Measurement of Central tendency

Measure of central tendency is a single value which represent the whole set of figure and all or individual terms concentrate on average

MCT

- Mean
  - Arithmetic mean
  - Geometric mean
  - Harmonic mean
- Median
- Mode

(i) Arithmetic mean : AM($\bar{x}$) It is defined as sum of absorbation divided by total no of items

$$\boxed{\bar{x} = \frac{\Sigma x}{N}}$$

Q 8, 1, 6 $= \bar{x} = \frac{15}{3} = 5$

Q calculate the weighted mean of from the following data.

Per table (x)	No of Tables sold (f)	fx
36	14	504
40	11	440
44	9	396
48	6	288
		$\Sigma fx = 1628$

$\boxed{\bar{x} = \Sigma fx / N}$

$\boxed{\Sigma f = N}$

$\bar{x} = \frac{1628}{40}$

$= 40.7$

The following gives the marks of two candidates and the weighted average mark. by what figure ld be the second candidate had to increase his marks of 'B' all other marks remaing the same in order candidates have the same place.

	weight (f)	X	Y	fx	fy
A	1	70	80	70	80
B	2	65	64	130	128
C	3	58	56	174	168
D	4	63	60	252	240

$\bar{x} = ?$

$\bar{y} = ?$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f = N}, \quad \bar{y} = \frac{\Sigma fy}{\Sigma f = N}$$

$$\bar{x} = \frac{626}{10}, \quad \bar{y} = \frac{616}{10}$$

$$\boxed{\bar{x} = 62.6} \quad \boxed{\bar{y} = 61.6}$$

$$2 \times (64 + k) = 128 + 2k = 626$$

$$2k = 626 - 128$$

$$2k = 598$$

$$k = \frac{598}{2}$$

$$616 + 2k = 626$$

$$2k = 10$$

$$\boxed{k = 5}$$

Q. Shortcut Method when nothing is given

$$\boxed{\bar{X} = A + \left(\frac{\Sigma d}{N}\right)} \quad \boxed{d = X - A}$$

where A → Assumed mean

Σd ⇒ diviation of mid value from the assum

N → no of items/products

Q. the pocket allowance of 10 students is given. Calculate the Arith mean by taking 40 as assumed mean.

S.No	X	X − A = d
1	15 − A	−25
2	20 − A	−20
3	30 − A	−10
4	22	−18
5	25	−15
6	18	−22
7	40	0
8	50	10
9	55	15
10	65	25
		$\boxed{\Sigma d = -60}$

given

$\boxed{A = 40}$ $N = 10$

$$\bar{X} = A + \left(\frac{\Sigma d}{N}\right)$$

$$\bar{X} = 40 + \left(\frac{-60}{10}\right)$$

$$\boxed{\bar{X} = 34}$$

$$\bar{X} = 40 - 6$$

$$\boxed{\bar{X} = 34}$$

* HW →

$$\boxed{\bar{X} = \frac{\Sigma x}{N}}$$ discrete individual

$$\bar{X} = \frac{\Sigma fx}{N}$$ weighted mean / average

$$\boxed{\bar{X} = A + \frac{\Sigma fd}{N}}$$ shortcut

wory $\bar{X} = \frac{\Sigma fm}{\Sigma f}$

→ • Inclus
• cum

nclusive
upper
nterval.
for ex
values

Exclus
or e
up
for

→ Inclusive & Exclusive ~~frequency~~ series
• Cummulative frequency

Inclusive series: An inclusive series includes the upper limit and the lower limit in the class interval.

for ex → A class Interval of 10 - 19 would include values from 10 to 19 (both 10 and 19 inclusive)

* Exclusive series: A exclusive series in which the upper limit is not included in that class and in included in upcoming class.

for example: 0 - 10
11 - 20
21 - 30

In this series, a value of 10 would be counted in the 11 - 20 class, not the 0 - 10 class. so 10 is exclusive value.

* Cummulative frequency: Cummulative frequency is a concept in statics that refers to the running total of frequencies within a set of data, grouped by class intervals. It tells you how many data points fall at or below a specific value in your dataset.

Find the arithmetic mean for the following data.

