Date / /
Page

Data :- collection of Row information

→ There are 4 types / classification
i) Qualitative - religion, gender
ii) Quantitative - Student above this height, weight
iii) Geographical - how many belongs to this area
iv) Temporal-time : how many stu taken admission upto June

Another way of Classification

One way Classification : classify the data based on only one variable

Two way classification: classify the data based on two variable
eg: boys with percentage upto 50% 40

Multi-way classification : classify the data based on three or more than three variable

Ungrouped data
Marks of Student out of 20
15, 20, 25, 26, 27, 7, 8, 7
This data is ungrouped or unorganised

7, 7, 8, 15, 20, 25, 26, 27 This is ungrouped data but Organised

Date......./......./.......
Page.......................

Q 7, 7, 9, 10, 11, 12, 15, 15, 16, 20

Ungrouped but organised

	f
7	11
9	1
10	1
11	1
12	1
15	11

Q 2, 1, 15, 12, 1, 8, 7, 21, 15, 1, 8, 8, 9, 8, 2, 15, 7, 3, 1, 20, 21, 9, 7, 9

1, 1, 1, 1, 2, 2, 3, 3, 7, 7, 7, 8, 8, 8, 9, 9, 9, 12, 15, 15, 15, 20, 21, 21

	f	
1	1111	4
2	11	2
3	11	2
7	111	3
8	111	3
9	111	3
12	1	1
15	111	3
20	1	1
21	11	2

Grouped data

	f	Tally	Cf
0 - 10	5	~~1111~~	5
10 - 20	6	~~1111~~ 1	11
20 - 30	9	1 X or ~~1111~~ 1111	20
30 - 40	3	111	23

$\Sigma f = 23$

Date / /
Page

Two ways to represent grouped data.
inclusive
Exclusive

inclusive	exclusive
↓	0-10
0-10	10-20
11-20	20-30
21-30	30-40
31-40	

Inclusive form

	f	
0-10	3	0-10.5
11-20	8	10.5-20.5
21-30	2	20.5-30.5

$10+\frac{1}{2}=\frac{20+1}{2}=\frac{21}{2}=10.5$

$0-\frac{1}{2}=\frac{2-1}{2}=\frac{1}{2}=0.5$

Lower limit : $m-\frac{1}{2}$

Upper limit : $m+\frac{1}{2}$

m = midvalue
i = Class length

class limit : A class is form within the two values these values are known as the class limits. the lower value is known as lower limit (L_1), upper value is known as upper limit (L_2)

- upper limit represent as L_2

Date......./......./.......
Page...........

Magnitude of class interval

height

The difference b/w the upper and lower limit of the class is called the magnitude and length of the class. it is denoted by (i)

Midvalue or class mark

Arithmatical average of two class limits that is lower limit and the upper limit is called the midvalue or class mark.

0 - 10 $\quad m = \frac{10+0}{2}, \frac{10+20}{2}$

10 - 20

class frequency

The Units of data belongs to anyone of the classes the total number of these units is known as the frequency.

10, 17, 15, 22, 11, 16, 19, 24, 29, 18, 25, 26, 32, 14, 17, 20, 23, 27, 30, 12, 15, 18, 24, 36, 18, 15, 21, 28, 33, 38, 34, 13, 10, 16, 20, 22, 29, 19, 23, 31

form the frequency table

i) by inclusive method (with 4 as class interval)

Date...../...../.....
Page.....

CI	f
10 - 13	5
14 - 17	8
18 - 21	8
22 - 25	7
26 - 29	5
30 - 33	4
34 - 37	2
38 - 41	1
	40

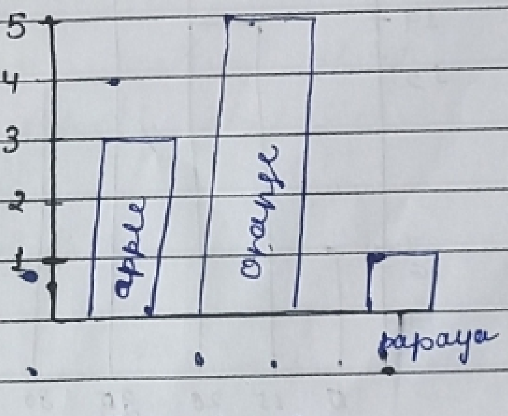
Bar Graph and Histogram

Bar Graph : Bar graph is a graph with rectangular bars each bar length is proportional to its represented value in other words the length or height of the bar is equal to the quantity within that category. The graph usually shows a comparision between different categories

Histogram: Histogram is another kind of graph that uses his bar in its display this type of graph is used in Quantitative data

→

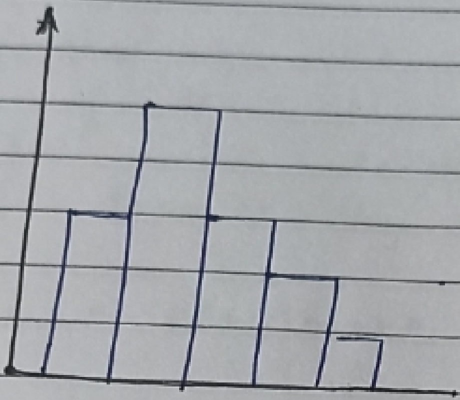
fruits	no
apple	3
orange	5
papaya	1



Bar graph Qualitative

Histogram

eg :



Quantitative dataset

Q Histogram

weekly wages	No. of workers
10 - 15	7
15 - 20	19
20 - 25	27
25 - 30	15
30 - 40	12
40 - 60	12
60 - 80	8

$\rightarrow \; 12 \times \frac{5}{10} = 6$

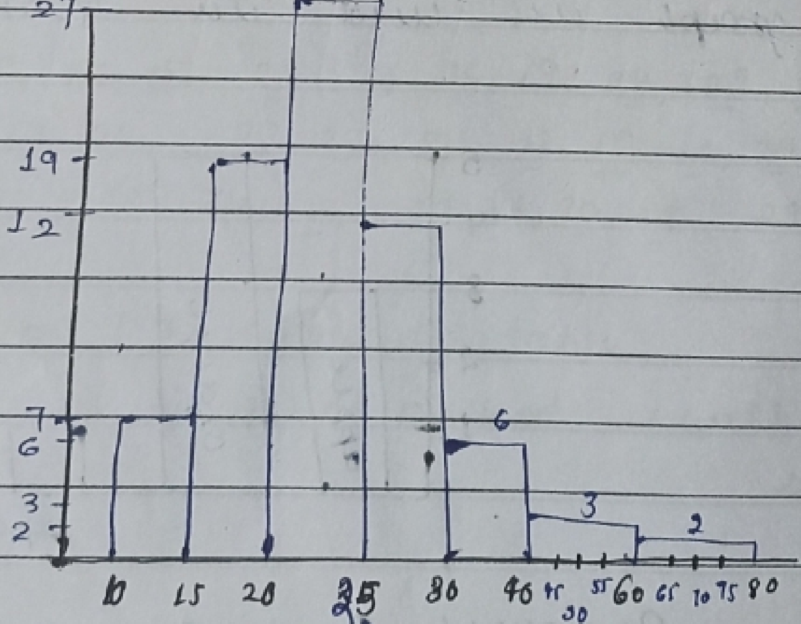
$12 \times \frac{5}{20} = 3$

$10 - 12$

$1 - \frac{12}{10}$

$5 - \frac{12}{?} \times 5 = 6$

$8 \times \frac{5}{20} = 2$

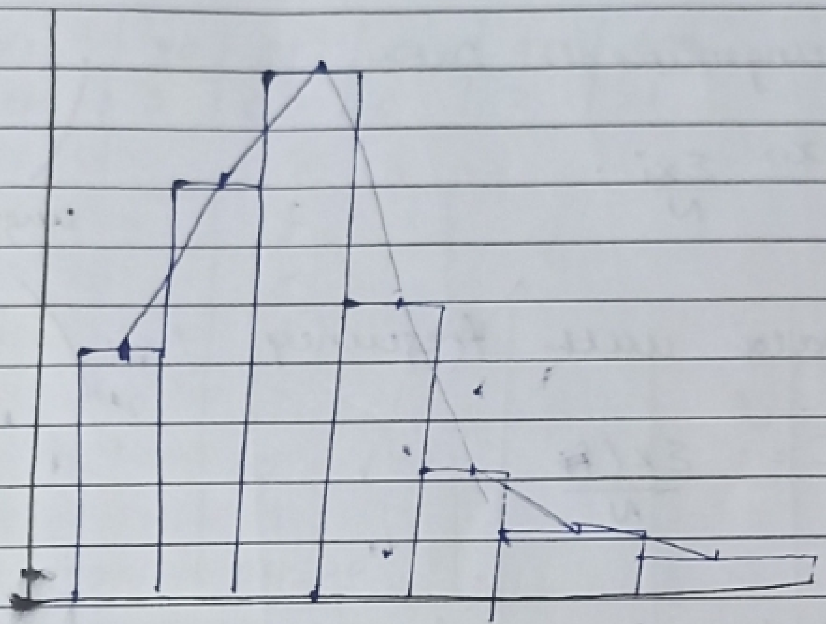


fx $\frac{\text{min size of CI}}{\text{lower CI}}$ (gap b/w CI)

Date / /

Page

frequency Polygon



* Question:

	weekly wages	no of workers	less than	more than
exclusive	10 - 15	7	7	100
	15 - 20	19	26	93
	20 - 25	2	53	74
	25 - 30	15	68	
	30 - 40	12	80	
	40 - 60	12	92	
	60 - 80	8	100	

~~Major~~ Measure of Central tendency.

