇒ Find $N/2$, where N = Total frequency

$\boxed{\text{Step4:-}}$ ⇒ Identify commulative frequency just greater than $N/2$

$\boxed{\text{Step5:-}}$ ⇒ Corresponding value of n is the median.

DATE: __/__/__
PAGE: ____

[8.] The following frequency table represents the starting salaries (in '000 Rupees/annum) of 30 graduating seniors of a liberal art college :-

Starting salary	25	30	35	40	45	48	52
No. of Students	5	7	10	15	6	3	1

Calculate the median starting salary of the graduates.

Soln :-

Starting salary (x)	No. of Students (f)	cumulative frequency
25	5	5
30	7	12
35	10	22
40	15	37
45	6	43
48	3	46
52	1	47
	47	

$N = 47$

$\frac{N}{2} = 23.5$

The cumulative frequency just greater than $N/2 = 23.5$ is 37

⇒ Therefore, the median starting salary value is 40

Median for Grouped Frequency :- If variable x is represented using class interval with grouped frequency then the median of x is calculated by using formula :-

$$\text{Median} = L + \left[\frac{\frac{N}{2} - m}{f}\right] \times c$$

where, $L \rightarrow$ is the lower limit of median class

$f \rightarrow$ is the frequency of median class

$c \rightarrow$ is the magnitude of median class

$m \rightarrow$ is the cumulative frequency of the class preceding the median class.

$N \rightarrow$ is total frequency

$$L + \frac{\left(\frac{N}{2} - m\right)}{f} * c$$

DATE: __/__/__
PAGE: ____

Find the median wage of the following distribution :-

wages (in '000)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of employees	3	7	10	12	8	4	2

wages (in '000)	No. of employees	cumulative frequency
10-20	3	3
20-30	7	10
30-40	10	20
40-50	12	32
50-60	8	40
60-70	4	44
70-80	2	46

$N = 46$, $\frac{N}{2} = 23$

Median class is 40-50

$L = 40$, $m = 20$

$f = 12$, $c = 10$

$$\text{Median} = 40 + \left[\frac{23-20}{12}\right] * 10$$

$$\text{Median} = 42.5$$

Mode :-

Mode is the most frequently occured observation in the given data.

Find the mode of the data given data :-

3, 5, 6, 7, 3, 6, 8, 9, 4, 3 ~~8~~ 10

Since observation 3 is repeated more no. of times, the mode of given data is 3

(X) → Not part of syllabus

DATE: / /
PAGE:

Mode for grouped data :- If variable x is represented using class interval with corresponding grouped frequency, then median of x is calculated using the formula :-

$$\text{Mode} = L + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times c$$

Where,

- $L \to$ lower limit of the modal class
- $f_1 \to$ frequency of modal class
- $f_0 \to$ preceding frequency of modal class
- $f_2 \to$ succeeding frequency of modal class
- $c \to$ width of modal class

[Q] The following distribution gives the pattern of overtime work done by 100 employees of a company.

overtime hours	10-15	15-20	20-25	25-30	30-35	35-40
No. of Employees	11	20	35	20	8	6

calculate the mode of overtime work done per employee

Soln :-

overtime (in hours)	No. of employees (f)
10-15	4
15-20	20
20-25	35
25-30	20
30-35	8
35-40	6

Since, the highest frequency is 35 the modal class is 20-25

$L = 20$, $f_1 = 35$, $f_0 = 20$, $f_2 = 20$, $c = 5$

$$\text{Mode} = 20 + \left[\frac{35-20}{2 \times 35 - 20 - 20}\right] \times 5$$

$$= 20 + \left[\frac{15}{30}\right] \times 5$$

$$= 22.5$$

DATE: / /
PAGE:

[Q.1] calculate the median of the following data that relates to the service time (in minutes) per customer for 7 customer at railway reservation counter :- 3.5, 4.5, 3, 3.8, 5.0, 5.5 and 4

Soln :- let n = service time per customer at railway reservation counter

Ascending order :- 3, 3.5, 3.8, 4, 4.5, 5.0, 5.5

Since, $n = 7$ is odd . $\left(\frac{7+1}{2}\right) \Rightarrow \left(\frac{8}{2}\right) \Rightarrow 4$

The 4 is the median.

[Q.2] The following data represents the no. of hours working by each employee using office computer in company.

7.5, 8, 9.6, 5.6, 4.5, 6.3, 5.6, 5.5, 6.0

Find the mode using the above data

Soln:- Since, the observation 5.6 is repeated more no. of times so, the mode for the given data is 3

[Q.3] The following distribution given the marks of 45 students in semester examination.

Marks	0-10	10-20	20-30	30-40	40-50
Number of students	2	5	10	20	8

calculate the mode.

Soln:-

Marks	No. of students (f)
0-10	2
10-20	5
20-30	10
30-40	20
40-50	8

Since, highest frequency is 20 the modal class is 30-40

$L = 30$, $f_1 = 20$, $f_0 = 10$, $f_2 = 8$, $C = 10$

$$\text{Mode} = 30 + \left[\frac{20-10}{2\times20-10-8}\right] \times 10$$

$$= 30 + \left[\frac{10}{22}\right] \times 10$$

$= 30 + 0.4545 \times 10$

$= 34.5454$

DATE: / /
PAGE:

Measure of Dispersion :- The Variability of observation from any given value is known as the dispersion of the given data. The measure of dispersion generally derived from its mean value.

Range :- The range is the difference b/w two extreme observation. if L = maxm value, & S = Minm value in the given data

$$\therefore \text{Range} = L - S$$

[Q] If the processing time (in hrs) of a particular code through to personal computer is 2, 5, 3.5, 4.2, 3, 2.9, 4.4, 5.1, 6.0 & 6.5. then find the range

Soln :- Range = $6.5 - 2 = 4.5$

☆ → The coefficient of range is known as relative measure

☆ → It is used to compare variability more than two groups

☆ → The formula to measure the coefficient range is

$$\therefore \text{coeff of Range} = \frac{L-S}{L+S}$$

☆ → The coefficient of range always lies b/w 0 & 1

DATE: / /
PAGE:

[Q] The following data represent the sales of firm for the last 10 months

Month	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct
Sales (000)	85	92	80	79	68	70	90	83	98	88

calculate the range & coefficient of range.

Sol^n :- Range $= 98 - 68 = 30$

coefficient of Range $= \frac{30}{166} = 0.18$

The standard deviation is always positive.

Variance :- The square of S.D is known as variance & denoted by σ^2. The variance is always (+ve)

$$\boxed{S.D = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}}$$

DATE: / /
PAGE:

[Q] The wholesale prices of a commodity for seven consecutive days in a month as follows :-

Days	1	2	3	4	5	6	7
commodity price/ quintal	240	260	270	245	255	286	264

calculate the variance & standard deviation.

Sol^n :- $\bar{x} = \frac{240+260+270+245+255+286+264}{7}$

$= \frac{1820}{7} \Rightarrow 260$

Days	x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
1	240	260	-20	400
2	260	260	0	0
3	270	260	10	100
4	245	260	-15	225
5	255	260	-5	25
6	286	260	26	676
7	264	260	4	16
				$\Sigma(x-\bar{x})^2 =$ 1442

rough

x	$(x-\bar{x})^2$
240	$(240-260)^2 =$
260	$(210-260)^2 =$

$\Sigma x = 1820$ $\quad \Sigma(x-\bar{x})^2 =$

$\bar{x} = \frac{1820}{7} = 260$

$\Sigma(x-\bar{x})^2 = 1442$

$\sigma = \sqrt{\frac{1442}{7}} = 14.35$

$$\boxed{\sigma = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$$

$\sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n} = \frac{1442}{7} \Rightarrow 206 \rightarrow$ variance

$\sigma = \sqrt{206} \Rightarrow 14.352 \rightarrow$ S.D

DATE: / /
PAGE:

[Q.] The scores of two batsman in their recent 5 innings are given below :-

Batsman A = 32, 45, 65, 67, 90

Batsman B = 100, 3, 45, 50, 10

using the above data identify the more consistent player.

Soln :-

Batsman A	Batsman B
$\overline{X_A} = \dfrac{32+45+65+67+90}{5}$	$\overline{X_B} = \dfrac{100+3+45+50+10}{5}$
$= 59.8$	$= 41.6$

S.D $\sigma_A = \sqrt{\dfrac{772.84 + 219.04 + 27.04 + 51.84 + 912.04}{5}} \Rightarrow \sqrt{\dfrac{1982.8}{5}}$

$= \sqrt{396.56} \Rightarrow 19.91$

For A

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
32	59.8	-27.8	772.84
45	59.8	-14.8	219.04
65	59.8	5.2	27.04
67	59.8	7.2	51.84
90	59.8	30.2	912.04
			$\Sigma(x-\bar{x})^2 = 1982.8$

For B

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
100	41.6	58.4	3410.56
3	41.6	-38.6	1489.96
45	41.6	3.4	11.56
50	41.6	8.4	70.56
10	41.6	-31.6	998.56
			$\Sigma(x-\bar{x})^2 = 5980.8$

S.D $\sigma_B = \sqrt{\dfrac{3410.56 + 1489.96 + 11.56 + 70.56 + 998.56}{5}} = \sqrt{\dfrac{5980.8}{5}}$

$= \sqrt{1196.24} \Rightarrow 34.586$

$CV_A = \frac{19.91}{57.8} \times 100$ $\qquad$ $CV_B = \frac{34.586}{41.6} \times 100$

$= 33.3\%$ $\qquad$ $= 83.1\%$

$\therefore$ Player A is more consistent

[Q.] A computer company received a rush order for as many home computers as could be shipped during a six week period. company records provide the following daily shipments:-

22, 65, 65, 67, 55, 50, 65, 77, 73, 30, 62, 54, 48, 65, 79, 60, 63, 45, 51, 68, 79, 83, 33, 41, 49, 28, 55, 61, 65, 75, 39, 87, 45, 50, 66, 65, 59, 25, 35 & 53