Marks	f	m	fm
0 - 10	20	5	100
10 - 20	24	15	360
20 - 30	40	25	1000
30 - 40	36	35	1260
40 - 50	20	45	900

$\Sigma = 3620$

Direct Method

$$\bar{x} = \frac{\Sigma fm}{N}$$

where $N = \Sigma f$

$$\bar{x} = \frac{3620}{140}$$

$$\boxed{\bar{x} = 25.85}$$

Short cut Method

Marks	f	m	m − A = d	fd
0 − 10	20	5	−20	−40
10 − 20	24	15	−10	−240
20 − 30	40	25	0	0
30 − 40	36	35	+10	360
40 − 50	20	45	+20	400

Assumed mean $= \frac{125}{5}$ $\quad \Sigma fd = 120$

$\boxed{A = 25}$

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

where $d = M - A$

$d = 25$

$$\bar{x} = 25 + \frac{120}{14}$$

$$\bar{x} = 25.85$$

Step diviation Method (when Nothing is given)

Marks	f	m
0 − 10	20	5
10 − 20	24	15
20 − 30	40	25
30 − 40	36	35
40 − 50	20	45

$10 - 20 = 10$

$i = 10$

$$\bar{x} = A + \frac{\Sigma fd'}{N} \times i$$

where

$d' = \frac{d}{i}$

$d = m - A$

$$\bar{x} = 25 + \frac{\frac{25}{10}}{14} \times 10$$

$$\bar{x} = 25 + \frac{25}{14} \quad \bar{x} = 26.78$$

* Cumulative frequency

while comulative frequency, It is necessary to convert the series into simple series.

marks	cf	f
less than 10	5	5
20	17	12
30	31	14
40	41	10
50	49	8

$\bar{X} = A + \frac{\Sigma f d'}{N} \times i$

$\bar{X} = 25 + \frac{4.9}{} \times 10$

$\bar{X} =$

$\Sigma f = 49$

$\Sigma f' = \frac{\Sigma f}{i} = \frac{49}{10}$

$\boxed{\Sigma f' = 4.9}$

* Calculate the ~~frequency~~ missing value when it's mean is given.

wages	f	Σfx
110	25	2750
112	17	1904
113	13	1469
117	15	1785
119	14	1400
125	8	1006
128	6	768
130	2	260

$\Sigma f =$

$\bar{X} = \frac{\Sigma fx}{\Sigma f}$

$\bar{X} = 115.86$

$a = 120$ Ans

$115.86 = \frac{9966 + 140}{100}$

* sum of deviation of certain no of item measured from 2.5 is 50 and from 3.5 is -50 then find N and x, y.

ans ⇒ $\boxed{\bar{X} = A + \frac{\Sigma d}{N}}$ $A = 2.5$, $d = 50$

$\bar{X} = 2.5 + \frac{50}{N}$ — (1)

$\bar{X} = 3.5 - \frac{50}{N}$ — (2)

$\bar{X} = 2.5 + \frac{50}{100}$

$\boxed{\bar{X} = 3}$

$\boxed{N = 100}$

* correcting incorrect values of the mean

the mean of 100 is 80 by mistake 1 item is miss read 92 instead of 29 find the correct mean.

* Combined arithmetic mean will be x

$$\bar{X}_{12} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2} \quad \bar{X}_{12} =$$

(1) the mean heigh of 25 male worker 61 and the height of female worker is find the combined mean of

20 from home

Mathematical

① sum of deviation of the items from the mean is always 0.

$$\Sigma(X-\bar{X})=0$$

X	$X-\bar{X}$
5	−5
10	0
15	+5
$\bar{X}=10$	6

यहाँ से

* **Median:** Median is another measurement of central tendency

Median is defined as the middle of the series when arranged either in ascending or descending order.

Individual series

$$\boxed{\text{for odd} \quad M = \text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ item}} \quad \therefore (\text{odd})$$

Calculate median

22, 16, 18, 13, 15, 19, 12, 20, 23

~~12, 13, 16, 18, 19, 20~~

any → ~~12, 13, 16, 18, 19, 20~~

12, 13, 15, 16, 18, 19, 20, 22, 23

1 2 3 4 5 6 7 8 9

$M = \left(\frac{9+1}{2}\right)$ $\boxed{M=5}$

Median = 18

5th
1 200
2 217
3 264
4 282
5 296
6 299
7 316
8 317

for even no of item

4.5th item

$M = \frac{N+1}{2}$ $M = \frac{8+1}{2}$ $M = \frac{9}{2}$

$\boxed{M = 4.5}$

$\frac{282 + 296}{2} = 289$

$\boxed{M = 289}$

* for even no of item

$\boxed{\text{average of } \left(\frac{N}{2}\right)th \left(\frac{N}{2}+1\right)th \text{ term}}$

$\rightarrow \frac{8}{2} = 4$ $\left(\frac{8}{2}+1\right) \rightarrow 4+1 = 5$

* level of known of high = Median

X:	10	12	14	16	18	20	22
f:	2	5	12	20	10	7	3

calculate the median from the following data.