Means to get the one value that is the representation of whole dataset

Three central tendency:

i) Mean

ii) Median

iii) Mode

Date......./......./.......
Page.......................

① Mean : represent average value of dataset

for ungrouped Data

$$x = \frac{\Sigma x_i}{N}$$

ungrouped

for Data with frequency

$$\bar{x} = \frac{\Sigma x_i f_i}{N}$$

Q Calculate mean of
9, 11, 13, 15, 17, 19

$N = 6$

$\Sigma x_i = 9 + 11 + 13 + 15 + 17 + 19 = 84$

$$\bar{x} = \frac{\Sigma x_i}{N} = \frac{84}{6} = 14$$

Q calculate mean ungrouped

x	10	11	13	15	16	19
f	4	5	8	6	4	3

$x_i f_i$	f
$10 \times 4 = 14$	4
$11 \times 5 = 55$	5
$13 \times 8 = 104$	8
$15 \times 6 = 90$	6
$16 \times 4 = 64$	4
$19 \times 3 = 57$	3
$\Sigma x_i f_i = 410$	$\Sigma f_i = 30$ or N

$$\bar{x} = \frac{\Sigma x_i f_i}{N}$$

$$= \frac{410}{30} = 13.6$$

Date / /
Page

Q Calculate mean for ungrouped data

marks	80	85	90	95	100
no of st	5	6	6	2	1

x_if_i	f
400	5
510	6
540	6
190	2
100	1
$\Sigma x_if_i = 1740$	$\Sigma f_i = 20$

$$\bar{x} = \frac{\Sigma x_if_i}{\Sigma f_i} = \frac{1740}{20} = 87 \text{ ans}$$

* Q for certain frequency table the mean is 146.

no of accidents	0	1	2	3	4	5	Total
freq (no of days)	46	? x	? y	25	10	5	200

Date / /
Page

Mean by Step deviation Method.

$$\bar{x} = A + \frac{\Sigma fd'}{N} \cdot h$$

where $d' = \frac{(m - A)}{h}$

CI	f	m	$d' = \frac{m-A}{h}$	fd'
300 - 350	5	325	-5	-25
350 - 400	14	375	-4	-56
400 - 450	23	425	-3	-69
450 - 500	20	475	-2	-40
500 - 550	50	525	-1	-50
550 - 600	52	575	0	0
600 - 650	25	625	1	25
650 - 700	22	675	2	44
700 - 750	7	725	3	21
750 - 800	2	775	4	8
	N = 220			$\Sigma fd' = -142$

$$\bar{x} = A + \frac{\Sigma fd'}{N} \cdot h$$

$$= 575 + \left(\frac{-142}{220}\right) \times 50$$

$$= 542.75$$

Median: is the middle value, which is arr in ascending or

15, 11, 2, 8, 20, 14

we need to organize this data eit in ascending or decending ord

Date / /
Page

2, 8, $\underbrace{11, 14}$, 15, 20
average

$\frac{11+14}{2} = 12.5$ Median

Q 3, 8, 5, 2, 7, 1, 6

1, 2, 3, 5, 6, 7, 8
Median = 5

Q find Median

(i) 23, 25, 29, 30, 39
29

(ii) 3, 4, 10, 12, 27, 32, 41, 49, 50, 55, 60, 63, 71, 75, 80
49 median

(iii) 29, 23, 25, 29, 30, 25, 28
23, 25, 25, 28, 29, 29, 30
28 median

Q Calculate Median (Ungrouped data)

x	f	Cf
5	3	3
10	6	9
$\boxed{15}$	10	19 → 16 lie in this
20	8	27
25	2	29
30	3	32

Hence 15 is the median

median → $\frac{N}{2} = \frac{32}{2} = 16$

Median for Continuous frequency distribu

formula:

$$\text{Median } (M) = L_1 + \frac{\frac{N}{2} - Cf}{f} \times h$$

where, → L_1 is the lower limit of class in which $\frac{N}{2}$ lies

→ Cf is the commulative frequency of prev class

→ f is the frequency of same class

→ h is the class interval.

Ques Calculate Median

age group	f	Cf
0 - 20	15	15
20 - 40	32	47
40 - 60	54	101 lies
60 - 80	30	131
80 - 100	19	150
		N = 150

$$\frac{N}{2} = \frac{150}{2} = 75$$

$$M = L_1 + \frac{\frac{N}{2} - Cf}{f} \times h$$

$$= 40 + \frac{75 - 47}{54} \times 20$$

$$= 40 + \frac{28 \times \cancel{20}^{10}}{\cancel{54}_{27}} = 40 + \frac{280}{27} = 50.37$$

Date / /
Page

Quartile. A data that is given to us then we have to equally divide into 4 parts.

1/4	1/2		
25%	25%	25%	25%

Q_1 Q_2 Q_3
↳ median

formula: $Q_i = LQ_i + \frac{(i)\frac{N}{4} - Cf}{f} \times h$

Q Calculate

	class interval	f	Cf
	2000 - 3000	2	2
Q_1	3000 - 4000	5	7 → lies
Q_2	4000 - 5000	6	13
Q_3	5000 - 6000	4	17
	6000 - 7000	3	20
		N = 20	

$\frac{N}{4} = \frac{20}{4} = 5$

→ $Q_1 = LQ_1 + \frac{\frac{N}{4} - C.f}{f} \times h$ → formula

$= 3000 + \left(\frac{5-2}{5}\right) \times 1000 = 3600$

→ $Q_2 = LQ_2 + \frac{\frac{N}{2} - Cf}{f} \times h$ → formula

$= 4000 + \left(\frac{10-7}{6}\right) \times 1000 = 4500$ median of data

→ $Q_3 = LQ_3 + \frac{3.\frac{N}{4} - Cf}{f} \times h$ — formula

$= 5000 + \frac{(15-13)}{4} \times 1000 =$

Pro Mean

calculate mean for:

CI	xi	fi	fixi
~~0-10~~	~~12~~	5	60
10-20	18	15	270
20-30	27	25	675
30-40	20	35	700
40-50	17	45	765
50-60	6	55	330

$\Sigma fi = 180$ $\Sigma fixi = 2730$

$$\bar{x} = \frac{\Sigma fixi}{\Sigma fi} = \frac{2730}{180} = 15.16$$

Calculate the mean of:

CI	fi	m	d = m - A
300-350	5	325	250
350-400	14	375	-200
400-450	23	425	-150
450-500	20	475	-100
500-550	50	525	-50
550-600	52	575	0
600-650	25	625	50
650-700	~~27~~ 22	675	100
700-750	7	725	150
750-800	2	775	200

class m = $\frac{350+300}{2}$

$= \frac{650}{2} = 325$

fidi
-1250
-2000
-3450
-2000
-25000
0
1250
2200
1030
400

$\Sigma fidi = 7100$

$$\bar{x} = A + \frac{\Sigma fidi}{\Sigma fi}$$

$$575 + \frac{-7100}{220}$$

$= 542.72$

Date / /
Page

Q1

Age	Population	Cf
0-15	60	60
15-30	70	130
30-45	45	175 - lies
45-60	52	227
60-75	71	298
75-90	42	340

find median

$$\frac{N}{2} = \frac{340}{2} = 170$$

$$M = L_1 + \frac{\frac{N}{2} - Cf}{f} \times h$$

$$= 30 + \frac{170 - 130}{\cancel{45}_{3}} \times \cancel{15}$$

$$= 30 + \frac{40}{3}$$

$$= \frac{90 + 40}{3} = \frac{130}{3} = 43.33$$

$$Q_1 = LQ_1 + \frac{\frac{N}{4} - Cf}{f} \times h \qquad \frac{N}{4} = \frac{340}{4} = 85$$

$$= 15 + \frac{85 - 60}{\cancel{70}_{14}} \times \cancel{15}^{3}$$

$$= 15 + \frac{25 \times 3}{14} = 15 + \frac{75}{14} = 20.35$$

$$Q_2 = LQ_2 + \frac{2 \times \frac{N}{4} - Cf}{f} \times h = 43.33 \rightarrow \text{median}$$

Date....../....../......