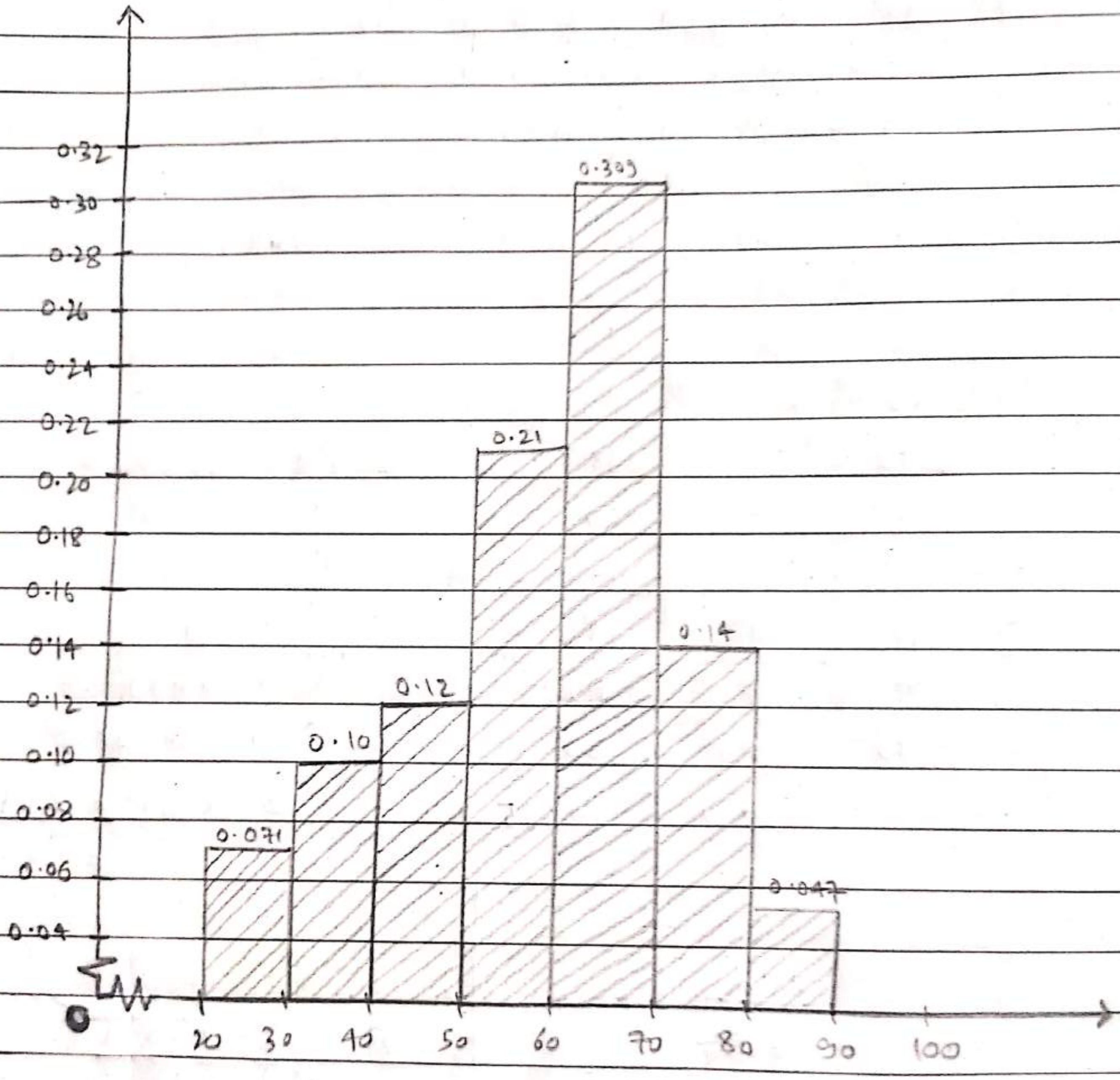
(i) construct a frequency distribution with the class interval 20-30, 30-40, ...

(ii) Determine the relative frequency

(iii) construct the histogram & comment on it

Soln:-

C.I	Tally marks	Frequency	Relative frequency
20-30	III	3	$\frac{3}{42} = 0.071$
30-40	IIII	4	$\frac{4}{42} = 0.$[redacted]095
40-50	卌	5	$\frac{5}{42} = 0.119$
50-60	卌 IIII	9	$\frac{9}{42} = 0.214$
60-70	卌 卌 III	13	$\frac{13}{42} = 0.$[redacted]309
70-80	卌 II	6	$\frac{6}{42} = 0.142$
80-90	II	2	$\frac{2}{42} = 0.047$
		42	1


0.32
0.30
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0
0.071
0.10
0.12
0.21
0.309
0.14
0.047
20
30
40
50
60
70
80
90
100

DATE: / /
PAGE:

[Q] The processing time of 10 personal computer are 56, 45, 32, 50, 50, 80, 66, 48, 35, 75. Determine the average & S.D of processing time.

Solⁿ :-

x	f	fx	$(x-\bar{x})^2$
32	1	32	$(32-53.7)^2 = 470.89$
35	1	35	$(35-53.7)^2 = 349.69$
45	1	45	$(45-53.7)^2 = 75.69$
48	1	48	$(48-53.7)^2 = 32.49$
50	2	100	$(50-53.7)^2 = 13.69 + 13.69$
56	1	56	$(56-53.7)^2 = 5.29$
66	1	66	$(66-53.7)^2 = 151.29$
75	1	75	$(75-53.7)^2 = 453.69$
80	1	80	$(80-53.7)^2 = 691.69$
	$\Sigma f = 10$	$\Sigma fx = 537$	$\Sigma (x-\bar{x})^2 = 2244.41 + 13.69 = 2258.1$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{537}{10} \Rightarrow 53.7$$

$$\sigma = \sqrt{\frac{2258.1}{10}} = 15.026 \quad \text{S.D}$$

Quartile Deviation :- If the three quartiles be Q_1, Q_2, Q_3 then the interquartile deviation and quartile deviation are calculated as :-

$IQD = Q_3 - Q_1$

$QD = \frac{Q_3 - Q_1}{2}$

$CQD = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

(i) interquartile deviation (IQD) $= Q_3 - Q_1$

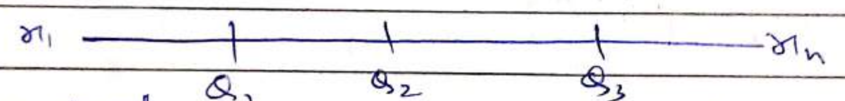
(ii) Quartile deviation (Q.D) $= \frac{Q_3 - Q_1}{2}$

(iii) coefficient of Quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

→ The coefficient of Quartile deviation always lies b/w 0 & 1

Quartile :- The quartile are numerical values which will divide the entire data into four equal parts.

let $x_1, x_2, \ldots, x_n$ be set of n observation of random variable x, which were arranged in ascending order of magnitude.



the total no. of observation divided into four equal parts with 25% of observation in each quarter.

Quartile :-

Step-1 :- Arrange the observation in ascending order

Step-2 :- use the following formulae to find the Quartiles

$$Q_i = \left[\frac{i(n+1)}{4}\right]^{th} \text{ value}, \quad i = 1, 2, 3, \text{etc.}$$

→ First Quartile :- $Q_1 = \left[\frac{1 \times (n+1)}{4}\right]^{th}$ value in the ordered series.

→ Second Quartile :- $Q_2 = \left[\frac{2 \times (n+1)}{4}\right]^{th}$ value in the ordered series.

→ Third Quartile :- $Q_3 = \left[\frac{3 \times (n+1)}{4}\right]^{th}$ value in the ordered series.

→ nth Quartile :- $Q_k = \left[\frac{k \times (n+1)}{4}\right]^{th}$ value in the ordered series.

If the no. is a decimal i.e

$Q_2 = 2.75^{th}$ value

$= (2+0.75)^{th}$ value

$=$ 2nd value $+ 0.75$ (3rd value $-$ 2nd value)

Percentiles

$$P_k = \left[\frac{k \times (n+1)}{100}\right]^{th} \text{ value} \qquad k = 1, 2, 3, \ldots, 99$$

Decile

$$D_k = \left[\frac{k \times (n+1)}{10}\right]^{th} \text{ value} \qquad k = 1, 2, 3, \ldots 9$$

[Q] Calculate the 1st 3 quartile from the following data. Also calculate quartile deviation.

5, 1, 9, 2, 15, 12, 8, 11, 19, 22

Soln:- 1, 2, 5, 8, 9, 11, 12, 15, 19, 22

$$Q_1 = \left[\frac{1 \times (10+1)}{4}\right] = \frac{11}{4} = 2.75^{th} \text{ value}$$

→ Q_1 = 2nd value $+ 0.75$ (3rd value $-$ 2nd value)

$= 2 + 0.75 \times (5-2)$

$= 2 + 0.75 \times 3$

$= 2 + 2.25 \Rightarrow 4.25$

→ $Q_2 = \left[\frac{2 \times (10+1)}{4}\right]^{th}$ value

$= \left[\frac{22}{4}\right]^{th}$ value

$= 5.5^{th}$ value

$= 5^{th}$ value $+ 0.5$ ($6^{th} - 5^{th}$)

$= 9 + 0.5(11-9)$

$= 9 + 0.5 \times 2$

$= 10$

DATE: / /
PAGE:

→ $Q_3 = \left[\frac{3 \times (10+1)}{4}\right]^{th}$ value

$= \left[\frac{33}{4}\right]^{th}$ value

$= 8.25^{th}$ value

$= 8^{th}$ value $+ 0.25\,(9^{th}$ value $- 8^{th}$ value$)$

$= 15 + 0.25(19-15)$

$= 15 + 1$

$= 16$

→ $Q.D = \frac{16 - 4.25}{2}$

$= \frac{11.75}{2} \Rightarrow 5.875$

[Q.] Find the three quartiles using the data :-

5, 1, 3, 8, 4, 10, 2, 15 & 20

Soln :- First arrange the data in ascending order. here $n = 9$

1, 2, 3, 4, 5, 8, 10, 15, 20

First Quartile $= Q_1 = \left[\frac{1 \times (9+1)}{4}\right]^{th}$ value $= 2.5^{th}$ value

Second Quartile $= Q_2 = \left[\frac{2 \times (9+1)}{4}\right]^{th}$ value $= 5^{th}$ value

third Quartile $= Q_3 = \left[\frac{3 \times (9+1)}{4}\right]^{th}$ value $= 7.5^{th}$ value

→ First Quartile $(Q_1) = 2.5^{th}$ value

$= 2^{nd}$ value $+ 0.5\,(3^{rd}$ value $- 2^{nd}$ value$)$

$= 2 + 0.5\,(3-2)$

$= 2 + 0.5$

$Q_1 = 2.5$

→ second Quartile $(Q_2) = 5^{th}$ value $= 5$

→ Third Quartile (Q_3) = 7.5^{th} value

$= 7^{th}$ value $+ 0.5$ (8^{th} value $- 7^{th}$ value)

$= 10 + 0.5(15 - 10)$

$= 10 + 2.5$

$= 12.5$

$\boxed{Q_3 = 12.5}$

→ Inter Quartile Range (I.Q.R) $= Q_3 - Q_1$

$= 12.5 - 2.5$

$= 10$

→ Quartile Deviation (Q.D) $= \dfrac{(Q_3 - Q_1)}{2}$

$= 10/2 \Rightarrow 5$

Deciles :-

step-1 :- Arrange the observations in ascending order

step-2 :- use the following formula to find the deciles

$$D_i = \left[\frac{i \times (n+1)}{10}\right]^{th} \text{ value}, \quad i = 1, 2, 3, \ldots 10$$

First decile $= D_1 = \left[\dfrac{1 \times (n+1)}{10}\right]^{th}$ value

5th decile $= D_5 = \left[\dfrac{5 \times (n+1)}{10}\right]^{th}$ value

Percentiles :-

step-1 :- Arrange the observations in ascending order

step-2 :- use the following formula to find Percentiles.

$$P_i = \left[\frac{i \times (n+1)}{100}\right]^{th} \text{ value}, \quad i = 1, 2, 3, \ldots 100$$

25th Percentile $= P_{25} = \left[\dfrac{25 \times (n+1)}{100}\right]^{th}$ value

50th Percentile $= P_{50} = \left[\dfrac{50 \times (n+1)}{100}\right]^{th}$ value

DATE: __/__/__
PAGE:

[Q] calculate 5^{th} decile, 30^{th} Percentile from following data :-

5, 1, 3, 8, 4, 10, 2, 15, 20

Soln :- 1, 2, 3, 4, 5, 8, 10, 15, 20

$D_5 = \left[\frac{5 \times (9+1)}{10}\right]^{th}$ value

$= 5^{th}$ value

$= 5$

$P_{30} = \left[\frac{30 \times (9+1)}{100}\right]^{th}$ value

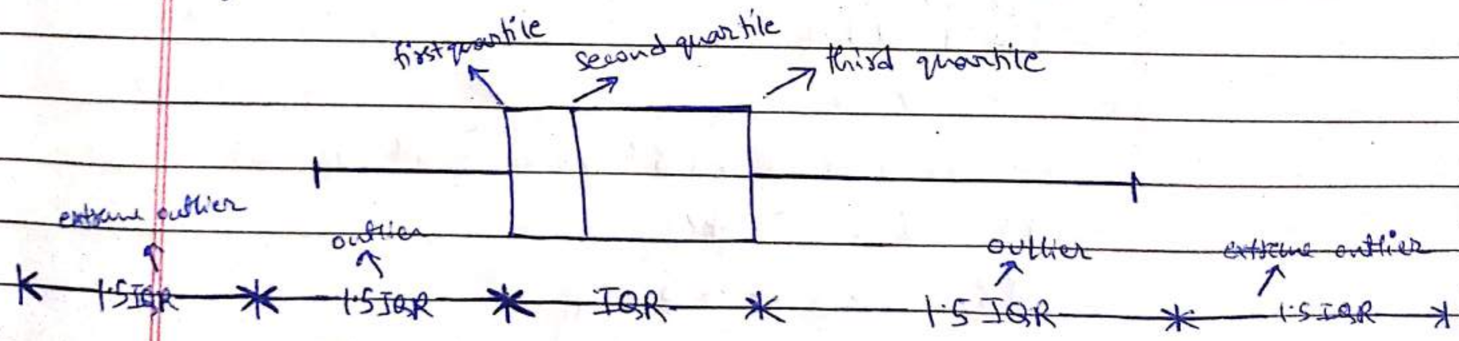
$= 3^{rd}$ Value

$= 3$

Box & whisker Plot :- The Box and whisker plot is graphically display that simultaneously describes several important features of data set. such as centre, spread, departure from symmetry & identification of unusual observation or outliers.

(i) left whisker = $Q_1 - 1.5 \times IQR$

(ii) Right whisker = $Q_3 + 1.5 \times IQR$.



DATE : / /
PAGE :

[9.] The following data represents the processing time of 10 personal computers in seconds 23, 46, 20, 48, 65, 70, 55, 39, 98, 81 construct Box & whisker plot & interpret the diagram.

Soln:- 20, 23, 39, 46, 48, 55, 65, 70, 81, 98

$Q_1 = \left[\frac{1 \times (10+1)}{4}\right]^{th}$ value

$= 2.75$

$= 2^{nd}$ value $+ 0.75$ (3^{rd} value $-$ 2^{nd} value)

$= 23 + 0.75 \; (39 - 23)$

$= 23 + 0.75 \times 16$

$= 23 + 12 \Rightarrow 35$

$Q_2 = \left[\frac{2 \times (10+1)}{4}\right]^{th}$ value

$= \left[\frac{22}{4}\right]^{th}$ value

$= 5.5^{th}$ value

$= 5^{th}$ value $+ 0.5$ (6^{th} value $-$ 5^{th} value)

$= 48 + 0.5 \; (55 - 48)$

$= 48 + 3.5 \Rightarrow 51.5$

$Q_3 = \left[\frac{3 \times (10+1)}{4}\right]^{th}$ value

$= 8.25^{th}$ value

$= 8^{th}$ value $+ 0.25 \times$ (9^{th} value $-$ 8^{th} value)

$= 70 + 0.25 \times (81 - 70)$

$= 70 + 2.75$

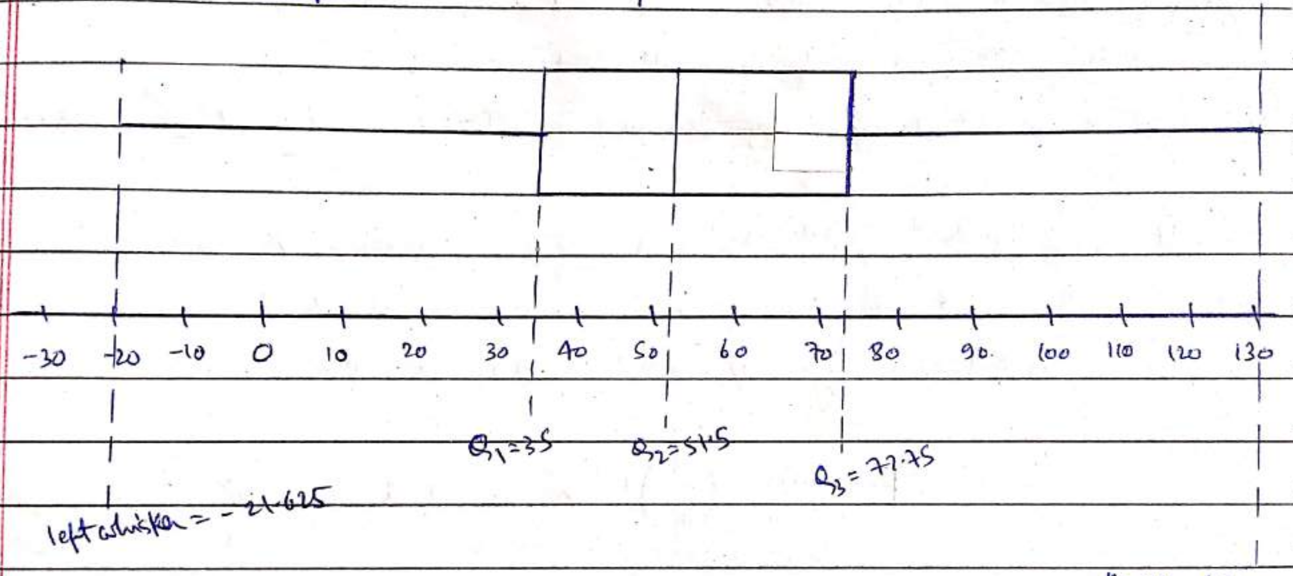
$= 72.75$

$IQR = Q_3 - Q_1$
$= 72.75 - 35$
$= 37.75$

left whisker $= Q_1 - 37.75 \times 1.5$
$= 35 - 37.75 \times 1.5$
$= -21.625$

Right whisker $= Q_3 + 1.5 \times 37.75$
$= 72.75 + 56.625$
$= 129.375$

→ Box & whisker plot



⇒ There is no extreme observation or outlier present in the given data

Random Variable

Discrete
↓
Discrete probability distribution
- → Bernoulli
- → Binomial
- → Poisson
- → Geometric
- → Uniform (Normal)

continuous
↓
Continuous probability distribution
- → Normal
- → exponential
- → uniform

DATE: / /
PAGE:

Bernoulli Distribution :- A random Variable X has Bernoulli distribution with parameter "p", where $0 \le p \le 1$. if its p.m.f $P(X=x)$ is given by

$$P(X=x) = \begin{cases} p & \text{if } x=1 \\ (1-p) & \text{if } x=0 \end{cases}$$

Note :-
(i) X assumes the value 1 & 0 for success & failure respectively
(ii) The mean of Bernoulli distribution ; $E(x) = p$
(iii) The variance of Bernoulli distribution : $Var(x) = p(1-p)$

Binomial distribution :- A discrete random Variable X has a Binomial distribution with parameter "n" & "p", where $n = 1, 2, 3, ...$ & $0 \le p \le 1$, if its p.m.f $P(X=x)$ is given by,

$$P(X=x) = \binom{n}{x} p^x q^{n-x}, \text{ where } q = 1-p$$
$$x = 1, 2, 3, \dots, n$$

Note :-
(i) if X is random Variable of following binomial distribution with Parameter "n" & "p" then it can written as :- $X \sim Binomial(n, p)$
(ii) The total probability is equal to 1
(iii) The mean of B.D :- $E(X) = np$
(iv) The Variance of B.D :- $Var(X) = np(1-p)$

[Q] Ten coins are thrown simultaneously. Find the probability of getting at least seven heads.

Soln :- $n = 10$ $\qquad X \sim BD(10, 0.5)$
let X = no. of heads
$p = \frac{1}{2} = 0.5$
$X = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$

$$\binom{10}{7} = \frac{10\times 9\times 8}{3\times 2\times 1} = 120 \qquad \binom{10}{9} = \frac{10}{1} = 10$$

$$\binom{10}{8} = \frac{10\times 9}{2\times 1} = 45 \qquad \binom{10}{10} = 1$$

DATE: __/__/__
PAGE: ______

$$P(X\geq 7) = P(X=7,8,9,10)$$

$$= P(X=7) + P(X=8) + P(X=9) + P(X=10)$$

$$= \binom{10}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3 + \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right) + \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0$$

$$= 120\times\left(\frac{1}{2}\right)^{10} + 45\left(\frac{1}{2}\right)^{10} + 10\left(\frac{1}{2}\right)^{10} + \left(\frac{1}{2}\right)^{10}$$

$$= \left(\frac{1}{2}\right)^{10}(120+45+10+1)$$

$$= \frac{176}{2^{10}} \Rightarrow \frac{176}{1024} \Rightarrow 0.171875$$

[Q.] Suppose 10% of new scooters will require warranty service within the 1st month of its sale. A scooter manufacturing company sells 1000 scooters in a month. Find the mean & S.D of scooter that require warranty service.

Solⁿ:-

$n = 1000$

$p = 10\% = \frac{10}{100} = 0.1$

Let X = no. of scooter which requires warranty service

$E(X) = \quad SD(X) = \sqrt{Var(X)} = ?$

$X \sim B.D(1000, 0.1)$

$E(X) = np$	$Var(X) = np(1-p)$	$SD(X) = \sqrt{Var(X)}$
$= 1000\times 0.1$	$= 100(1-0.1)$	$= \sqrt{90}$
$= 100$	$= 100\times 0.9$	$= 3\sqrt{10}$
	$= 90$	$= 9.487$
		≈ 9 Scooters

[Q.] The following data represents the processing time of 10 personal computer in seconds.

65, 72, 50, 68, 65, 70, 55, 49, 80, 85.