Discrete series

$\left(\frac{N+1}{2}\right) = M$

X	f	cf
10	2	2
12	5	7
14	12	19
16	20	39
18	10	49
20	7	56
22	3	59

$\boxed{N = \Sigma f = 59}$

$\boxed{N = 59 \text{ odd}}$

$M = \left(\frac{N+1}{2}\right)th$ item

$M = \frac{60}{2}$ $\boxed{M = 30}$

$\boxed{M = 16}$

Continue series

	f	cf
0 – 5	6	6
5 – 10	12	18
10 – 15	17	35
15 – 20	30	65
20 – 25	10	75
25 – 30	10	85
30 – 35	8	93
35 – 40	5	98
40 – 45	2	100

≈ 50

$\boxed{\Sigma f = N = 100}$

$$\boxed{M = l + \frac{\left(\frac{N}{2} - cf\right)}{f} \times i}$$

l = lower limit of median class

cf = Cumulative fre of class

f = Frequency of median class

i = Size of class interval

Calculate the median from the following data:

Daily wages :	210	160	200	150	170	180	230	220	140	190
No. of employee :	15	20	18	35	27	23	13	12	18	19

ans ⇒ firstly arrange it into ~~ascending order~~ ascending order

X	f	cf
140	18	18
150	35	53
160	20	73
170	27	100
180	23	123
190	19	142
200	18	160
210	15	175
220	12	187
230	13	200

$N = \Sigma f = 200$

for even

Average of $\left(\frac{N}{2}\right)$th and $\left(\frac{N}{2}+1\right)$th item

$\left(\frac{200}{2}\right) = 100 \quad \left(\frac{200}{2}+1\right) = 101$

$= \frac{170+180}{2} =$

$\boxed{M = 175}$

for → continue series

* Calculate the median from the following data:

values	frequency
less than 10	4
" 20	16
" 30	40
" 40	76
" 50	96
" 60	112
" 70	120
" 80	125

ans ⇒

Interval	value	Frequency cf	f
0 – 10	less than 10	4 →	4
10 – 20	"	16	12
20 – 30	"	40 cf	24
30 – 40	"	76	36 f
40 – 50	"	96	20
50 – 60	"	112	16
60 – 70	"	120	8
70 – 80	"	125	5

N = 125

$$M = l_1 + \left(\frac{\frac{N}{2} - cf}{f}\right)$$

$\frac{N}{2} = \frac{125}{2} = 62.5$

$l_1 = 30$

$cf = 40$

$f = 36$

$i = 10 - 0 = 10$ $\boxed{i = 10}$

$$M = 30 + \left(\frac{62.5 - 40}{36}\right) \times 10$$

$$M = 30 + \left(\frac{62.5 - 40}{36}\right) \times 10 \quad M = 30 + \left(\frac{22.5}{36} \times 10\right)$$

$$M = 30 + 6.25 \quad \boxed{M = 36.25}$$

* Calculate the median from the following data:

value : 1-10 11-20 21-30 31-40 41-50

Frequency : 4 12 20 9 5

interval	frequency	cf
0.5 – 10.5	4	4
10.5 – 20.5	12	16 cf
20.5 – 30.5	20	36
30.5 – 40.5	9	45
40.5 – 50.5	5	50
	~~N = 50~~	N = 50