Page..........

$$Q_3 = L_{Q_3} + \frac{3 \cdot \frac{N}{4} - Cf}{f} \times h \qquad \frac{N}{3} = \frac{340}{3} = 113$$

$$= 60 + \left(\frac{255 - 227}{71} \right) \times 15$$

$$= 60 + \frac{28 \times 15}{71}$$

$$= 60 + 5.91$$

$$65.91$$

Decile

$$D_i = L_{D_i} + \left(\frac{i \frac{N}{10} - c \cdot f}{f} \right) \times h$$

Percentile

$$P_i = L_{P_i} + \left(\frac{i \frac{N}{100} - c \cdot f}{f} \right) \times h$$

Ques

	Class interval	f	cf
	2000 - 3000	2	2
	3000 - 4000	5	7
	4000 - 5000	6	13
	5000 - 6000	4	17
d_9	6000 - 7000	3	20
			N = 20

$$9 \times \frac{N}{10} = \frac{9 \times 20}{10}$$

$$P_{95} = 95 \times \frac{20}{100} = 19$$

Date / /
Page

→ $D_9 = ?$

$$D_9 = L_9 + \frac{\left(9 \times \frac{N}{10} - Cf\right)}{f} \times h$$

$$= 6000 + \left(\frac{18 - 17}{3}\right) \times 1000$$

$$= 6000 + \frac{1000}{3}$$

$$= \frac{18000 + 1000}{3} = 6333.33$$

→ $P_{95} = ?$

$$L_{95} + \frac{\left(95 \times \frac{N}{100} - C.f\right)}{f} \times h$$

$$= 6000 + \left(\frac{19 - 17}{3}\right) \times 1000$$

$$= 6000 + \frac{2000}{3} = \frac{18000 + 2000}{3}$$

$$= 6666.\dot{6}\dot{6}$$

Ques	Class	f	Cf
	4-8	11	11
Q_1	8-12	19	30
	12-16	23	53
	16-20	15	68
Q_3	20-24	12	80
	24-28	11	91
	28-32	8	99

$N = 99$

Date / /
Page

→ $Q_1 = LQ_1 + \frac{\frac{N}{4} - cf}{f} \times h$ $\quad \frac{N}{4} = \frac{99}{4} = 24.$

$8 + \left(\frac{24.75 - 11}{19}\right) \times 4$

$8 + \frac{13.75 \times 4}{19}$

$= 8 + \frac{55}{19} = \frac{152 + 55}{19} = \frac{207}{19} = 10.89$

→ $Q_3 = LQ_3 + \frac{\left(3 \times \frac{N}{4} - cf\right)}{f} \times h$ $\quad 3 \times \frac{N}{4} = \frac{3 \times 99}{4} = 74.2$

$20 + \frac{(74.25 - 68)}{13} \times 4$

$= 20 + \frac{6.25}{3}$

$= \cancel{26.25} = 22.08$

→ ~~$D_3 = L_{D3} + \frac{\left(3 \times \frac{N}{10} - cf\right)}{f} \times h$~~ ~~$3 \times 99$~~

Ques

	class	f	cf
	4-8	11	11
d_3	8-12	19	30
P_{40}	12-16	23	53
	16-20	15	68
P_{75} d_7	20-24	12	80
	24-28	11	91
	28-32	9	100

$N = 100$

Date...../...../.....

Page.........

→ $D_3 = LD_3 + \left(\dfrac{3 \times \frac{N}{10} - Cf}{f}\right) \times h$ $\qquad 3 \times \frac{100}{10} = 30$

$= 8 + \left(\dfrac{30 - 11}{19}\right) \times 4$

$= 8 + \dfrac{19}{19} \times 4 = 12$

$7 \times \frac{100}{10} = 70$

→ $D_7 = LD_7 + \left(\dfrac{7 \times \frac{N}{10} - Cf}{f}\right) \times h$

$20 + \dfrac{70 - 68}{\cancel{12}\ 3} \times 4$

$20 + \dfrac{2}{3} = \dfrac{60 + 2}{3} = \dfrac{62}{3} = 20.66$

→ $P_{40} = l_{P_{40}} + \left(\dfrac{40 \times \frac{N}{100} - Cf}{f}\right) \times h$ $\qquad 40 \times \frac{100}{100} = 40$

$= 12 + \left(\dfrac{40 - 30}{23}\right) \times 4$

$= 12 + \dfrac{40}{23} =$

→ $P_{75} = l_{P_{75}} + \left(\dfrac{75 \times \frac{N}{100} - Cf}{f}\right) \times h$ $\qquad 75 \times \frac{100}{100} = 75$

$20 + \dfrac{75 - 68}{12} \times 4 =$

Date....../....../......
Page..........

Mode :- represent the frequency occurring value

Mode is also a measure of central tendency defined as the value which ~~often~~ occurs most often. If the data is arranged in the form of frequency table the corresponding class to the maximum frequency is called model class.

Ques 7, 4, 5, 1, 7, 3, 4, 6, 7

The value that is repeated most number of time is 7
↳ mode

Ques 19, 20, 21, 24, 27, 30
no mode

Ques 12, 15, 11, 12, 19, 15, 24, 27, 20, 19, 12, 15
12 and 15 both are mode

Ques find mode

wage	no of employees
45	12
50	11
55	14 → maximum / mode
60	13
65	12
70	10
75	9

Date ___/___/___

Page ______

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

Ques

family size	no. of families	
1-3	7	$\to f_0$
$\to$ 3-5	8	$\to f_1$
5-7	2	$\to f_2$
7-9	2	
9-11	1	

$$\text{Mode} = 3 + \left(\frac{8-7}{2\times 8 - 7 - 2}\right) \times 2$$

$$= 3 + \frac{1}{16-7-2} \times 2$$

$$3 + \frac{2}{7} = \frac{21+2}{7} = \frac{23}{7} = 3.28$$

Ques

class	f	
5-11	22	
11-17	14	f_0
$\to$ 17-23	24	f_1
23-29	11	f_2
29-35	23	
35-41	20	

$$\text{mode} = 17 + \left(\frac{24-14}{2\times 24 - 14 - 11}\right) \times 6$$

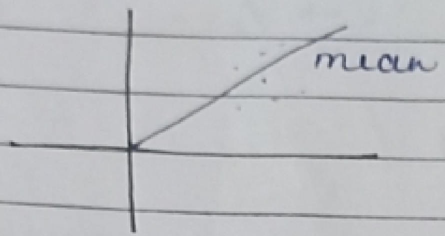
Date: 6^{th} September
Page

$$= 17 + \frac{10}{23} \times 6$$

$$= 17 + \frac{60}{23} = \frac{451}{23} = 19.60$$

Measures of dispersion

2, 4, 7, 9, 1, 5



Range : Difference b/w highest value to the lowest value

Q find range
4, 8, 6, 6, 2, 9, 9, 6, 9, 13
$13 - 2 = 11$

Quartile Range or interquartile Range

$Q_3 - Q_1$

Semi-Interquartile Range (Quartile Deviation)

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

Average mean deviation about mean

$$\frac{\Sigma(x_i - \mu)}{n}$$ for ungrouped data

Date / /
Page

$\frac{\Sigma f_i (x_i - m)}{n}$ for grouped data.

Que 4, 8, 1, 6, 6, 2, 9, 3, 6, 9 find average mean deviation

Mean $\mu = \frac{\Sigma x_i}{n} = \frac{54}{10} = 5.4$

$x_i - \mu$	
-1.4	
-2.6	by formula
-4.4	
0.6	$= \frac{\Sigma (x_i - \mu)}{n}$
0.6	
-3.4	
3.6	$= \frac{5.2}{10}$
-2.4	
0.6	$= 0.52$ ans
7.6	

$\Sigma x_i - \mu =$ -5.2

Variance (σ^2)

formula: $\frac{\Sigma (x_i - \mu)^2}{n}$ or $\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$

Standard deviation

$$\sqrt{\frac{\Sigma (x_i - \mu)^2}{n}}$$

variance

$$\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2 \qquad d = x_i - A \qquad \text{ungrouped}$$

or

~~Standard~~

$$\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2 \qquad \text{grouped}$$

Question find variance

x	$x_i - \mu$	$(x_i - \mu)^2$	x^2
9	2	4	81
8	-1	1	64
6	-1	1	36
5	-2	4	25
8	1	1	64
6	-1	1	36

$\Sigma x = 42$ $\qquad \Sigma (x_i - \mu)^2 = 12 \qquad \Sigma x^2 = 306$

$$\mu = \frac{9+8+6+5+8+6}{6} = \frac{42}{6} = 7$$

by formula $\sigma^2 = \frac{\Sigma (x_i - \mu)^2}{n} = \frac{12}{6} = 2$ ans

or

by another formula we need to do x^2

$$\sigma^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$$

$$= \frac{306}{6} - \left(\frac{42}{6}\right)^2 = 51 - 49$$

$$= 2 \text{ ans}$$

$A = 25$

Class	Profits	f	x or m	$d = m - A$	d^2	fd^2	fd
	0-10	8	5	-20	400	3200	-160
	10-20	12	15	-10	100	1200	-120
	20-30	20	25 A	0	0	0	0
	30-40	30	35	10	100	3000	300
	40-50	20	45	20	400	8000	400
	50-60	10	55	30	900	9000	300
		$N = 100$				$\Sigma fd^2 = 24400$	$\Sigma fd = 720$

by formula

$$\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2$$

$$= \frac{24400}{100} - \left(\frac{720}{100}\right)^2$$

$$= 244 - 51.84 = 193.16$$

$$S.D = \sqrt{193.16} = 13.89$$

Q* The arithmetic mean & standard deviation of series of 20 items are 20 cm and 5 cm, But while calculation, an item with value 13 was misread as 30, find the correct mean & correct S.D

$\frac{30}{13}$

$n = 20$

mean $\rightarrow \frac{\Sigma x_i}{n} = 20 \Rightarrow \Sigma x_i = 400$

$400 - 30 + 13 = \underline{383}$
new or corrected total sum

$$\text{Corrected mean} = \frac{\text{corrected sum}}{n} = \frac{383}{20} = 19.15$$

Date....../....../......