Construct Box & Whisker plot & interpret the diagram

Sol^n:- Arrange data in ascending order :- 49, 50, 55, 65, 65, 68, 70, 72, 80, 85

$Q_1 = \left[\frac{1\times(10+1)}{4}\right]^{th}$ value

$= 2.75$

$= 2^{nd}$ value $+ 0.75\,(3^{rd}$ value $- 2^{nd}$ value$)$

$= 50 + 0.75\,(55-50)$

$= 50 + 0.75 \times 5$

$= 53.75$

$Q_2 = \left[\frac{2\times(10+1)}{4}\right]^{th}$ value

$= \left[\frac{22}{4}\right]^{th}$ value

$= 5.5^{th}$ value

$= 5^{th}$ value $+ 0.5\,(6^{th}$ value $- 5^{th}$ value$)$

$= 65 + 0.5\,(68-65)$

$= 65 + 0.5 \times 3$

$= 66.5$

$Q_3 = \left[\frac{3\times(10+1)}{4}\right]^{th}$ value

$= 8.25^{th}$ value

$= 8^{th}$ value $+ 0.25 \times (9^{th}$ value $- 8^{th}$ value$)$

$= 72 + 0.25\,(80-72)$

$= 72 + 0.25 \times 8$

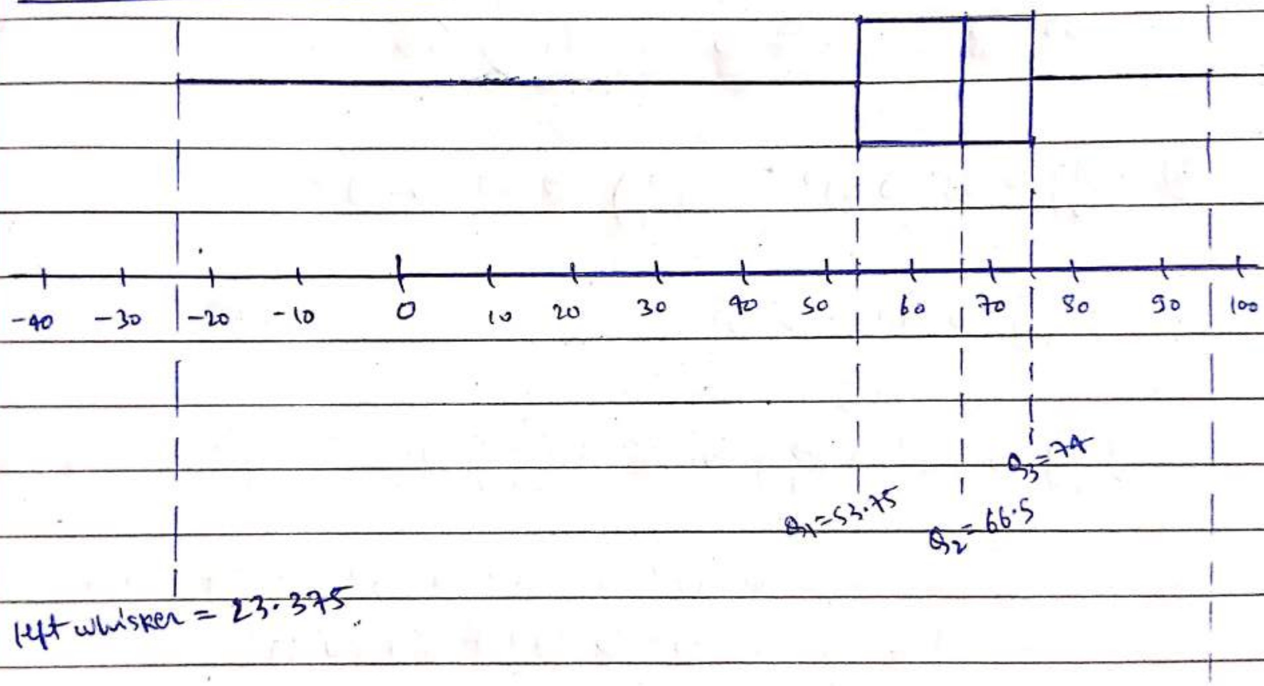
$= 74$

$IQR = Q_3 - Q_1$
$= 74 - 53.75$
$= 20.25$

Left whisker $= Q_1 - 1.5 \times IQR$
$= 53.75 - 1.5 \times 20.25$
$= 23.375$

Right whisker $= Q_2 + 1.5 \times IQR$
$= 66.5 + 1.5 \times 20.25$
$= 96.875$

⇒ Box & Whisker plot



⇒ There is no extreme outlier observation or outlier present in the given data.

[Q.] A brokerage survey reports that 30 per cent of individual investors have used a discount broker i.e one which does not charge the full commission. In a random sample of 9 individuals, what is the probability that

(a) exactly two of the sampled individuals have used a discount broker?
(b) not more than three have used a discount broker
(c) at least three of them have used a discount broker

$\binom{9}{0} = 0$ $\qquad 10C_{10} = 1$

$\binom{9}{1} = 1$

DATE:
PAGE:

Solⁿ:- let X = The no. of investors used the discount broker

$X = 0, 1, 2, 3, 4, \ldots, 9$, $n = 9$

$P = 30\% \Rightarrow \frac{30}{100} \Rightarrow 0.3$

$$\binom{9}{3} = \frac{\cancel{9} \times \cancel{8} \times 7 \times \cancel{6} \times 5 \times 4}{\cancel{6} \times 5 \times 4 \times \cancel{3} \times 2 \times 1}$$

$X \sim BD(9, 0.30)$

$$\binom{9}{2} = \frac{9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3}{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = 36$$

(a) $\binom{9}{2}(0.3)^2(0.3)^7$

$= 36(0.3)^9 \Rightarrow 7.08588 \times 10^{-4} \rightarrow 0.000708588$

(b) $\binom{9}{0}(0.3)^0(0.3)^9 + \binom{9}{1}(0.3)^1(0.3)^8 + \binom{9}{2}(0.3)^2(0.3)^7 +$

$\binom{9}{3}(0.3)^3(0.3)^6$

$= \binom{9}{0}(0.3)^9 + \binom{9}{1}(0.3)^9 + \binom{9}{2}(0.3)^9 + \binom{9}{3}(0.3)^9$

$= 0 \cdot (0.3)^9 + 9(0.3)^9 + 36 \cdot (0.3)^9 + 84(0.3)^9$

$= 0 + 9(0.3)^9 + 36 \cdot (0.3)^9 + 84(0.3)^9$

$= (0.3)^9 (9 + 36 + 84)$

$= (0.3)^9 (129) \Rightarrow 0.002539107$

$\Rightarrow 2.539107 \times 10^{-3}$

(c) $[1 - P(X \le 2)]$

$= 1 - \left[\binom{9}{0}(0.3)^0(0.3)^9 + \binom{9}{1}(0.3)^1(0.3)^8 + \binom{9}{2}(0.3)^2(0.3)^7\right]$

$= 1 - \left[\binom{9}{0}(0.3)^9 + \binom{9}{1}(0.3)^9 + \binom{9}{2}(0.3)^9\right]$

$= 1 - [0 \cdot (0.3)^9 + 9 \cdot (0.3)^9 + 36(0.3)^9]$

$= 1 - (0.000885735)$

$= 1 - 8.85735 \times 10^{-4}$

$= 0.9991142$

$\simeq 0.99$

DATE: ___ PAGE: ___

p.m.f of poisson distribution

Poisson distribution :- A random Variable X has poisson distribution with parameter "λ", where $\lambda > 0$ if its p.m.f $P(X=x)$ is given by :

$$P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad \text{where } X = 0,1,2,\dots$$

[Some application of poisson distribution :- No. of accident on particular road during a week]

Note :- (i) If X is random variable of following poission distribution with parameter λ, then it can written as :- $X \sim \text{Poisson}(\lambda)$

(ii) The total probability is equal to 1

★ (iii) The mean of poisson distribution :- $E(X) = \lambda$

★ (iv) The variance of poisson distribution :- $Var(X) = \lambda$

[Q.] The no. of vehicles joining in a fuel queue in a petrol bunk at particular minute has a poisson distribution with parameter 5.8

$\frac{29}{5} = 5.8$

Find the probability that

(i) No vehicle joins the fuel queue in a particular minute.

(ii) Two or more vehicle joins the fuel queue in particular minute

(iii) Exactly 4 vehicle joins the fuel queue.

Soln :- let X = No. of vehicles joins the fuel queue in a particular minute.

$\lambda = 5.8$

(i) $P(X=0) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

$= \frac{e^{-5.8} \times 5.8^{0}}{0!}$

$= 0.0030$

(ii) $P(X \geq 2) = 1 - P(X < 2)$

$= 1 - P(X = 0,1)$

$= 1 - P(x=0) + P(x=1)$

$= 1 - \left[0.0030 + \frac{e^{-5.8} \times 5.8^{1}}{1!}\right]$

$= 1 - (0.0030 + 0.0174)$

$= 1 - 0.02$

$= 0.98$

0.9796

(iii) $P(X=4) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-5.8}\, 5.8^{4}}{4!}$

$= \frac{(0.0030)(1131.65)}{4!}$

$= 0.1428$

DATE: / /
PAGE:

[Q] On an average a school kid makes 3 mistakes in one page of note. using poisson distribution find the probability that a randomly observed page is free of mistakes. Among 200 pages, in how many pages would you expect mistakes?

Soln:- $\lambda = 3$

X be no. of mistakes

$P(X=0) = \frac{e^{-3} \times 3^0}{0!}$
$= 0.0497$

$P(X \geq 1) = 1 - P(X<1)$
$= 1 - 0.0497$
$= 0.95$

$\Rightarrow 200 \times 0.95 \Rightarrow 190$ pages

[Q] 12% of items produced by machine are defective. what is the probability that out of a sample of 5 items produced by a machine

(a) no defectives
(b) at least one defective
(c) at most two are defective

Soln:- X be the no. of defective items

$X = 1, 2, 3, 4, 5$

$\lambda = 12\% \Rightarrow \frac{12}{100} \Rightarrow 0.12$

$\lambda = np$
$= 5 \times 0.12 = 0.6$

(a) $P(X=0) = \frac{e^{-0.12 \times 5} \times (0.12 \times 5)^0}{0!}$
$= 0.5488$

(b) $P(X \geq 1) = P(X = 1, 2, 3, 4, 5)$
$= 1 - P(X<1)$
$= 1 - P(X=0)$
$= 1 - 0.5488$
$= 0.4512$

DATE: / /
PAGE:

(c) $P(X \le 2) = P(X = 0, 1, 2)$

$$= P(X=0) + P(X=1) + P(X=2)$$

$$= 0.5488 + \frac{e^{-0.6} \times (0.6)^1}{1!} + \frac{e^{-0.6} \times (0.6)^2}{2!}$$

$$= 0.5488 + 0.3292 + 0.0987 \Rightarrow 0.9767$$

[Q.] Suppose the probability that an item produced by certain machine will be defective is 10.1%. Find the probability that a sample of 10 items will contain at most one defective item if the quality of successive items is independent.

Solⁿ :- X be no. of defective items

$\lambda = 0.1 \times 10 = 1$

$$P(X \le 1) = P(X=0) + P(X=1)$$

$$= \frac{e^{-1} \times 1^0}{0!} + \frac{e^{-1} \times 1^1}{1!}$$

$$= 0.3678 + 0.3678$$

$$= 0.7356$$

Normal Distribution :-

if a random variable X is following normal distribution with A random variable X is said to have normal distribution with parameter μ & σ^2 if its p.d.f is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; -\infty < (x, \mu) < +\infty$$

$$; \sigma^2 > 0$$

here μ and σ^2 are the parameter of the normal distribution and are also called as mean & variance of normal distribution.