$\frac{N}{2} = 25$

$l_1 = 20.5$

$f = 20$

$cf = 16$

$i = 10.5 - 0.5$

$\boxed{i = 10}$

$$M = l_1 + \left(\frac{\frac{N}{2} - cf}{f}\right) \times i$$

$$M = 20.5 + \left(\frac{25 - 16}{20}\right) \times 10$$

$$M = 20.5 + \frac{9}{\cancel{20}_{2}} \times \cancel{10}$$

$$M = 20.5 + 4.5$$

$$\boxed{M = 25}$$

Meadian class = $\left(\frac{N}{2}\right)$ item = $\frac{100}{2}$ = 50th item

Meadian class = 15-20 = 5 = i

cf = 35

$$M = 15 + \left(\frac{50-35}{30}\right) \times 5 \Rightarrow \boxed{M = 17.5}$$

~~Inclusive series~~ inclusive series

↓

Mid value series

M.	F	X
m	15	0-10
5	7	10-20
15	11	20-30
25	10	30-40
35	13	40-50
45	8	50-60
55	20	60-70
65	16	70-80
75		

$L = m - \frac{i}{2}$

$M = m + \frac{i}{2}$

$= 5 + \frac{10}{2} = 10$

$\boxed{M = 10}$

* find the missing frequency

Marks	f	cf
0-10	10	10
10-20	f1	10+f1
20-30	25	35+f1
30-40	30	65+f1
40-50	f2	65+f1+f2
50-60	10	75+f1+f2
N = 100		

Cal or

N = 100

M = 30

f1+f2+75 = 100

f1+f2 = 25

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

$$\cancel{30} = \cancel{30} + \left(\frac{50-(35+f1)}{\cancel{100}\ 10}\right)\times 10$$

given
$M = 30$
$L1 = 30$
$f =$

$10 = 50 - 35 - f1$

$10 = 15 - f1$

$f1 = 15 - 10$

$\boxed{f1 = 5}$

$f1 + f2 + 75 = 100$

$f2 = 100 - 75 - f1$

$f2 = 100 - 75 - 5$

$\boxed{f2 = 20}$

Mode

Mode is the ~~important measurement of~~ central tendancy it is the value which occure more frequently

8, 8, 12, 8, 12, 8, 4, 5 individual series

X	f
4	1
5	1
8	4
12	2

Discrete series mode: (i) inspection method
(ii) Grouping method (PPT 69 sem)

110	120	130	140	150	160
2	4	8	10	5	4

$\boxed{\text{Mode} = 140}$

~~Grouping Method~~ Continuous series

$$\boxed{Z = l_1 + \frac{f_1 - f_0}{2f_1 - (f_0 - f_2)} \times i}$$

Z is the Mode

l_1 = lower ~~middle~~ limit of the ~~lower~~ Modal class

f_0 → frequency of the pre mode of class

f_1 → frequency of the ~~post~~ in front of mode of class

f_2 → frequency of the post mode of class

$\boxed{\text{Mode} = 3\,\text{median} - 2\bar{X}}$ 2 marks

X	f		
0-5	3		
5-10	7		
10-15	15	f_0	
15-20	30	f_1	
20-25	20	f_2	
25-30	10		
30-35	5		

max frequency

Z = $l_1 = 15$

$f_1 = 30$

$f_0 = 15$

$f_2 = 20$

$i = 5$

Calculate the mode from the following data

Ans ⇒

$$Z = \left(15 + \frac{30 - 15}{(2 \times 30 - 15 - 20)}\right) \times 5$$

$$Z = 15 + \frac{15}{(60 - 35)} \times 5$$

$$Z = 15 + \frac{15}{25} \times 5$$

$$\boxed{Z = 18}$$

Marks	cf	f
10 & 15	4	4
10 & 20	12	8
10 & 25	30	18
10 & 30	60	30
10 & 35	80	20
10 & 40	90	10
10 & 45	95	5
10 & 50	97	2

27.7
cm →

* Geomitric Mean : It is nth root of the product of all the n values of the variable.

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

GM of 4 & 9.

$$GM = (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}}$$

taking log both side

$$\log GM = \log (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}}$$

$$\log GM = \frac{1}{n} \log [x_1 \cdot x_2 \cdots x_n]$$

$$\log GM = \frac{1}{n} [\log x_1 + \log x_2 \cdots \log x_n]$$

$$\log GM = \text{antilog}\left(\frac{\Sigma(\log x)}{n}\right)$$

this is for individual

X	log X
180	2.2553
190	2.28
240	2.38
386	2.58
492	2.69
662	2.82
	sum =

sum = $\frac{\quad}{n} = \frac{\quad}{6}$ after that

antilog = antilog( ) =

Geometric Mean

Measure of Dispersion : Spread or variation

Central tendency : How to find average and it gives central value and How the value revolve around a center but it doesn't show the correct picture of variability dispersition of data, dispersion mean distance from the average.