Page..............

Correct SD = $\sqrt{\text{variance}}$

SD = 5 => v = 25

$$25 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$$

$$25 = \frac{\Sigma x^2}{20} - (20)^2$$

$$\Sigma x^2 = 20(25 + 400)$$
$$20 \times 425$$
$$\Sigma x^2 = 8500$$ uncorrected

Σx^2 (corrected)
$= 8500 - 30^2 + 13^2$
$= 8500 - 900 +$
$= 7769$

$$\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$$

Corrected variance = $\frac{7769}{20} - (19.15)^2$

$= 21.73$

Corrected S.D = $\sqrt{\text{variance}}$

$= \sqrt{21.73} = 46615$

Q. find variance and SD

C I	f	M	d = M - A	d^2	fd^2	fd
0-6	23	3	-18	324	7452	-414
6-12	27	9	-12	144	3888	-324
12-18	22	15	-6	36	792	-132
18-24	14	21 = A	0	0	0	0
24-30	11	27	6	36	396	66
30-36	19	33	12	144	2736	228
36-42	21	39	18	324	6804	378

N = 137 $\quad \Sigma fd^2 = 22068 \quad \Sigma fd = -198$

$$\sqrt{\sigma^2} = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

$$= \sqrt{\frac{22068}{137} - \left(\frac{-198}{137}\right)^2}$$

$$= \sqrt{\frac{22068}{137} - \left(\frac{\cancel{2377764}\ 39204}{18769}\right)}$$

$$= \sqrt{161.08 - (-1.44)^2}$$

$$\sqrt{161.08 - 2.07} = 159 \text{ variance}$$

$$SD = \sqrt{159}$$

$$= 12.60 \text{ ans}$$

Date......../......../

Page........

Product A	Product B	P.A $x_i - \bar{x}$	P.B $x_i - \bar{x}$
59	150	11.15	
75	200	27.15	
27	125	-20.15	
63	310	15.15	
27	330	24.15	
28	250	-19.85	
56	225	8.15	
= 335			

find mean, SD for Product A and Product B

$\bar{x} = \frac{\Sigma x_i}{N} = \frac{335}{7} = 47.85$ Product A mean

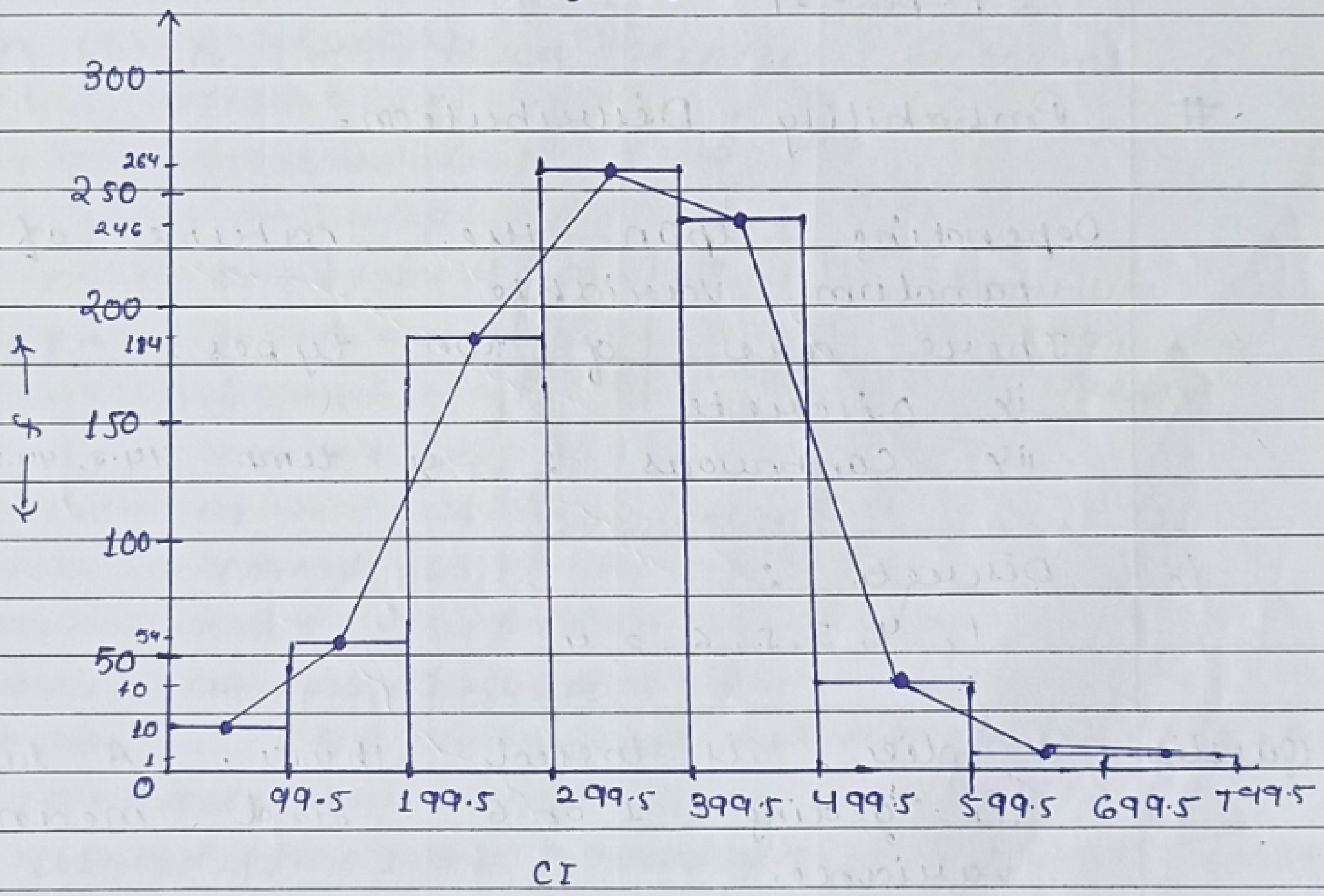
$\bar{x} = \frac{\Sigma x_i}{N} = \frac{1590}{7} = 2227.14$ Product B mean

$\sigma = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{N}}$

1Q Draw the histogram and frequency polygon for the following distribution also calculate the arithmetic mean

Class Interval	frequency	m / x	Class Boundary
0 - 99	10	49.5	0 - 99.5
100 - 199	54	149.5	99.5 - 199.5
200 - 299	184	249.5	199.5 - 299.5
300 - 399	264	349.5=A	299.5 - 399.5
400 - 499	246	449.5	399.5 - 499.5
500 - 599	40	549.5	499.5 - 599.5
600 - 699	1	649.5	599.5 - 699.5
700 - 799	1	749.5	699.5 - 799.5

$\Sigma f = 800$



Arithmetic mean:

$d = \frac{x-A}{h}$	fd'
-4	-40
-3	-162
-2	-368
-1	-264
0	0
1	40
2	2
3	3
	$\Sigma fd' = -789$

here $h = 100$

let $A = 449.5$

$N = \Sigma f = 800$

$$\bar{x} = 449.5 + \left(\frac{-789}{800}\right) \times 100$$

$$= 449.5 - 98.625$$

$$\bar{x} = 350.875 \quad \text{ans}$$

*Q2 The following marks were given to a batch of candidate:

66	62	45	79	32	51	56	60	51	49
25	42	54	54	58	70	43	58	50	52
38	67	50	59	48	65	71	30	46	55
82	51	63	45	53	40	35	56	70	52
67	55	57	30	63	42	74	58	44	55

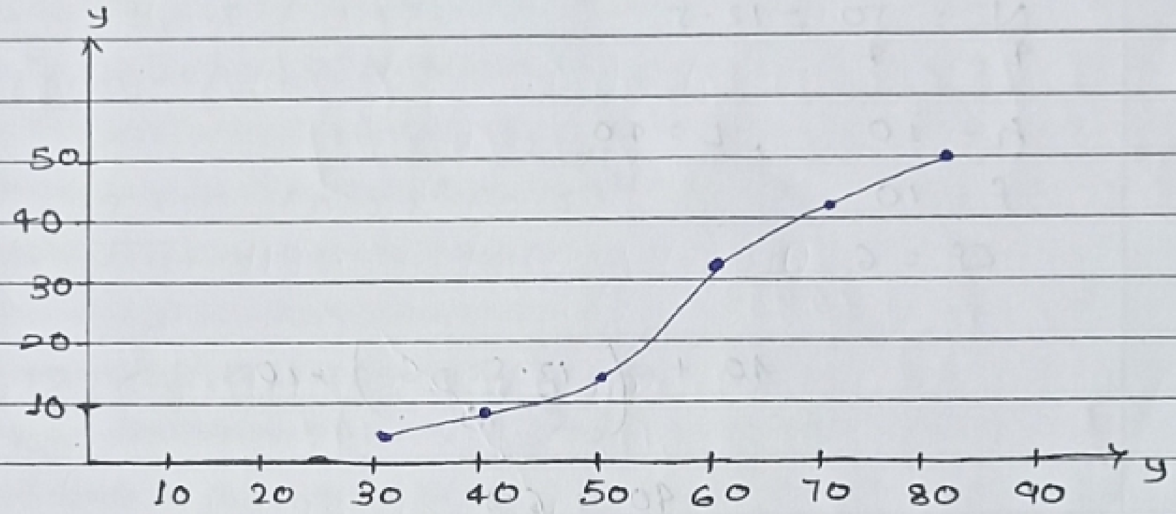
Draw a C.f curve. Hence find the proportion of candidate securing more than 50 marks also mark off the median, the first and third quartile.