DATE : / /
PAGE :

Note :- A random variable X is said to have normal distribution with mean μ & variance σ^2, then it can represented as,

$$X \sim Normal(\mu, \sigma^2)$$

Standard Normal Variate :- if a random variable X is following a normal distribution with mean μ and variance σ^2 then,

$$Z = \frac{X-\mu}{\sigma}$$ is known as standard normal variate & it follows normal distribution with

(i) Mean $= E(Z) = 0$
(ii) Variance $= Var(Z) = 1$

[Q] let a random variable X is normally distributed with mean 12 and variance 16. determine the following

(a) $P(X \geq 20)$
(b) $P(X \leq 20)$
(c) $P(0 \leq X \leq 20)$

Soln :- Variance $= 16$ (σ^2)
$\sigma = 4$ **(S.D)**
Mean $= 12$ (μ)

$X \sim Normal(12, 16)$
SD $(x) = \sqrt{Var(x)} = \sqrt{16} = 4$
Variance $= V(x) = \sigma^2 = 16$
Mean $= \mu = 12$

* upper bound − lower bound

DATE: / /
PAGE:

(i) $P(X \geq 20) = P\left[\frac{X-\mu}{\sigma} \geq \frac{20-\mu}{\sigma}\right]$

$= P\left[Z \geq \frac{20-12}{4}\right]$

$= P[Z \geq 2]$

$= P[2 \leq Z \leq \infty]$

$= P[Z=\infty] - P[Z=2]$

$= 1 - P[Z=2]$

= [redacted]

= [redacted] $\frac{3}{0}$

$= 1 - 0.97725 \Rightarrow 0.02275$

(ii) $P(X \leq 20) = P\left[\frac{X-\mu}{\sigma} \leq \frac{20-\mu}{\sigma}\right]$

$= P\left[Z \leq \frac{20-12}{4}\right]$

$= P[Z \leq 2]$ $\frac{1}{6}$

$= P[-\infty \leq Z \leq 2]$

$= P[Z=2] - P[Z=-\infty]$

$= P[Z=2]$

$= 0.97725$ $\to \frac{1}{2}$

(iii) $P(0 \leq X \leq 20) = P\left[\frac{0-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{20-\mu}{\sigma}\right]$

$= P\left[\frac{0-12}{4} \leq Z \leq \frac{20-12}{4}\right]$

$= P[-3 \leq Z \leq 2]$

$= P(Z=2) - P(Z=-3)$

$= 0.97725 - 0.00135$

$= 0.9759$

2 PAGE

[Q.] The average monthly sales of 2000 firms are normally distributed with mean Rs. 26000 and S.D of Rs. 6000. Find

(a) The no. of firms for which the sales exceed Rs. 32,000

(b) The no. of firms with sales b/w Rs. 28,000 & Rs. 32,000

Soln:- X = monthly sales

$E(X) = \mu = 26000$ $\quad\quad \sigma^2 = (6000)^2$

$\sigma = 6000$ $\quad\quad$ S.D $= \sigma = 6000$

$X \sim \text{Normal}(26000, 6000^2)$

(a) P[The firm sales exceed Rs. 32,000] $= P[X \geq 32{,}000]$

$$= P\left[\frac{X-\mu}{\sigma} \geq \frac{32000 - 26000}{6000}\right]$$

$$= P[Z \geq 1]$$

$$= 1 - P[Z \leq 1] \Rightarrow 1 - P[Z=1]$$

$$= 1 - 0.84134$$

$$= 0.15866$$

No. of firms $= 2000 \times 0.15866$

$= 317.32$

$\simeq 317$

(b) $P[28000 \leq X \leq 32000] = P$[[redacted] $\frac{28000-26000}{6000} \leq \frac{X-\mu}{\sigma} \leq \frac{32000-26000}{6000}$]

$$= P[0.33 \leq Z \leq 1]$$

$$= P[Z=1] - P[Z=0.33]$$

$$= 0.84134 - 0.62930$$

$$= 0.21204$$

No. of firms $= 2000 \times 0.21204$

$= 424.08$

DATE: / /
PAGE:

[Q.] The lifetimes of certain kinds of electronic devices have a mean of 300 hours and standard deviation of 25 hours. Assuming that the distribution of these lifetimes, which are measured to the nearest hour, can be approximated closely with a normal curve

(a) Find the probability that any one of these electronic devices will have a lifetime of more than 350 hours.?

(b) what percentage will have lifetimes of 300 hours or less?

(c) what percentage will have lifetime from 220 or 260 hours?

Solⁿ:- (a) $P(X \geq 350)$ | $X \sim ND(\mu, \sigma^2)$

(b) $P(X \leq 300)$ | $E(X) = \mu = 300$ hrs

(c) $P(220 \leq X \leq 260)$ | $SD(X) = \sigma = 25$ hrs

(a)

$$P(X \geq 350) = P\left[\frac{X-\mu}{\sigma} \geq \frac{350-\mu}{\sigma}\right]$$

$$= P\left[Z \geq \frac{350-300}{25}\right]$$

$$= P[Z \geq 2]$$

$$= P[2 \leq Z \leq \infty]$$

$$= P[Z=\infty] - P[Z=2]$$

$$= 1 - P[Z=2]$$

$$= 1 - 0.97725 \Rightarrow 0.02275$$

(ii)

$$P(X \leq 300) = P\left[\frac{X-\mu}{\sigma} \leq \frac{300-\mu}{\sigma}\right]$$

$$= P[Z \leq 0]$$

$$= P[-\infty \leq Z \leq 0]$$

$$= P[Z=-\infty] - P[Z=0]$$

$$= 1 - 0.5050 \Rightarrow 0.50 \Rightarrow 0.50 \times 100 \Rightarrow 50\%$$

(iii) $P(220 \le X \le 260) = P\left[\frac{220-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{260-\mu}{\sigma}\right]$

$= P\left[\frac{220-300}{25} \le z \le \frac{260-300}{25}\right]$

$= P[-3.20 \le z \le -1.60]$

= upper limit − lower limit

$= P[z = -1.60] - P[z = -3.20]$

$= 0.05480 - 0.00069$

$= 0.05411$

$= 5.411\%$

[Q.] An aptitude test for selecting officers in a bank was conducted on 1000 candidates. The average score is 42 and standard deviation score is 24. Assuming normal distribution for the score, find :-

(a) The no. of candidates whose scores exceed 58

(b) the no. of candidates whose scores lie b/w 30 & 66.

Sol^n:- $X \sim ND(\mu, \sigma^2)$

$E(X) = \mu = 42$

$SD(X) = \sigma = 24$

(a) $P(X \ge 58)$

(b) $P(30 \le X \le 66)$

DATE: / /
PAGE:

Sol^n :- (a) $P(X \geq 58) = P\left[\frac{X-\mu}{\sigma} \geq \frac{58-42}{24}\right]$

$= P\left[z \geq \frac{16}{24}\right]$

$= P[z \geq 0.666]$

$= P[0.666 \leq z \leq \infty]$

$= P[z = \infty] - P[z = 0.666]$

$= 1 - 0.7454$

$= 0.2546$

$\therefore$ Aptitude test conducted for officer in a bank on 1000 candidates $= 0.2546 \times 1000$

$= 254.6$

≈ 254

(b) $P(30 \leq X \leq 66) = P\left[\frac{30-42}{24} \leq \frac{X-\mu}{\sigma} \leq \frac{66-42}{24}\right]$

$= P[-0.5 \leq z \leq 1]$

$= P[z=1] - P[z=-0.5]$

$= 0.8413 - 0.3085$

$= 0.5328$

$\therefore$ Aptitude test conducted for officer in bank on 1000 candidates $=$

0.5328×1000

$= 532.8$

$\simeq 533$

DATE: / /
PAGE:

Uniform distribution [continuous case] :-

(1) A random Variable x is said to be uniformly distributed over the interval $[\alpha, \beta]$ if its p.d.f is given by,

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \le x \le \beta \\ 0 & \text{, otherwise} \end{cases}$$

(2) The uniform distribution arises in practice when we suppose a certain random variable is equally likely to near any value in interval $[\alpha, \beta]$

(3) if x lies b/w interval $[\alpha, \beta]$ & probability of x lies b/w the sub interval $[a, b]$ is $P[a \le x \le b] = \frac{b-a}{\beta - \alpha}$

Simple linear Regression Model

Regression :- Regration is statical tool for investigating the relationship b/w variables. It is frequently used to predict the future & understand which factors cause an outcome.

Regression analysis :- Regression analysis is mathematical measure of average relationship b/w 2 a more Variable in term of original units of data.

Types of variables in regression analysis. there are 2 types of Variables

(1) Dependent Variable
(2) Independent Variable.

(1) Dependent Variable :- Data that cannot be controlled directly.
(2) Independent Variable :- Data that can be controlled directly.

★ Linear regression can be used in market research studies & customer survey analysis.

★ Least Square regression is used to model causal relationships b/w parameters in biological systems.

Model :- A model is adequate summary of Variables comprising the statical assumption that help us to express dependent variables as a function of Independent Variables.

Parameter :- Parameter are measurable quantities of model for estimating the output.

Concept :- In linear regression models, linear implies arrange or extending the straight line. linear suggest that the relationship b/w dependent & Independent variable can be expressed in straight line

$$Y = mx + c$$

(i) Y is dependent variable; ie the variable needs to be estimated & Predicted.

(ii) x is the independent variable basically it is input.

(iii) m is the slope. It determine the angle of line. It is parameter denoted as β.

(iv) c is the intercept, A constant that determines the value of Y when x is 0.

Regression model :- The regression model is given by, $Y = a + bX$

where, Y is response variable

X is Independent Variable

a is the intercept

b is the regression coefficient of X.

DATE: ___ PAGE: ___

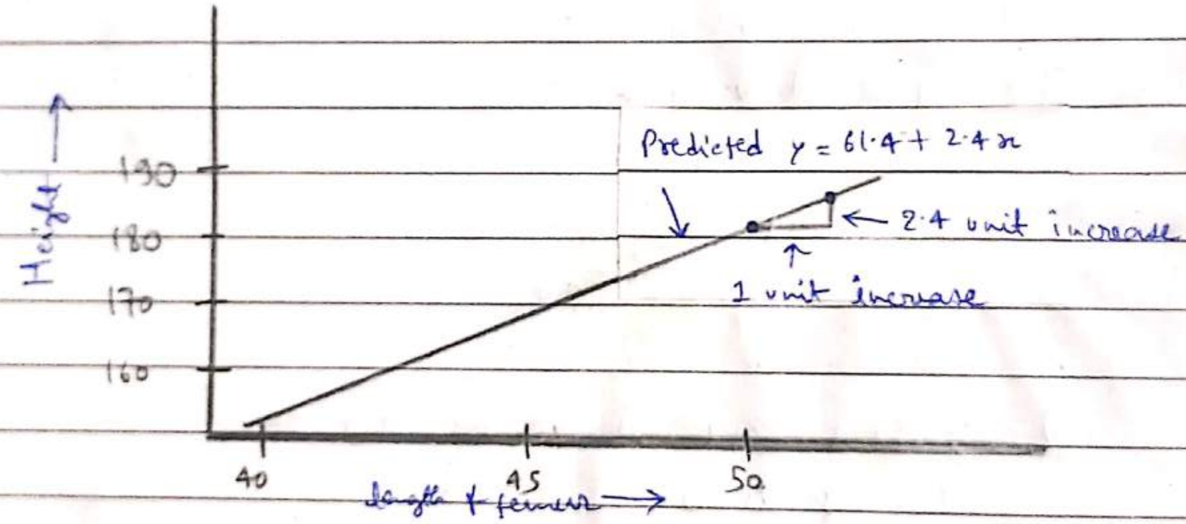
ex :- Height Based on Human Remains

- Regression Equation :- It is predicted height and is the length of a femur (thighbone), measured in centimeters.

use the regression equation to predict the height of person whose ~~femur length was~~ 50 centimeters.