49, 50, 51 → $\bar{x}_1$ $\qquad \bar{x}_1 = \frac{\Sigma x}{N} = \frac{1050}{3} = 50$

10, 60, 80 → $\bar{x}_2$ $\qquad \bar{x}_2 = \frac{150}{3} = \boxed{\bar{x}_2 = 50}$

variability means how the Aborsution of the set are scerated around the average.

dispersion means the extend of uniformity and degree of variation, more the uniformity less is the degree of dispersion.

Absolute value : exact value.

Relative value : with respect to another value.

Dispersion

- Range
- QD Quartile Deviation
- Inter Quartile Range
- mean Deviation
- Standard Deviation

Range : Range is the difference b/w ~~lower height~~ Hiegest and lowest value in the series.

Individual series

Find the range and coefficient range from the following series

~~3, 8, 21, 51, 61,~~ 3, 8, 2, 5, 6, 9, 10

Range = 10 − 2 = 8

coef range = $\frac{10-2}{10+2} = \frac{8}{12} = \frac{2}{3}$

discrete series

X	f
10	3
20	5
30	2
40	6
50	3

range = 50 − 10 = 40

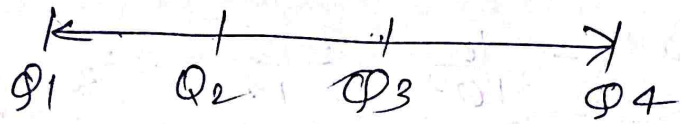
coe range = $\frac{50-10}{50+10} = \frac{40}{60} = \frac{2}{3}$

continue series

X	f
10 − 20	6
20 − 30	3
30 − 40	8
40 − 50	5
50 − 60	6

* IQR → Inter Quartile Range: It is diffe. b/w upper Q & lower quartile.

$$\boxed{Q = Q_{up} - Q_{low}}$$



* Quartile deviation: ~~Hand~~ Half

$$\boxed{Q_D = \frac{Q_3 - Q_1}{2}}$$

* Coeficient of the Quartile deviation:

$$\boxed{Coef = \frac{Q_3 - Q_1}{Q_3 + Q_1}}$$

Q find Q1 & Q3 of the following data

3, 8, 5, 2, 6, 9, 4, 10

before → arrange ascending order

→ 2, 3, 4, 5, 6, 8, 9, 10 (1 2 3 4 5 6 7 8)

$$\boxed{Q_1 = \left(\frac{N+1}{4}\right)^{th}} \quad \boxed{Q_2 = 2\left(\frac{N+1}{4}\right)^{th}} \quad \boxed{Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{ item}}$$

$$\boxed{Q_4 = 4\left(\frac{N+1}{4}\right)^{th} \text{ item}}$$

$$Q_1 = \left(\frac{8+1}{4}\right) = 2.25 \text{ th item}$$

$$\boxed{2\text{th item} + 0.25 \text{ item} (3\text{rd item} - 2\text{th item})}$$

$$2 + 0.25(4 - 3)$$

$$2 + 0.25(1) = \boxed{3.25}$$

$Q_3 = 3\left(\frac{N+1}{2}\right)th$

$Q = 3\left(\frac{8+1}{2}\right) = 3\left(\frac{9}{2}\right) = 3 \times 2.25 = \underline{6.75}$

~~[illegible]~~

or $\Rightarrow$ 6th + 0.75 ~~[illegible]~~ (7th − 6th)

or $\rightarrow$ 8 + 0.75 (9 − 8) = 8.75

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

$$Q\ Coe = \frac{Q_3 - Q_1}{Q_3 + Q_1} =$$

Assignment:

① Define statistics ② Application of Stat (5) ③ Differ primary data & Secondary data (8 points).

* Mean deviation/~~Calendar~~ Average deviation: It is defined as the arithemetic Average of deviation of various items of the series computed from sum measures of central tendency.

Mean deviation from the mean will be

$$\boxed{M.D_M = \frac{\Sigma|x - M|}{N}}$$

$$\boxed{M.D_{\bar{x}} = \frac{\Sigma|x - \bar{x}|}{N}}$$

Coefficient of mean deviation =

$$\boxed{Coeff\ M.D_{\bar{x}} = \frac{\text{mean Deviation}}{\text{Mean}}}$$

$$\boxed{Coeff\ M.D_M = \frac{\text{Median Deviation}}{\text{Median}}}$$

for Discrete

$\frac{\Sigma f|x - M|}{N}$ and $\frac{\Sigma f|x - \bar{x}|}{N}$

calculate the mean deviation from the mean of median and coeff of mean deviation for following data.