Date ___/___/___
Page ______

2ans

class interval	f	c.f
20 - 30	1	1
30 - 40	5	6
40 - 50	10	16
50 - 60	20	36
60 - 70	8	44
70 - 80	5	49
80 - 90	1	50
	50	

cumulative frequency curve



candidates who securing more than 50 marks

total number of candidate = 50

candidate with more than 50 marks = 32

$$\text{proportion} = \frac{\cancel{32}^{16}}{\cancel{50}_{25}} = 64\%$$

Median

$\frac{N}{2} = \frac{50}{2} = 25$, $h = 10$, $l = 50$

$cf = 16$, $f = 20$

Date / /
Page

median = $L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times h$

$= 50 + \left(\frac{25 - 16}{20}\right) \times 16$

$= 50 + \frac{9}{2}$

$= 50 + 4$

54 ans

Quartile 1 :- $LQ_1 + \left(\frac{\frac{N}{4} - Cf}{f}\right) \times h$

$\frac{N}{4} = \frac{50}{4} = 12.5$

$h = 10$, $L = 40$

$f = 10$

$Cf = 6$

$40 + \left(\frac{12.5 - 6}{10}\right) \times 10$

$40 + 6.5$

46.5 ans

Quartele 3:- $\frac{3 \cdot N}{4} = \frac{3 \times 50}{4} = 37.5$, $LQ_3 = 60$

$h = 10$, $cf = 36$, $f = 8$

$60 + \left(\frac{37.5 - 36}{3}\right) \times 10$

$= 60 + \frac{1.5}{8} \times 10$

$= 60 + \frac{15}{8} = 60 + 1.875 = 61.875$ ans.

Date ___/___/___
Page ______

Q3 find the mean, median, mode for the following

Mid value x	frequency	$x_i f_i$	Cf	
15	2	30	2	
20	22	110	24	
25	19	475	43	lies c
30	14	420	57	
35	3	105	60	
40	4	160	64	
45	6	270	70	
50	1	50	71	
55	1	55	72	
	$\Sigma f = 72$	$\Sigma x_i f_i = 2005$		

mean $\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f} = \frac{2005}{72} = \cancel{28.54}$

$= 27.84$

Median:

$\frac{N}{2} = \frac{72}{2} = 36$ lies in $cf = 43$

$\because$ Thus the median is 25

Mode: The mode is the mid value with the highest frequency.
The mid-value 20 has the highest frequency of 22

$\because$ Mode = 20

Date / /
Page

Q4 Calculate mean, median and mode of the following data relating to weight of 120 article.

weight in gm	no of article f	x	$x_i f_i$	cf
0-10	14	5	70	14
10-20	17	15	255	31
20-30	22 → f_0	25	550	53 → cf
30-40	26 → f_1	35	910	79 ←←
40-50	23 → f_2	45	1035	102
50-60	18	55	990	120
	N = 120		$\Sigma x_i f_i$ = 3810	

→ mean $\bar{x} = \frac{\Sigma x_i f_i}{N} = \frac{3810}{120} = 31.75$ gm

→ median : $\frac{N}{2} = \frac{120}{2} = 60$ lies in 30-40

$$L_1 + \left(\frac{\frac{N}{2} - cf}{f}\right) \times h$$

$$= 30 + \left(\frac{60 - 53}{26}\right) \times 10$$

$$= 30 + \frac{70}{26}$$

$$= 30 + 2.69 = 32.69 \text{ gm}$$

→ Mode : $l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$

Date / /
Page

$$\Rightarrow 30 + \left(\frac{26 - 22}{2\times 26 - 22 - 23}\right) \times 10$$

$$\Rightarrow 30 + \left(\frac{4}{7}\right) \times 10$$

$$\Rightarrow 30 + \frac{40}{7}$$

$$\Rightarrow 30 + 5.71$$

$$\Rightarrow 35.71 \text{ gm}$$

5Q following table given the comulative frequency of the age of a group of 199 teachers. find the mean and median age of the group.

age in year	Cf	x	f	f_ix_i
20-25	21	22.5	21	472.5
25-30	40	27.5	19	522.5
30-35	90 cf	32.5	50	1625
35-40	130 le	37.5	40	1500
40-45	146	42.5	16	680
45-50	166	47.5	20	950
50-55	176	52.5	10	525
55-60	186	57.5	10	575
60-65	195	62.5	9	562.5
65-70	199	67.5	4	270
	N = 199			7682.5 Σf_ix_i

$$\text{Mean } \bar{x} = \frac{\Sigma x_i f_i}{\Sigma f} = \frac{7682.5}{199} = 38.60$$

the mean age is 38.60 years

Date / /
Page

→ Median :

The median class lies at $\frac{N}{2} = \frac{199}{2} = 99.5$

99.5 lies in 35-40 , $f = 40$, $h = 5$

by formula:- $L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times h$

$35 + \left(\frac{99.5 - 90}{\cancel{40}\ 8}\right) \times 5$

$35 + \frac{9.5}{8}$

$= 35 + 1.18$

$= 36.18$

So, the median age is approximately 36.18 y

6Q Compute the quartile deviation and Standard deviation for the following

x	f
100 - 109	15
110 - 119	44
120 - 129	133
130 - 139	150
140 - 149	125
150 - 159	82
160 - 169	35
170 - 179	16

N = 600

wrong.

Date...../...../.....

Page.........

f	m	cf	d = x - A	fd	f.d²
15	104.5	15	-30	-450	13500
44	114.5	59	-20	-880	17600
133	124.5	192	-10	-1330	13300
150	134.5 = A	342	0	0	0
125	144.5	467	10	1250	12500
82	154.5	549	20	1640	32800
35	164.5	584	30	1050	31500
16	174.5	600	40	640	25600
				$\Sigma fd = 1920$	$\Sigma fd^2 = 146800$

$$Q_1 = LQ_1 + \left(\frac{\frac{N}{4} - Cf}{f}\right) \times h$$

here $\frac{N}{4} = \frac{600}{4} = 150$ lies in 120-129

$f = 133$, $L = 120$, $Cf = 59$, $h = 09$

$$Q_1 = 120 + \left(\frac{150 - 59}{133}\right) \times 9$$

$$= 120 + \frac{91}{133} \times 9$$

$$= 120 + 6.16$$

$$= 126.16$$

$$Q_3 = L + \left(\frac{3 \times \frac{N}{4} - Cf}{f}\right) \times h$$

here; $\frac{3N}{4} = \frac{3 \times 600}{4} = 450$ lies in 140-149

$f = 125$, $L = 140$, $Cf = 342$, $h = 9$

Date......./......./.......
Page.......

$$Q_3 = 140 + \left(\frac{450 - 342}{125}\right) \times 9$$

$$= 140 + \left(\frac{108}{125}\right) \times 9$$

$$= 140 + 7.78$$

$$= 147.78$$

Quartile deviation : $\frac{Q_3 - Q_1}{2}$

$$= \frac{147.78 - 126.16}{2}$$

$$= \frac{21.62}{2} = 10.81$$

Standard deviation

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

$$= \sqrt{\frac{146800}{600} - \left(\frac{1920}{600}\right)^2}$$

$$= \sqrt{244.67 - 10.24}$$

$$= \sqrt{234.43}$$

$$= 15.31 = SD$$

Q7 The Scores obtained by two batsman A and B in 10 matches are given below

A :	30	44	66	62	60	34	80	46	20	38
B :	34	46	70	38	35	48	60	34	45	30

Calculating mean, SD and coefficient of variation for each, determine who is more efficient and who is more consistent.