$$\hat{y} = 61.4 + 2.4(50) = 181.4$$ here $\hat{a} = 61.4$, $\hat{b} = 2.4$

A 1 cm increase in femur length results in 2.4 cm increase in predicted height

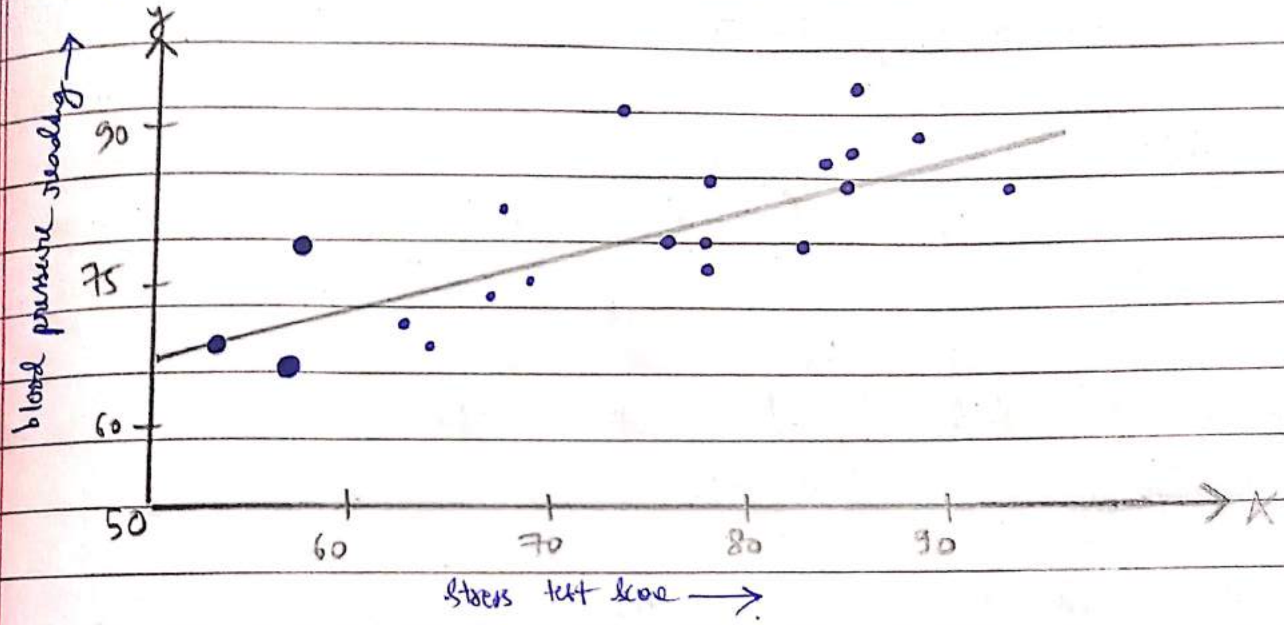


Three regression lines are possible, showing positive association (slope > 0), negative association (slope < 0) & no association (slope = 0).

[Q.] Suppose a university medical centre is investing the relationship b/w stress and blood pressure. Assume that both a stress test score and a blood pressure reading have been recorded for a sample of 20 patient. The data are shown graphically in the figure below, called a scatter diagram, Values of the independent Variable, stress test score, are given on the horizontal axis, and values of the dependent Variable, blood pressure, are shown on the vertical axis.

Sol^n:- The line passing through the data points is the graph of the estimated regression equation : $\hat{y} = 42.3 + 0.49x$

The parameter estimates, $\hat{a} = 42.3$ & $\hat{b} = 0.49$



Residuals Measure the size of Prediction Errors :- Residuals measure the size of prediction errors, the vertical distance b/w point & regression line. Each observation has a residual

⇒ calculation for each residual :- $y - \hat{y}$

· larger residual indicates an unusual observation.

Method of analysis [ordinary least Square method (OLS)] :-

The least square method is statical procedure to find the best fit for set of data point by minimizing the sum of residual of points from the plotted curve. Each point of data represents the relationship b/w known independent variable & unknown dependent variable. least square regression is used to predict the behaviour of dependent variables.

It mainly used in weather data analysis

DATE: / /

\# The estimates $\hat{a}$ & $\hat{b}$ are determined in such as way that:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \text{ is least}$$

$$\sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i)^2 = S \text{ is least}$$

\# Estimate slope b is,

$$\hat{b} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

\# Estimate of intercept a is given by,

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

\# Coefficient of determination :- The coefficient of determination is given by the square of correlation coefficient r i.e

$\therefore$ coefficient of determination = r^2

when r is not known, then coefficient of determination is given by,

$$r^2 = \frac{\text{explained Variation}}{\text{total Variation}}$$

$$r^2 = \frac{\Sigma(\hat{y} - \bar{y})^2}{\Sigma(y - \bar{y})^2}$$

$\Rightarrow$ coefficient of determination is denoted as R^2

\# Interpretation of r^2 :-

(i) r^2 always lies b/w 0 & 1

(ii) r^2 close to 1 indicates strong correlation b/w x & y.

(iii) r^2 can be used to measure how well indipendent variable explains the dependent Variable.

eg :- 0.80 implies 80% of variation in dependent Variable . y has been explained by independent Variable x & fitted model can be said as good fit

$b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$, $b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$

$\Rightarrow$ Regression eqn Y on X is:

$$Y - \bar{Y} = b_{yx}(X - \bar{X})$$

$\Rightarrow$ Regression eqn X on Y is:

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

[Q.1] Marks obtained by 10 students in Economic & Statistics are given below :-

Marks in Eco = 25, 28, 35, 32, 31, 36, 29, 38, 34, 32

Marks in Stat = 43, 46, 49, 41, 36, 32, 31, 30, 33, 39

Find

(i) The regression eqn of Y on X

(ii) estimate the marks in statistics when the marks in Economics is 30.

Solⁿ:- let the marks in Economics be denoted by X & Statistics by Y.

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
320	380	0	0	140	398	-93

$\bar{X} = \frac{\Sigma X}{n} = \frac{320}{10} \Rightarrow 32$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{380}{10} \Rightarrow 38$

$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{-93}{140}$

$= -0.664$

DATE: ___ / ___ / ___
PAGE: ___

(i) Regression eqⁿ Y on X is,

$$Y - \bar{Y} = b_{yx} (X - \bar{X})$$
$$Y - 38 = -0.664 (X - 32)$$
$$Y - 38 = -0.664X + 21.248$$
$$Y = -0.664X + 21.248 + 38$$
$$Y = -0.664X + 59.248$$
$$Y = 59.248 - 0.664X$$

(ii) To estimate the marks in Statistics (Y) for given marks Economics (X). put $X = 30$, in above eqⁿ we get,

$$Y = 59.248 - 0.664(30)$$
$$= 59.248 - 19.92$$
$$= 39.33$$

cm

[Q.2] The height in [cm] and weight (in kg) of 10 basketball players on a team are :-

Height (X) = 186, 189, 190, 192, 193, 193, 198, 201, 203, 205

weight (Y) = 85, 85, 86, 90, 87, 91, 93, 103, 100, 101

calculate :-

1. The regression line of Y on X
2. The coefficient of correlation & coefficient of determination
3. The estimated weight of player who measures 208

Soln :-

DATE: / /
PAGE:

X	Y	$x = X-\bar{X}$	$y = Y-\bar{Y}$	x^2	y^2	xy
186	85	-9	-7.1	81	50.41	63.9
189	85	-6	-7.1	36	50.41	42.6
190	86	-5	-6.1	25	37.21	30.5
192	90	-3	-2.1	9	4.41	6.3
193	87	-2	-5.1	4	26.01	10.2
193	91	-2	-1.1	4	1.21	2.2
198	93	3	0.9	9	0.81	2.7
201	103	6	10.9	36	118.81	65.4
203	100	8	7.9	64	62.41	63.2
205	101	10	8.9	100	79.21	89
1950	921	0	0	368	430.9	376

$\bar{X} = \frac{\Sigma X}{n} = \frac{1950}{10} \Rightarrow 195$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{921}{10} \Rightarrow 92.1$

$b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$

$= \frac{376}{368} \Rightarrow 1.0217$

(i) $Y-\bar{Y} = b_{yx}(X-\bar{X})$

$Y - 92.1 = 1.0217\,(X-195)$

$Y - 92.1 = 1.0217x - 199.2315$

$Y = 1.0217x - 199.2315 + 92.1$

$Y = 1.0217x - 107.139$

$Y = -107.139 + 1.0217x$

$\therefore \hat{a} = -107.139$

$\hat{b} = 1.0217$

(ii) $r = \frac{\Sigma xy}{\sqrt{\Sigma(x^2) \times \Sigma(y^2)}}$

$= \frac{376}{\sqrt{360 \times 430.9}}$

$= \frac{376}{393.8578}$

$= 0.95465$

and

$r^2 = r \times r$

$= 0.95465 \times 0.95465$

$\Rightarrow 0.91135$

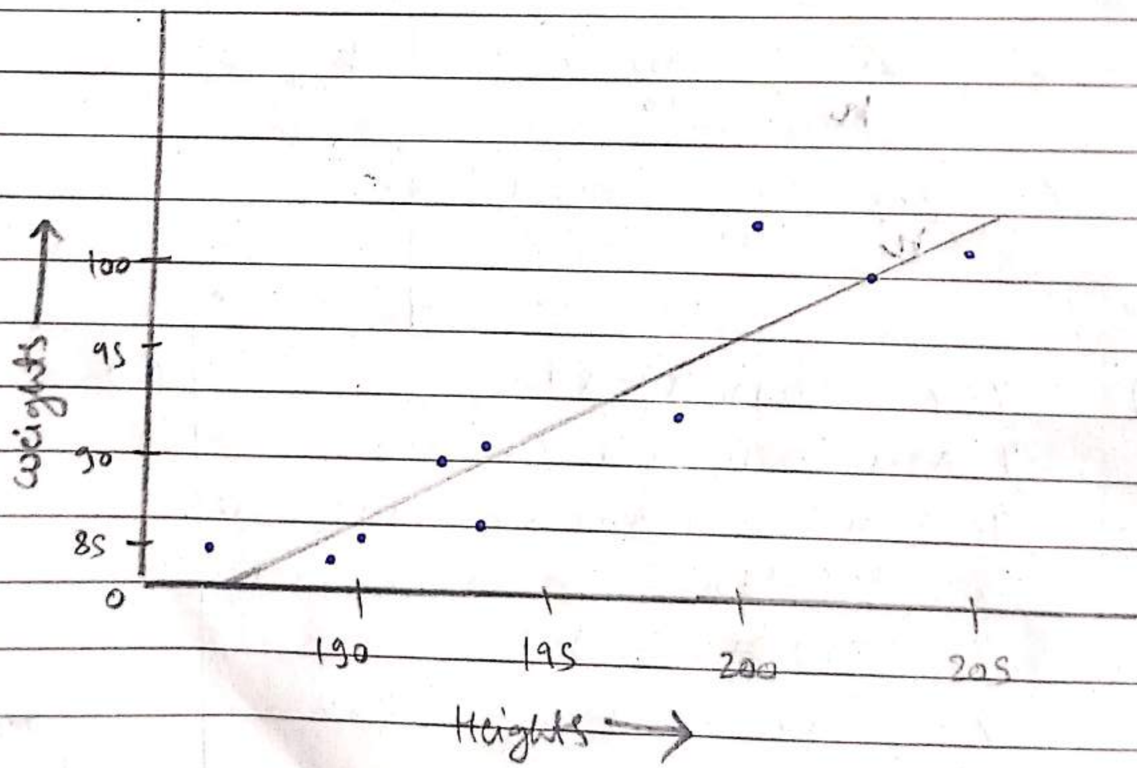
(iii) The estimated weight of a player who measures 208 cm.

$$Y = -107.139 + 1.0217(208)$$

$$Y = -107.139 + 212.5136$$

$$Y = 105.3746$$

Regression Plot :-



[6.C] An investment company speculates about the relationship b/w family income (X) & their allocation for investment (Y). A survey of 8 randomly selected families gives the following data.