x	$\lvert x-\bar{x}\rvert$
20	$\lvert 20-45\rvert = 25$
22	$\lvert 22-45\rvert = 23$
25	20
38	7
40	5
50	5
65	15
70	20
75	25

$$M.D_{\bar{x}} = \frac{\Sigma\lvert x-\bar{x}\rvert}{N}$$

$$= \Sigma\lvert$$

$$MD_M = \frac{\Sigma\lvert x-M\rvert}{N}$$

ans=

$\bar{x} = \frac{\Sigma x}{N}$ $\bar{x} = 45$

* Standard deviation (σ): variance $= (\sigma)^2$

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

First used by Karl Pearson in 1893, it is also known as root mean square deviation, it is defined as the square root of the arithmetic mean of square of deviation of the values taken from the mean.

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

$$\boxed{\text{Coefficient of } \underset{(\sigma)}{S.d} = \frac{\sigma}{\bar{x}}}$$

Std. deviat^n
By direct method $\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$ $\hookrightarrow \bar{x}$

* Difference b/w Mean deviation and standard deviation

~~+ve any~~	~~+ve, -ve~~

Q calculate the standard deviation from following series

X	$\lvert X-\bar{X}\rvert$	$(X-\bar{X})^2$
10	-20	400
20	-10	100
30	0	0
40	10	100
50	20	400

$\bar{X}=30$ $\Sigma =$

$\boxed{\bar{X}=\frac{\Sigma X}{N}}$

$\sigma = \sqrt{\frac{1000}{5}} = \sqrt{200} =$

$\sigma = 14.14$

* Short cut method : $\boxed{\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}}$

where $\boxed{d = X - A}$

$d =$

$A = 30$

X	$X-A=d$	d^2
10		
20		
30		
40		
50		

$\sigma = \sqrt{200}$

* Step deviation method $\sigma = \sqrt{\left(\frac{\Sigma d'^2}{N}\right) - \left(\frac{\Sigma d}{N}\right)^2} \times i$ $d' = \frac{d}{i}$

~~for continuous series~~

* Discrete Series

* Calculate the standard deviation from the following data

X	f	X−A = d	fd²
3	7	−3	63
4	8	−2	32
5	10	−1	10
6	12	0	0
7	4	1	4
8	3	2	12
9	2	3	18

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

$A = 6$

$\boxed{N = \Sigma f = 46}$

fd'
−21
−16
−10
0
4
6
6

$\boxed{\Sigma fd' = -31}$

$\Sigma fd^2 = 139$ $\quad i = 1$

$$\frac{\Sigma fd^2}{N} = \frac{139}{46} = 3.02$$

$$= \left(\frac{\Sigma fd}{N}\right)^2 = \left(\frac{-31}{46}\right)^2 = \left(\frac{961}{2116}\right) = 0.45$$

$$\sigma = \sqrt{3.02 - 0.45}$$

$$\sigma = \sqrt{2.570}$$

$$\boxed{\sigma = 1.603}$$

Q calculate Standard deviation (σ) & Mean deviation & coeff of deviation

X	f
0 – 10	15
10 – 20	32
20 – 30	51
30 – 40	78
40 – 50	97
50 – 60	109

Question [illegible] है, तो coeff [illegible] है। don't worry

$$\text{coeff of deviation} = \frac{\sigma}{\bar{X}} \times 100$$

ans = 15.48

Sem

Use step deviation method

$$\sigma = \sqrt{\left(\frac{\Sigma fd'^2}{N}\right) - \left(\frac{\Sigma fd'}{N}\right)^2} \times i \quad = \boxed{d = \frac{d}{i}}$$

where $\boxed{d = X - A}$

calculate $\sigma = ?$

$\bar{X} = ?$

coeff of deviation = ?

PYQ 4M The mean and standard deviation of 200 items are found to be 60 and 20 respectively. if ~~at the time~~ of calculations, two items were wrongly taken as 3 and 67 instead of 13 and 17, detect the correct mean and standard deviation. verify the correct coefficient of variation.