Batsman A

$$\text{mean}_A = \frac{30+44+66+62+60+34+80+46+20+38}{10}$$

$$= \frac{480}{10} = 48$$

x_i	$x_i - \mu$	$(x_i - \mu)^2$
30	-18	324
44	-4	16
66	18	324
62	14	196
60	12	144
34	-14	196
80	32	1024
46	-2	4
20	-28	784
38	-10	100

$\Sigma(x_i-\mu)^2 =$ 3112

Given $N = 10$

Date / /
Page

$$\text{variance}_A = \frac{\Sigma (x_i - \mu)^2}{N}$$

$$= \frac{3112}{10}$$

$$= 311.2$$

$$SD_A = \sqrt{311.2} = 17.64$$

coefficient of variation

$$CV_A = \frac{SD}{Mean} \times 100$$

$$= \frac{17.64}{48} \times 100$$

$$= 36.75\ \%$$

Batsman B

$$mean_B = \frac{34 + 46 + 70 + 38 + 55 + 48 + 60 + 34 + 45 + 30}{10}$$

$$= \frac{460}{10} = 46$$

x_i	$x_i - \mu$	$(x_i - \mu)^2$
34	12	144
46	0	0
70	24	576
38	-8	64
55	9	81

48	2	4
60	14	196
34	-12	144
45	-1	1
30	-16	256

$\Sigma(x_i-\mu)^2 = 1466$

$$\text{variance}_B = \frac{\Sigma(x_i-\mu)^2}{N}$$

$$= \frac{1466}{10}$$

$$= 146.6$$

$$SD_B = \sqrt{146.6} = 12.10$$

coefficient of variation

$$CV_A = \frac{SD}{Mean} \times 100$$

$$= \frac{12.10}{46} \times 100$$

$$= 26.30\ \%$$

Thus Bateman A is more efficient, but Batsman B is more consistent

mean Efficiency: Batsman A has higher mean score (48) compa[red] to Bateman B (46), making Batsman A slightly more [e]

coff vari ↓ Consistency: Bateman B have lower C.V (26.30%) compa[red]
low to Batsman A (36.75%), Batsman B is more consistent

Date 19, Sep 21.
Page

Random Variable: If we perform some task that have

→ If a real variable x be associated with the outcome of random experiment, then since the value which x takes depend on chance it is called random variable

eg: coin to be tossed < Head → $\frac{1}{2}$, Tail → $\frac{1}{2}$

$$\boxed{\Sigma p_i = 1}$$

Probability Distribution.

Depending upon the nature of random variable
There are of two types:
i) Discrete
ii) Continuous : eg temp 44.3, 44.23,

i) Discrete :
1, 4, 3, 5, 6, 2, 4

Question A die is tossed thrice. A success is "getting 1 or 6". find mean & variance

by formula : $\mu = \Sigma p_i x_i$

$\sigma^2 = \Sigma p_i x_i^2 - \mu^2$

Date / /
Page

no of possible outcomes of a dice: 6

So, Probability of 1 or 6 (success) $= \frac{2}{6} = \frac{1}{3}$

probability of no likely outcome (failure) $= \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{8}{27}$

probability of neither 1 or 6 (failure) $= 1 - \frac{1}{3} = \frac{2}{3}$

${}^3C_1$ Probability of one likely outcome (success)

$= {}^3C_1 \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \cancel{\frac{4}{27}}\ \frac{4}{9}$

${}^3C_2$ Probability of two likely outcomes

$= {}^3C_2 \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$

${}^3C_3$ Probability of three likely outcome

$= {}^3C_3 \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27}$

x_i	0	1	2	3
$P(x_i)$	$\frac{8}{27}$	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{1}{27}$

Probability distribution Table

${}^nC_r = \frac{n!}{r!\,(n-r)!}$ → combination

n = total no of events

r = successful event, $\mu = 0 \times \frac{8}{27} + 1 \times \frac{4}{9} + 2 \times \frac{2}{9} + 3 \times$

$= 0 + \frac{4}{9} + \frac{4}{9} + \frac{1}{9} = \frac{\cancel{9}}{\cancel{9}} =$

$\mu = \Sigma p_i x_i = 1$

$\sigma^2 = \Sigma p_i x_i^2 - \mu =$

$= 1$

$\left(0 + \frac{4}{9} + \frac{8}{9} + \frac{\cancel{9}}{\cancel{27}}\frac{1}{3}\right) - 1 = \frac{15}{9} - 1 = \frac{6}{9} =$

40

Binomial Distribution first page

Date / /
Page

$(q-p)^n = q^n + {}^nC_1 pq^{n-1} + {}^nC_2 p^2 q^{n-2} +$

${}^nC_3 p^3 q^{n-3} - {}^nC_n p^n = 1$

Mean of B.D = np
variance of B.D = npq

Standard deviation of B.D = $\sqrt{npq}$

Repetative trials

two outcomes

Success (p) — failure (q)

$p + q = 1$

$p = 1 - q$ or $q = 1 - p$

for n no of trials, prob p of success and q prob. of failure then, out of 'n' attempts 'r' attempts should be success ${}^nC_r p^r q^{n-r}$

Q let for event, probability of success is 0.75 find the prob of 7 success out of 12 attempts

$n = 12$, $r = 7$, $p = 0.75$, $q = 0.25$

$\to$ ${}^nC_r p^r q^{n-r}$

${}^{12}C_7 (0.75)^7 (0.25)^5$

with 50 such set

$$50\ ({}^{12}C_7\ (0.75)^7\ (0.25)^5$$

$$ {}^{n}C_r = \frac{n!}{r!\,(n-r)!} = \frac{12!}{7!\,(12-7)!}(0.75)^7(0.25)^5$$

$$= \frac{12\times 11\times 10\times 9\times 8\times \cancel{7!}}{\cancel{7!}\times 5!}(0.75)^7(0.25)^5$$

*Q A die was thrown 9000 times and a throw of 5 or 6 was obtained 3240 times. On assumption of random throwing, do data indicate an unbaised die.

ans $n = 9000$,

Prob that 5 or 6 will come $= \frac{2}{6} = \frac{1}{3}$

$= P = \frac{1}{3}$

$q = 1 - P = 1 - \frac{1}{3} = \frac{2}{3}$

$$z = \frac{x - np}{\sqrt{npq}}$$

$$= \frac{3240 - (9000 \times \frac{1}{3})}{\sqrt{\cancel{9000}^{3000} \times \frac{1}{\cancel{3}} \times \frac{2}{3}}}$$

$$= \frac{3240 - 3000}{\sqrt{2000}} = 5.36$$

for 5%, $z = 5.36 > 1.96$

Assumption wrong, baised die

Date / /
Page

for 1% $z = 5.36 > 2.58$

fail

Binomial distribution

Also called as Bernoulli distribution beca[use] it is based on bernoullis process

- It is a type of probability distribution in which probability of any event have two possible results such as success or failure

eg if a coin is tossed, what is the probability of coming head.

if head → success
if tail → failure

$P(H) = \frac{1}{2} = 0.5$ success denoted by p

failure $(q) = 1 - p$
$q = 1 - 0.5 = 0.5$

Terms in B.D

p → Probability of success
q → Prob. of failure
n → no of trials

formula → $\boxed{P(r) = {}^nC_r\, p^r q^{n-r}}$

Date / /
Page

here

p = Prob of success in single trial
q = Prob of failure $(1-p)$
n = no of trials
r = no of success in n trials

$$^nC_r = \frac{n!}{r!\,(n-r)!}$$

formulae

mean $(\bar{x}) = np$
S.D $(\sigma) = \sqrt{npq}$
Variance $(\sigma^2) = npq$

Q The mean of B.D is 40 and SD is 6 Calculate n, p, q

mean $\bar{x} \Rightarrow np = 40$

SD = 6
$\sqrt{npq} = 6$

$\sqrt{npq} = 6$
Sq both side

$(\sqrt{npq})^2 = 6^2$
$npq = 36$
$40q = 36$
$q = \frac{\cancel{36}}{\cancel{40}} = \frac{9}{10}$

$q = \frac{9}{10} = 0.9$

$q = 1 - p$
$p = 1 - q$
$p = 1 - 0.9$
$p = 0.1$

$np = 40$
$n \times 0.1 = 40$
$n = \frac{40}{\frac{1}{10}} = \frac{40}{1}$ = 4

eg Multipunch tablet machine produce 12% defective tablets. what is the prob. that out of random sample of 20 tablets produced by machine 4 are defective

$n = 20$

$p = 12\% = \frac{12}{100} = 0.12$ success

$q = 1 - p$

$= 1 - 0.12 = 0.88$

$r =$ no of success in n term $= 4$

$P(r) = {}^{n}C_{r}\ p^{r} q^{n-r}$

${}^{20}C_{4}\ p^{4} q^{20-4}$

$= \frac{20!}{4!(20-4)!} \times (0.12)^{4} \times (0.88)^{16}$

$= \frac{\overset{5}{\cancel{20}} \times 19 \times \overset{\cancel{6}\,3}{\cancel{18}} \times 17 \times \cancel{16!}}{\cancel{4!} \times \cancel{16!} \times \cancel{3} \times \cancel{2}} \times 0.00020736 \times 0.1293$

$= 0.0866$

Date / /
Page

Binomial distribution

Q4 * In 256 sets of 12 threes of a coin in how many sets one can expect 8 heads and 4 tails?

$p = \frac{1}{2}$ $n = 12$

$r = 8$

$q = \frac{1}{2}$

$^nC_r (p)^r (q)^{n-r}$; $^nC_r = \frac{n!}{r!(n-r)!}$

$\Rightarrow$ $^{12}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4$

$\Rightarrow$ $\frac{12!}{8! \times 4!} (0.5)^8 (0.5)^4$

$256 \times {}^{12}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4$

$= 256 \times 0.1188$

$= 30.41$

* Q In a sampling of large part manuf-actured, mean no of defective in a sample of 20 is 2. out of 1000 such samples, how many would be expected to contain at least 3 defective parts.