Annual Income (X) (in 000 Rs.)	18	21	19	34	23	30	36	39
Percent allocation for investment (Y)	10	15	11	20	15	40	45	55

(i) develop a simple regression model by considering percent allocation as dependent variable & annual income as independent variable.

(2) Estimate the percentage of allocation for Investment when the annual Income in Rs. 27,500.

DATE: / /
PAGE:

Solⁿ:-

X	Y	$x = X-\bar{X}$	$y = Y-\bar{Y}$	x^2	y^2	xy
18	10	−9.5	−16.375	90.25	268.140	155.5625
21	15	−6.5	−11.375	42.25	129.390	73.9375
19	11	−8.5	−15.375	72.25	236.390	130.687
34	20	6.5	−6.375	42.25	40.640	−41.4375
23	15	−4.5	−11.375	20.25	129.390	51.1875
30	40	2.5	13.625	6.25	185.640	34.0625
36	45	8.5	18.625	72.25	346.890	158.3125
39	55	11.5	28.625	132.25	819.390	329.1875
$\Sigma X = 220$	$\Sigma Y = 211$	$\Sigma x = 0$	$\Sigma y = 0$	$\Sigma x^2 = 478$	$\Sigma y^2 =$ 2155.87	$\Sigma xy =$ 891.499

(i) $\bar{X} = \frac{\Sigma X}{n} \Rightarrow \frac{220}{8} \Rightarrow 27.5$

$\bar{Y} = \frac{\Sigma Y}{n} \Rightarrow \frac{211}{8} \Rightarrow 26.375$

$b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$

$= \frac{891.499}{478}$

$= 1.8650$

$Y - \bar{Y} = b_{yx}(X - \bar{X})$

$Y - 26.375 = 1.8650(X - 27.5)$

$Y - 26.375 = 1.8650x - 51.2875$

$Y = 1.8650x - 51.2875 + 26.375$

$\boxed{Y = 1.8650x - 24.9125}$

(ii) Estimate the % of allocation for Investment when the annual income in Rs. 27,500. put (27.5) in above eqⁿ we get;

$Y = 1.8650(27.5) - 24.9125$

$Y = 51.2875 - 24.9125$

$Y = 26.375$

[6.6] The data given below is from a chemical experiment to prepare a standard curve for the determination of formaldehyde by the addition of chromatropic acid and concentrated sulphuric acid and the reading of the resulting purple color on a spectrophotometer.

carbohydrate (ml) (X)	0.1	0.3	0.5	0.6	0.7	0.9
optical density (Y)	0.08	0.27	0.45	0.54	0.63	0.78

(i) Fit a linear model for optical density modelled by carbohydrate.

(ii) Estimate the optical density when the carbohydrate is 1.2 ml.

(iii) Make a scatterplot with a regression line.

Soln :-

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
0.1	0.08	−0.4166	−0.3783	0.1735	0.1431	0.1575
0.3	0.27	−0.2166	−0.1883	0.0469	0.0354	0.0407
0.5	0.45	−0.0166	−0.0083 [redacted]	0.0002755	0.0000688	0.000137
0.6	0.54	0.0834	0.0817	0.00695	0.00667	0.00681
0.7	0.63	0.1834	0.1717	0.0336	0.0294	0.0314
0.9	0.78	0.3834	0.3217	0.1469	0.1034	0.1233
$\Sigma X = 3.1$	$\Sigma Y = 2.75$	$\Sigma x = -0.1664$ $4 \times 10^{-4} \Rightarrow 0.0004$	$\Sigma y = 0.0002$	$\Sigma x^2 = 0.4081$	$\Sigma y^2 = 0.3180$	$\Sigma xy = 0.3598$

$$\bar{X} = \frac{\Sigma x}{n} = \frac{3.1}{6} \Rightarrow 0.5166$$

$$\bar{Y} = \frac{\Sigma y}{n} = \frac{2.75}{6} \Rightarrow 0.4583$$

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{0.3598}{0.4081}$$

$$= 0.8816$$

(i) $Y - \bar{Y} = b_{yx} (X - \bar{X})$

$Y - 0.4583 = 0.8816 (X - 0.5166)$

$Y - 0.4583 = 0.8816X - 0.4554$

$Y = 0.8816X - 0.4554 + 0.4583$

$Y = 0.8816X - 0.0029$

$\therefore \hat{b} = 0.8816$

$\hat{a} = 0.0029$

(ii) Estimate the optical density when the carbohydrate is 1·2 ml

$Y = 0.8816X - 0.0029 \quad \therefore [\text{put } X = 1.2]$

$Y = 0.8816(1.2) - 0.0029$

$Y = 1.05792 - 0.0029$

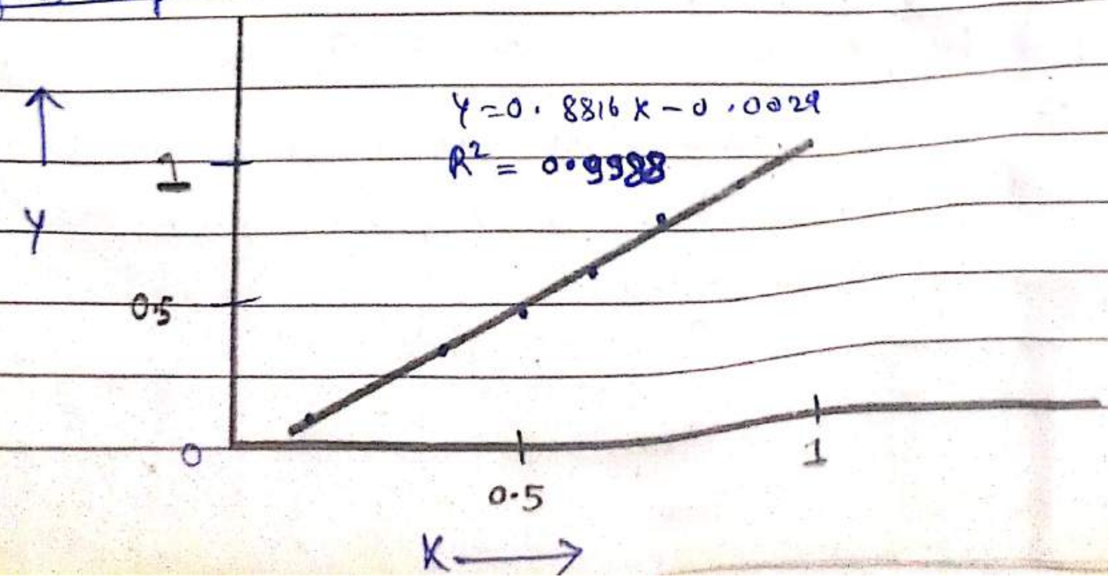
$Y = 1.05502$

& $r^2 = \dfrac{\Sigma xy}{\sqrt{\Sigma(x^2) \times \Sigma(y^2)}}$

$= \dfrac{0.3598}{\sqrt{0.4081 \times 0.3180}}$

$= \dfrac{0.3598}{\sqrt{0.1297}} \Rightarrow \dfrac{0.3598}{0.3602} \Rightarrow 0.9988$

(iii) # Regression plot :-



Simple linear regression

DATE: / /
PAGE:

⇒ If X & Y are any two random variables assuming n-pairs of values and there is a linear relationship b/w X & Y. we can represent simple regression model as :- $\boxed{Y = a + bx + e}$

where, Y = dependent Variable
X = Independent Variable
and a & b are regression coefficient
and e is the error term

⇒ Fitting regression line means estimating the unknown values of a & b. The estimated value of a & b is denoted as $\hat{a}$ & $\hat{b}$

∴ fitted regression line is

$$Y = \hat{a} + \hat{b}x$$

$$\boxed{\hat{b} = \frac{\text{cov}(x,y)}{\text{Var}(x)} = \frac{\frac{\Sigma xy}{n} - \bar{x}\bar{y}}{\frac{\Sigma x^2}{n} - (\bar{x})^2}}$$

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

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

$\boxed{\hat{a} = \bar{Y} - \hat{b}\bar{x}}$

[Q] The following data represent the income & expenditure level of 10 employee. [Data in thousands]

income	15	20	24	35	42	30	42	52	48	50
expenditure	10	15	15	23	35	28	32	40	29	41

Fit a Simple regression model b/w income & expenditure. Also estimate the value of expenditure of an employee when the income is Rs.20

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
15	10	-20.8	-16.6	432.64	275.56	345.28
20	15	-15.8	-11.8	249.64	139.24	186.44
24	15	-11.8	-11.8	139.24	139.24	139.24
35	23	-0.8	-3.8	0.64	14.44	3.04
42	35	6.2	8.2	38.44	67.24	50.84
30	28	-5.8	1.2	33.64	1.44	-6.96
42	32	6.2	5.2	38.44	27.04	32.24
52	40	16.2	13.2	262.44	174.24	213.84
48	29	12.2	2.2	148.84	4.84	26.84
50	41	14.2	14.2	201.64	201.64	201.64
$\Sigma X = 358$	$\Sigma Y = 268$	$\Sigma x = 0$	$\Sigma y = 0.2$	$\Sigma x^2 = 1545.6$	$\Sigma y^2 = 1044.92$	$\Sigma xy = 1192.44$

$\bar{X} = \frac{\Sigma X}{n} = \frac{358}{10} \Rightarrow 35.8$

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{268}{10} \Rightarrow 26.8$

$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{1192.44}{1545.6}$

$= 0.77150$

(i) $Y - \bar{Y} = b_{yx}(X - \bar{X})$

$Y - 26.8 = 0.77150\,(X - 35.8)$

$Y - 26.8 = 0.77150X - 27.6197$

$Y = 0.77150X - 27.6197 + 26.8$

$Y = 0.77150X - 0.8197$

$\hat{b} = 0.77150$

$\hat{a} = -0.8197$

(ii) estimate the value of expenditure of an employee when the income is Rs.20

[put X=20]

$Y = 0.77150(20) - 0.8197$

$= 14.6103$

≈ 15

$\therefore$ therefore the estimated expenditure of employee when income is Rs. 20,000 is Rs. 15,000

→ This [Q] written twice times

[Q.1] A computer company received a rush order for as many home computers as could be shipped during a six week period. Company records provide the following daily shipments :-

22, 65, 66, 47, 56, 56, 61, 74, 73, 36, 62, 54, 48, 61, 74, 46, 65, 45,
57, 68, 71, 82, 35, 41, 44, 28, 58, 61, 45, 75, 55, 75, 34, 54, 45, 80,
68, 58, 56, 85, 35 & 53.

(i) Construct the frequency distribution with the class interval , 20 – 30 , 30 – 40 etc.

(ii) Determine the relative frequency.