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2} = \sqrt{\frac{\sum x^2}{200} - (60)^2} = 20$$

$\bar{x}$

calculating incorrect $\sum x^2$

$$\Rightarrow \frac{\sum x^2}{200} - 3600 = 400$$

$$\Rightarrow \sum x^2 = 4000 \times 200$$

$$\Rightarrow \sum x^2 = 8 \times 10^5$$

Incorrect	Correct
3	13
67	17

$$\Rightarrow \sum x^2 = 8 \times 10^5 - (3)^2 - (67)^2 + (13)^2 + (17)^2 \Rightarrow \text{correct } \sum x^2$$

$$\Rightarrow 8 \times 10^5 - 9 - 4489 + 169 + 289$$

$$\Rightarrow 795,960$$

$x_1 = 2$

$x_2 = 3 \quad 7$

$x_3 = 4$

$\sum x^2 = 2^2 + 3^2 + 4^2 = 50$

$2^2 + \cancel{3^2} + 4^2 - \cancel{3^2} + 7^2 =$

$$\boxed{\frac{\sum x}{n} = \bar{x}}$$

Incorrect $\sum x = 60 \times 200$

correct $\sum x = 12000 - 3 - 67 + 13 + 17$

$\sum x = 11960$

correct $\bar{x} = \frac{11960}{200} = 59.8$

$$CV = \frac{20.09}{59.8} \times 100$$

33.59%

$$\text{correct } \sigma = \sqrt{\frac{795,960}{200} - (59.8)^2} = \sqrt{3979.8 - 3576.04}$$

$$\sigma = \sqrt{403.8} = 20.09$$

* Skewness and kurtosis

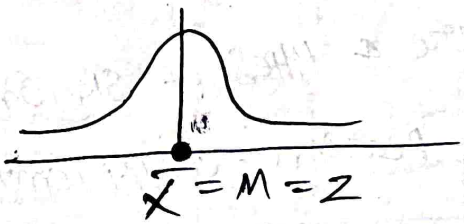
* Skewness : When a series is not symmetrical, it is said to be asymmetrical or skewed. Skewness refers to asymmetry or lack of symmetry in the shape of frequency.

* distribution : A distribution is said to be skewed, when mean, median falls at different points in the distribution.

Distribution → symmetry
Distribution → Asymmetry

* symmetry distribution : No skewness is present, Mean, Median and mode are ~~win~~ coincides (equal).

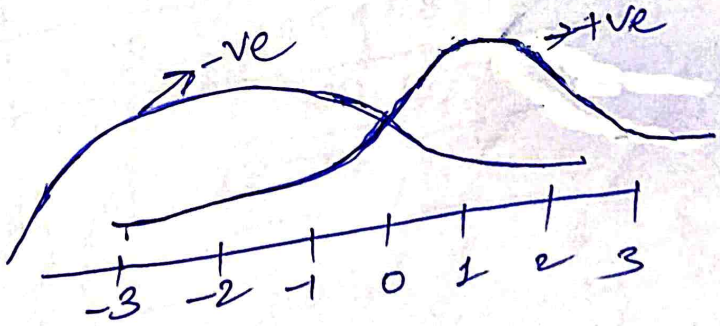
$$\boxed{\bar{X} = M = Z}$$



* Asymmetry distribution : When a distribution is not symmetrical, then it is skewed or Asymmetry distⁿ.

There are 2 types of skewness.

① +vely skewed ② Negatively skewed.



* diff b/w dispersion and skewness.

dispersion	skewness
(i) Dispersio is concerned with the amount of variation	(ii) skewness is concerned with direction of variation.
(ii) It focus on spread of data.	It focus on symmetry of data.
(iii) It helps to judge the reliability of central tendency.	(iii) It helps to identify deviation from normality.
(iv) Example: Range, variance, IQR	Ex ⇒ Skewness coefficient.

when $\bar{x} > z \rightarrow$ (+vely skewed)

$\bar{x} < z \rightarrow$ (−vely skewed)

There are 4 types of skewness (in Marks)

(i) karl pearson coefficient of skewness:

$$\boxed{\text{Karl C.O.S} = \frac{\bar{x} - z}{\text{std. deviation}}} \quad \boxed{\text{Karl C.O.S} = \frac{\bar{x} - z}{\sigma}}$$

(PYQ) Generate Karl Pearson's Coefficient of skewness from the following data.

profit RS·L	70-80	80-90	90-100	100-110	110-120	120-130	130-140	140-150
No of cos	12	18	35	42	50	45	30	8

∵ Karl co·s = $\frac{\bar{X} - Z}{\sigma}$, $\bar{X} = A + \frac{\Sigma fd}{N}$ ∴ $d = X - A$