$np = 2$

mean $= 2$

$20 \times p = 2$

$p = \frac{2}{20} = 0.1$

Date / /
Page

$np = 2$ $\qquad p \to$ no of defective pieces

$20 \times p = 2$

$\boxed{p = 0.1}$

$p \to$ no of defective pieces

$q = 1 - 0.1 = 0.9$ (prob. of non defective piece)

$n = 20,$

$\Rightarrow$ $1 -$ (prob. of 0 defective + prob. of 1 defective + prob of 2 defective)

$$\Rightarrow 1 - \left({}^{20}C_0 (0.1)^0 (0.9)^{20} + {}^{20}C_1 (0.1)^1 (0.9)^{19} + {}^{20}C_2 (0.1)^2 (0.9)^{18} \right)$$

$$\Rightarrow {}^{20}C_0 (0.1)^0 (0.9)^{20}$$

$$= \frac{20!}{0!(20-1)!} \times (0.1)^0 \times (0.9)^{20} = \frac{20! \times 1 \times (0.9)^{20}}{19!} = \frac{20 \times \cancel{19!}}{\cancel{19!}} =$$

$$= 0.121$$

$$\Rightarrow {}^{20}C_1 (0.1)^1 \times (0.9)^{19}$$

$$\Rightarrow \frac{20!}{1!(20-1)!} \times (0.1)^1 \times (0.9)^{19}$$

$$= \frac{20 \times \cancel{19!}}{\cancel{19!}} \times 0.01 \times 0.13$$

$$= 0.26 \approx 0.27$$

$^{20}C_2 \times (0.1)^2 \, (0.9)^{18}$

$\Rightarrow \dfrac{20!}{2!\,(20-2)!} \times (0.1)^2 \times (0.9)^{18} = 0.285$

$\Rightarrow$

for 1000

1000×0.324

$= 324$

$\Rightarrow 1 - 0.121 + 0.27 \times 0.285$

$= 0.324$

Q let x be random variable that follows Binomial Distribution, with Expectation $E(x) = 7$, variance, $V(x) = 6$

find Probability of Sources.

$E(x) = np$, $V(x) = npq$ $\Rightarrow \dfrac{7}{6} = \dfrac{np}{npq}$

here $\dfrac{1}{q} = \dfrac{7}{6}$

$q = \dfrac{6}{7}$

Normal Distribution

np : mean

$\sqrt{npq}$: S.D

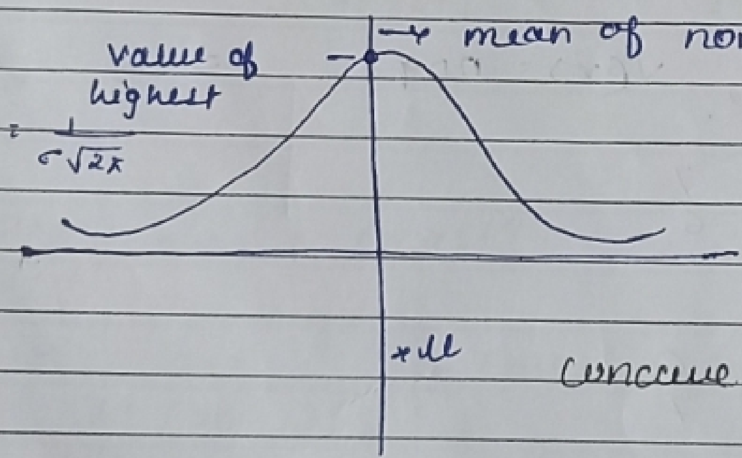
$$z = \frac{x - np}{\sqrt{npq}}$$

where x is binomial variate. p is probability of success. q is the prob. of failure for large value of n and not very small value of p and q the variable z approximate to normal variable

σ = SD

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma}}$$

properties of normal distribution



normal dis.. is symmetric about mean.

concave downward

The graph of the normal distribution is bell shape

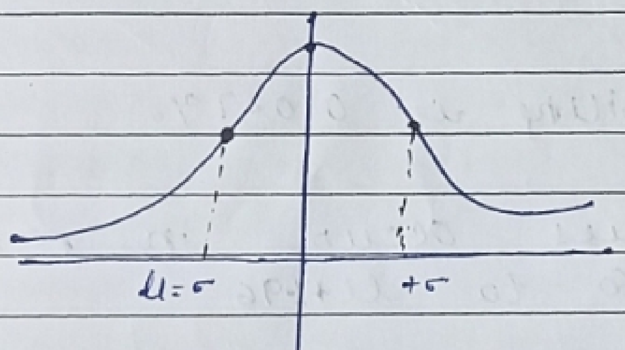
μ: mean of normal dis

$P(Z \le (-10))$ or $P(Z \le 12)$ is same

- It is unimodel which ordinate decreasing both side about its mean.
- The maximum ordinate is $\frac{1}{\sigma\sqrt{2\pi}}$
- value of highest peak inversely proportional to standard deviation
- Since the normal dis. is symmetric its mean median and mode have the same value

point of inflextion.

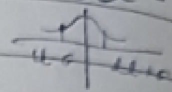
The point in the graph at which the curvature is changing either from concave upward to downward and concave downward to upward.



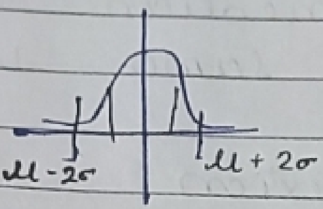
Mean deviation about mean = μ:

$$\boxed{\sigma = \frac{4}{5}}$$

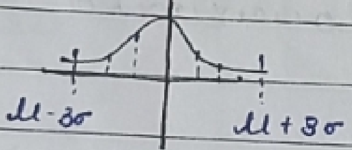
- The area under the normal curve b/w the range $\mu + \sigma$ to $\mu - \sigma$ is 0.6826 or 68.26%.



- The area under the normal curve b/w the limits $\mu - 2\sigma$ to $\mu + 2\sigma$ the probability is 95.5%



- from the limit $\mu - 3\sigma$ to $\mu + 3\sigma$ the probability of having event is 99.73%.



outside probability is 0.027%.

Hi. 95% of values occur in range $x = \mu - 1.96\sigma$ to $\mu + 1.96$

- 99.9% values occur in the limit $x - 3.29\sigma$ to $x + 3.29\sigma$

Note:-
- z for 5% sig — 1.96
- z for 1% significance — 2.58

Sum of all prob = 1

$\Sigma p = 1$

5% significance on $\mu - 2\sigma$ to $\mu + 2\sigma$

1% sig $\mu - 3\sigma$ to $\mu + 3\sigma$,

Q1. x is a normal variate with mean 30 and S.D → 5, find the prob

i) $26 \le x \le 40$

ii) $x \ge 45$

iii) $|x - 30| > 5$

(I) $x = 26$, $z = \frac{x - \mu}{\sigma}$

$= \frac{26 - 30}{5} = -0.8$

$x = 40$, $z = \frac{40 - 30}{5} = 2$

$-0.8 \le z \le 2$

$\underbrace{-0.8 \le z \le 0}_{\downarrow} + 0 \le z \le 2$

$0 \le z \le 0.8$

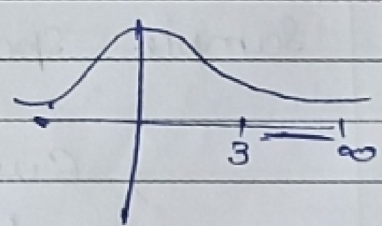
$0.2881 + 0.4772$

=

(ii) $x = 45$, $z = \frac{45 - 30}{5} = 3$

$0.5 - P(0 \le z \le 3)$

↓ Probability



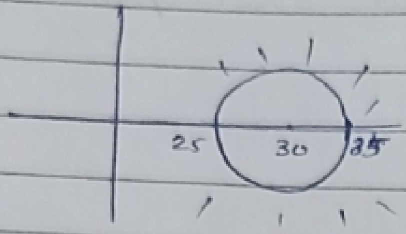
$P(z \ge 3) = 0.5 - P(0 \le z < 3)$

$0.5 - 0.4987$

$= 0.0013$

Maths.

(iii) $|x - 30| > 5$ general eqn of circle, $|z - a| = r$



find outside prob outside the circle

$1 - P(\text{inside the circle})$

$P|x - 30| > 5 = 1 - P||x - 30| < 5$

$1 - [25 \le x \le 35]$

$1 - [-1 \le z \le 1]$

$= 1 - 2 \times p(0 \le z \le 1)$

$= 1 - 2 \times 0.3413$

$= 0.3174$

Probability and Set Notation

- Random experiments
- Sample space
- Event

exhaustive event | Mutually exclusive ~~output~~ event | equaly likely event

Date...../...../.....

Page.....

Random Experiment

Experiment which ~~have~~ are perform essentally under same condition but the result hat been predicted is known as Random Experiment

Sample space

the set of all possible Outcome of a Random experiment is known as sample space

Event

the Outcome of a random experiment is called an event.

dice = { 1, 2, 3, 4, 5, 6 }

P(3) → event

Thus event is the subset of Sample space.

Exhaustive event

- A set of events is said to be exhaustive if it include all the possible event.