(iii) Construct the histogram & comment on it

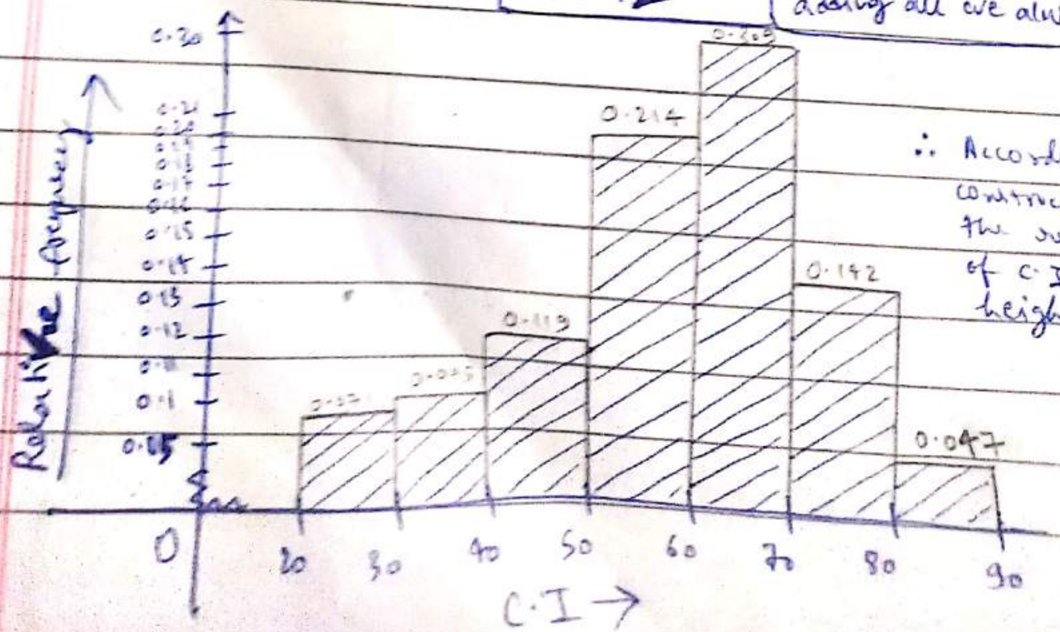
Sol^n :- (i) & (ii)

20 – 30 → until 20 to 29

30 – 40 → 30 to 39

C-I	Tally mark	frequency	Relative frequency
20 – 30	III	3 ✓	$\frac{3}{42} = 0.071$
30 – 40	IIII	4 ✓	$\frac{4}{42} = 0.095$
40 – 50	~~IIII~~	5 ✓	$\frac{5}{42} = 0.119$
50 – 60	~~IIII~~ IIII	9 ✓	$\frac{9}{42} = 0.214$
60 – 70	~~IIII~~ ~~IIII~~ III	13 ✓	$\frac{13}{42} = 0.309$
70 – 80	~~IIII~~ I	6 ✓	$\frac{6}{42} = 0.142$
80 – 90	II	2	$\frac{2}{42} = 0.047$
		$\Sigma f = 42$	adding all we always get 1

(iii)



∴ According to constructed histogram the relative frequency of C.I 60 – 70 is heigher.

DATE: __/__/__
PAGE: ______

[Q.27] The processing time of 10 personal computer (in seconds) are :- 56, 45, 32, 50, 50, 80, 66, 48, 35 & 75. determine the average & S.D of processing time

Soln :- (i) Avg of processing time = $\frac{56+45+32+50+50+80+66+48+35+75}{10}$

$= \frac{536}{10} \Rightarrow 53.6$

(ii) S.D of processing time

$\bar{x}$, Mean = 53.6

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
32	53.6	−21.6	466.56
35	53.6	−18.6	345.96
45	53.6	−8.6	73.96
48	53.6	−5.6	31.36
50	53.6	−3.6	12.96
50	53.6	−3.6	12.96
56	53.6	2.4	5.76
66	53.6	12.4	153.76
75	52.6	21.4	457.96
80	53.6	26.4	696.96
			$\Sigma(x-\bar{x})^2 = 2258.2$

$n = 10$

$$S.D = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{2258.2}{10}} = \sqrt{225.82}$$

$$= 15.0273$$

[Q.3] The following data represents the marks of the students in class. Determine the avg marks of students.

Marks	10	15	20	25	30	35	40	45
No. of students	2	3	5	8	6	4	3	1

Solⁿ :-

Marks (x)	No. of students (f)	fx
10	2	$10 \times 2 = 20$
15	3	$15 \times 3 = 45$
20	5	$20 \times 5 = 100$
25	8	$25 \times 8 = 200$
30	6.	$30 \times 6 = 180$
35	4.	$35 \times 4 = 140$
40	3	$40 \times 3 = 120$
45	1	$45 \times 1 = 45$
	$\Sigma f = 32$	$\Sigma fx = 850$

$\bar{x} = \frac{850}{32}$

$= 26.5625$

∴ Avg marks of students is 26.5625

already done

[Q.4] The following distribution gives the pattern of overtime work....

[Q.5] Two cricket players Sam & Ram has following scores in their five innings

Sam	50	32	60	40	50
Ram	65	100	20	30	75

(i) Determine the avg scores

(ii) Determine the S.D of scores

(iii) determine the coefficient of variation & hence identify the most consistent player.

DATE: / /
PAGE:

Sol^n :- (i) For Sam, avg scores $= \frac{50+32+60+40+50}{5} \Rightarrow \frac{232}{5} \Rightarrow 46.4$

For Ram, avg scores $= \frac{65+100+20+30+75}{5} \Rightarrow \frac{290}{5}$

$\Rightarrow 58$

(ii) determine the S.D of scores

$\Rightarrow$ $\boxed{\text{for Sam}}$ for Sam, $\bar{x} = 46.4$

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
50	46.4	3.6	12.96
32	46.4	-14.4	207.36
60	46.4	13.6	184.96
40	46.4	-6.4	40.96
50	46.4	3.6	12.96
			$\Sigma(x-\bar{x})^2 =$ 459.2

$\therefore n = 5$

$\therefore S.D = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{\frac{459.2}{5}} \Rightarrow \sqrt{91.84}$

$\Rightarrow 9.5833$

$\Rightarrow$ $\boxed{\text{for Ram}}$ for Ram, $\bar{x} = 58$

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
65	58	7	49
100	58	42	1764
20	58	-38	1444
30	58	-28	784
75	58	17	289
			$\Sigma(x-\bar{x})^2 =$ 4330

$\therefore n = 5$

$\therefore S.D = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{4330}{5}}$

$= \sqrt{866}$

$\Rightarrow 29.4278$

(iii) For Sam:

$\therefore C.V = \frac{SD}{\bar{x}} \times 100$

$= \frac{9.5833}{46.4} \times 100$

$= 20.65$

For Ram:

$\therefore C.V = \frac{SD}{\bar{x}} \times 100$

$= \frac{29.4278}{58} \times 100$

$= 50.7375$

$\Rightarrow$ The coefficient of variation of Sam is less than Ram. Therefore Sam is more consistent player

Statistics

DATE : / /
PAGE :

[Q.] The winning amounts of two players in game of 6 rounds is given below :-

Player A	50	90	80	90	70	40
Player B	30	54	48	54	42	24

(i) Determine the coefficient of variation

(ii) which two player has consistent winning.

Sol^n :- (i)

For A, $\bar{x}$ Mean $= \frac{50+90+80+90+70+40}{6} \Rightarrow \frac{420}{6} \Rightarrow 70$

For B, $\bar{x}$ Mean $= \frac{30+54+48+54+42+24}{6} \Rightarrow \frac{252}{6} \Rightarrow 42$

$\Rightarrow$ For A

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
50	70	−20	400
90	70	20	400
80	70	10	100
90	70	20	400
70	70	0	0
40	70	−30	900
			$\Sigma(x-\bar{x})^2 = 2200$

$\Rightarrow$ For B

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
30	42	−12	144
54	42	12	144
48	42	6	36
54	42	12	144
42	42	0	0
24	42	−18	324
			$\Sigma(x-\bar{x})^2 = 792$

$\therefore SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{2200}{6}} \Rightarrow 19.1485$

$C.V = \frac{SD}{\bar{x}} \times 100 \Rightarrow \frac{19.1485}{70} \times 100$

$\Rightarrow 27.355$

$\therefore SD = \frac{\Sigma(x-\bar{x})^2}{n} = \sqrt{\frac{792}{6}}$

$\Rightarrow 11.4891$

$C.V = \frac{SD}{\bar{x}} \times 100$

$= \frac{11.4891}{42} \times 100$

$= 27.355$

(ii) Both the player consistent winning

Statistics

DATE: / /
PAGE:

[Q.] A survey was conducted to determine the age (in years) of 120 automobiles. The result of such a survey is as follows.

Age of Auto	0-4	4-8	8-12	12-16	16-20
Number of Autos	13	29	48	22	8

What is the median age of the autos?

Sol^n:-

Age of Auto	Number of Autos (f)	cumulative frequency
0-4	13	13
4-8	29	42
8-12	48	90
12-16	22	112
16-20	8	120

$N = 120$ $\quad \frac{N}{2} = \frac{120}{2} = 60$

Median class is 8-12

$L = 8$, $m = 42$, $f = 48$

$c = 4$

$$\text{median} = L + \left[\frac{\frac{N}{2} - m}{f}\right] \times c$$

$$= 8 + \left[\frac{60 - 42}{48}\right] \times 4$$

$$= 8 + 0.375 \times 4$$

$$= 9.5$$

[Q.] A bowler's scores for 8 games were 55, 60, 74, 45, 80, 90, 22 & 35 using these data as a sample, compute the following descriptive statistics.

1. Avg score
2. median score
3. Standard deviation of score

Sol^n:-

① Avg score $= \frac{55+60+74+45+80+90+22+35}{8}$

$= 57.625$

② Ascending order:- 22, 35, 45, 55, 60, 74, 80, 90

$n = 8$ i.e even,

then $\left(\frac{n}{2}\right) = \left(\frac{8}{2}\right) \Rightarrow 4$

$\left(\frac{n+2}{2}\right) = \left(\frac{10}{2}\right) \Rightarrow 5$

Median $= \frac{55+60}{2}$

$= 57.5$

Scanned with OKEN Scanner

(3) $\bar{x}$, Mean = $\frac{55+60+74+45+80+90+22+35}{8} \Rightarrow 57.625$

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
22	57.625	-35.625	1269.190
35	57.625	-22.625	511.890
45	57.625	-12.625	159.390
55	57.625	-2.625	6.890
60	57.625	2.375	5.640
74	57.625	16.375	268.190
80	57.625	22.375	500.640
90	57.625	32.375	1048.40
			$\Sigma(x-\bar{x})^2 = 3769.87$

$$S^2 = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{3769.87}{7}} \Rightarrow \sqrt{538.55}$$

this because n

$\boxed{S = 23.2068}$

[Q.] The no. of employee, wages per employee and the variance of the wages of employees for two factories are given below

No. of employees	Factory A	Factory B
Avg wage per employee/month	1250	1075
Variance of the wages (in Rupee)	86	115

using coefficient of variation, identify in which factory there is a greater variation in the distribution of wages per employee

∴ $C.V_A < C.V_B$. So factory B has greater variation in distribution of wages per employee

Sol^n:-

Factory A

$x = 1250$

$\sigma x^2 = 86$ $\boxed{\sigma x = \sqrt{86} = 9.273}$

$C.V_A = \frac{\sigma x}{\bar{x}} \times 100$

$= \frac{9.273}{1250} \times 100 \Rightarrow 0.7418$

Factory B

$\bar{x} = 1075$

$\sigma x^2 = 115$ $\boxed{\sigma x = \sqrt{115} = 10.7238}$

$C.V_A = \frac{\sigma x}{\bar{x}} \times 100$

$= \frac{10.7238}{1075} \times 100 \Rightarrow 0.9975$