Equally Likely

if one of the events cannot be expected to happen in preference to another

Date / /
Page

then such event is known as equally likely event

Q An urn contains 5 red and 10 black ~~box~~ balls. Eight of them are placed in another urn, what are chances, that it contains 2 red and 6 black balls

5R 10B		2R 6B

15

$$^{15}C_{8} = \frac{15!}{8!\,(15-8)!}$$

$$= \frac{\cancel{15}^{5} \times \cancel{14}^{2} \times 13 \times \cancel{12}^{2} \times 11 \times 10 \times 9 \times \cancel{8!}}{\cancel{8!} \times \cancel{7} \times \cancel{6} \times \cancel{5} \times \cancel{4}_{2} \times \cancel{3} \times \cancel{2} \times 1}$$

$$= \frac{13 \times 11 \times 10 \times 9}{2} = \frac{12870}{2}$$

$$= \cancel{\tfrac{128700}{20}} = \cancel{715} = 6435$$

$$^{5}C_{2} = \frac{5!}{2!\,(5-2)!}$$

$$= \frac{5 \times \cancel{4}^{2} \times \cancel{3!}}{\cancel{2!} \times \cancel{3!}} = 10$$

$$^{10}C_{6} = \frac{10!}{6!\,(10-6)!} = \frac{10!}{6! \times 4!}$$

Date / /
Page

$$= \frac{10 \times \overset{3}{\cancel{9}} \times \overset{2}{\cancel{8}} \times 7 \times \cancel{6}}{\cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2}}$$

$$= 70 \times 3$$

$$= 210$$

Probability = $\frac{\text{num of event}}{\text{sample space}} = \frac{n(E)}{S}$

$$P = \frac{n(E)}{S}$$

$$= \frac{10 \times 210}{6435}$$

$$= \frac{2100}{6435} = \frac{140}{429} = 0.326 \text{ ans}$$

Independent event

two event are said to be independent if happening and failure of 1 event does not effect the happening & failure of another event

Dependent event

two event are said to be dependend if happening & failure of 1 event affet the happening & failure of another event

Conditional probability

for two dependent events the prob of

Occurance of B, given that A has already been occured is

$$P\left(\frac{B}{A}\right) \quad \text{or} \quad P\left(\frac{A}{B}\right)$$

\# $P(A \cap B) = P(A) \cdot P\left(\frac{B}{A}\right)$

$= P(B) \cdot P\left(\frac{A}{B}\right)$ } This is a case of dependent event

Q Two cards are drawn in succession from pack of 52 cards, find the chances, that first card is king and second card is queen
i) Replaced
ii) Not Replaced.

$$P(K) = \frac{4}{52} = \frac{1}{13}$$

$$P(Q) = \frac{4}{52} = \frac{1}{13}$$

(i) $P(K) \cdot P(Q) = \frac{1}{13} \times \frac{1}{13} = \frac{1}{169}$

(ii) $P(K) = \frac{4}{52} = \frac{1}{13}$, $P(Q) = \frac{4}{51}$

$$P(K) \cdot P(Q) = \frac{1}{13} \times \frac{4}{51} = \frac{4}{663} = 0.007$$

Date / /
Page

Q A pair of dice is tossed twice. Find probability of 7 points
a) once
b) at least once
c) twice

P = probability of first toss + Prob of second toss

$\frac{1}{6} \times \frac{5}{6} + \frac{5}{6}$

Q A coin is tossed 400 times, and head turned up 216 times. Test the hypothesis that coin is unbiased at 5% level of significance

Let us assume coin is unbaised

$p = \frac{1}{2}$

$q = 1 - \frac{1}{2}$

$= 1 - \frac{1}{2}$

$= \frac{1}{2}$

no of time head appear by Prob.

$P(H) = \frac{1}{2}$, $n = 400$, $E(H) = \frac{1}{2} \times 400 = 200$

but head appears = 216

$z = \frac{x - np}{\sqrt{npq}}$

$npq = 400 \times \frac{1}{4} \times \frac{1}{2}$

$= \frac{216 - 200}{\sqrt{100}}$

$= \frac{16}{10}$

$= 1.6$

Date / /
Page

Sampling and Inference

Introduction

- we know that a small section selected from the population is called a sample.

- The process of drawing a sample is called sampling.

- It is essential that a sample must be a random selection so that each member of the population has the same chance of being included in the sample. Thus the fundamental assumption underlaying theory of sampling is Random sampling

Objective of Sampling

Sampling aims to gathering the maximum information about the population with the minimum effort, cost and time. The object of sampling studies is to obtain the best possible values of the parameters under specific conditions.

Date / /
Page

Simple Sampling

A special case of random sampling in which each event has the same probability P of success and the chance of success of different ~~success~~ event ~~Event~~ are independent is known as simple sampling

Sampling distribution

Consider all possible samples of size n which can be drawn from the given population at random. for each sample, we can compute the mean. the mean of all the samples will not be identical. if we group their different mean according to their frequency, the frequency distribution so formed is known as sampling distribution of mean.

while drawing each sample be put back the previous sample so that the parent population remain the same this ~~is~~ called sampling with replacement.

Standard error.

the standard deviation of the sampling distribution is called standard error (S.E)

Date / /
Page

Similarly, the S.D of the Sa

Similarly, the S.D error of sampling distribution of mean is called standard error of mean

the reciprocal of standard error is called precision

- if the sample size is ≥ 30 $(n \geq 30)$, the sample is considered as large

- the sampling distribution of large samples is assumed to be normal.

Testing the Hypothesis

To reach the decision about population on the basis of sample information we make certain assumptions about the population involved, such assumptions, which may be true or may not be true, are called Statistical Hypothesis.

Coefficient of Correlation

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

where $X = x - \bar{x}$

$Y = y - \bar{y}$

Q4 A Test of intelligence and Engineering ability applied to 10 students with following data. calculate coefficient of Correlation.

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

$$\bar{x} = \frac{105 + 104 + 102 + 101 + 100 + 99 + 98 + 96 + 93 + 92}{10}$$

$$= \frac{990}{10} = 99$$

$$\bar{y} = \frac{101 + 103 + 100 + 98 + 95 + 96 + 104 + 92 + 97 + 94}{10}$$

$$= \frac{980}{10} = 98$$

Date __/__/__

Page ____

$X = x - \bar{x} =$ $105 - 99 = 6$
$104 - 102 = 2$
$102 - 99 = 3$
$101 - 99 = 2$
$100 - 99 = 1$
$99 - 99 = 0$
$98 - 99 = -1$
$96 - 99 = -3$
$93 - 99 = -6$
$92 - 99 = -7$

$Y = y - \bar{y} =$ $101 - 98 = 3$
$103 - 98 = 5$
$100 - 98 = 2$
$98 - 98 = 0$
$95 - 98 = -3$
$96 - 98 = -2$
$104 - 98 = 6$
$92 - 98 = -6$
$97 - 98 = -1$
$94 - 98 = -4$

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	XY	X^2	Y^2
105	101	6	3	18	36	9
104	103	2	5	10	4	25
102	100	3	2	6	9	4
101	98	2	0	0	4	0
100	95	1	-3	-3	1	9
99	96	0	-2	0	0	4
98	104	-1	6	-6	1	36
96	92	-3	-6	18	9	36
93	97	-6	-1	6	36	1
92	94	-7	-4	28	49	16
				$\Sigma XY = 92$	$\Sigma X^2 = 170$	$\Sigma Y^2 = 140$

Date...../...../.....
Page.....

$\Sigma xy = 92$

$\Sigma x^2 = 170$

$\Sigma y^2 = 140$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

$$= \frac{92}{\sqrt{170 \times 140}}$$

$$= \frac{92}{\sqrt{23800}}$$

$$= \frac{92}{154.272486} = 0.59$$

positive correlation

T-Test

test of significance

- T-test: $n < 30$
- Z-test: $n \geq 30$

$$t = \frac{\bar{x} - \mu}{S.D} \sqrt{n-1}$$

There are three type of T-test

T-test

- Simple Sample
- Independent
- paired

Date ___/___/___

Page ______

- if we are picking only one sample sets thes are called simple Sample

- if we are picking two sample sets which is not related to each other is called independent

- if we are picking two or more Sample sets which is related to each other is called paired

formula

$$\boxed{t = \frac{\bar{x} - \mu}{S \cdot D} \sqrt{n-1}}$$

here, $\bar{x} \to$ sample mean
$\mu \to$ population mean
$n \to$ Sample size
$S \cdot D \to$ standard deviation

Q A machinist is making engine parts with axle diameter of 0.7 inch a random Sample of 10 parts shows mean diameter of 0.742 inch with standard deviation of 0.04 inch on the basis of this sample, would you say the work is inferior?

$\bar{x} = 0.742$ $\quad$ $S \cdot D = 0.04$

$\mu = 0.7$ $\quad$ $n = 10$

$5\% = 0.05$

Date........ /........ /

Page..............

$$t = \frac{\bar{x} - \mu}{SD} \sqrt{n-1}$$

$$= \frac{0.742 - 0.7}{0.04} \sqrt{10-1}$$

$$= \frac{0.042}{0.04} \cdot \sqrt{9}$$

$$= \frac{0.126}{0.04} = 3.15$$

$n = 10$

degree of freedom $= n-1 = 10-1 = 9$

2.26

t critical > t calculate = hypothesis accept

t critical < t calculate $\Rightarrow$ hypothesis reject

5% $\rightarrow$ $2.26 < 3.15$ = Reject

Q A certain stimulus administered to each of 12 patients results in increase in B.P

5, 2, 8, −1, 3, 0, −2, 1, 5, 0, 4, 6

can it be concluded that stimulus will be accompanied by increase in B.P