$Z = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times i$

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

$d = x - A$, $d' = \frac{d}{i}$

profit	f	d = X-A	X	fd	d'	d'²	fd'	fd'²
70-80	12	-35	75	-420	-3·5	12·25	-42	159·25
80-90	18	-25	85	-450	-2·5	6·25	-45	112·5
90-100	35	-15	95	-525	-1·5	2·25	-52·5	78·75
100-110	42	-5	105	-210	-0·5	0·25	-21	10·5
110-120	50	5	115	250	0·5	0·25	25	12·5
120-130	45	15	125	675	1·5	2·25	67·5	101·25
130-140	30	25	135	750	2·5	6·25	75	187·5
140-150	8	35	145	280	3·5	12·25	28	98

$N = 880$ $\Sigma fd = 350$ $L = 110$ $f_0 = 42$ $f_1 = 50$ $f_2 = 45$

$A = 110$ $i = 10$ $\Sigma fd' = 35$ $\Sigma f \cdot d'^2 = 760 \cdot 25$

$\bar{X} = 110 + \left(\frac{350}{880}\right)$

$\bar{X} = 110 \cdot 3977$

$Z = 110 + \left(\frac{50 - 42}{2 \times 50 - 42 - 45}\right) \times 10$

$Z = 110 + \left(\frac{8}{13}\right) \times 10 \Rightarrow Z = 110 + 6 \cdot 15$ $Z = 116 \cdot 15$

$$\boxed{\frac{\Sigma fd'}{N} = \frac{35}{880} = 0.039}$$

$$\boxed{\left(\frac{\Sigma fd'}{N}\right)^2 = 0.001521}$$

$$\frac{\Sigma f \cdot d'^2}{N} = \frac{760.25}{880} = 0.8639$$

$$\sigma = \sqrt{0.8639 - 0.001521}$$

$$\sigma = \sqrt{0.86}$$

$$\boxed{\sigma = 0.92}$$

$$\text{Karl P.C.O.S} = \frac{\bar{X} - Z}{\sigma} = \frac{110.39 - 116.15}{0.92}$$

$$= \frac{-5.76}{0.92} = -6.26$$

$$\boxed{\text{Karl P.C.O.S} = -6.26}$$

* find Karl Pearson Coefficient of Skewness

X	58	59	60	61	62	63
f	10	18	30	42	35	38

Mode(Z) = 61

$\bar{X} = 61.08$

$\sigma = -0.045$

Ans ⇒

X	f	fX
58	10	580
59	18	1062
60	30	1800
61	42	2562
62	35	2170
63	38	2394

$N = 173$

$\Sigma fx = 10568$

$\boxed{z = 61}$

$\boxed{\bar{x} = 61.08}$

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

Unit 2 ⇒ Sampling Distribution & Testing of Hypothesis

* **Population:** Population means entire field under investigative about which knowledge is sort.

Ex ⇒ if we want to collect information about the monthly expenditure of 2000 students of a college then the entire agricate of a two thousand student will be termed as universe or population.

population is of two type

① finite population ② Infinite population.

* **Sample:** Sample is a part of population.

(PYQ) (4M) What is sampling? different ~~types of sampling~~ b/w probability and ~~Non-probability~~ Non-probability sampling.

Sampling is a method of select a sample from a given population is known as sampling.

probability sampling	Non-probability sampling
(i) probability sampling is a method in ~~which~~ which there is equal opportunity to be selected as a representative.	Non-probability sampling is a method in which there is no equal opportunity to be selected as representative.
(ii) It is Also known as random sampling.	Also known as non-random sampling.
(iii) In this selection is random.	In this selection is arbitary
(iv) It is objective method.	It is subjective method.
(v) Hypothesis is ~~Jemo~~ tested.	Hypothesis is generated.
(vi) Produces unbiased result.	produces biased result.

Sampling

probability sampling
1. simple Random sampling
2. straitified " sampling
3. systematic " sampling
4. cluster " sampling

Non-probability sampling
1. convenience sampling
2. Judgement sampling
3. Quota sampling
4. snowball sampling

① Simple Random sampling : Every member of the population has an equal chance of selection.

Ex → lottery method / Ludo method / Spin method | for example 50 people are playing lottery so 50 lottery ticket are colleted in bucket and you are selecting randomly on ticket so any one can selected in b/w 50.

~~(ii) stratified random sampling~~

Merit
- ① free from personal buyers
- (ii) Save time money or labour.

demerit
- Time consuming
- Not suitable for certain populations.

② Stratified Random sampling : In this sampling population is devided into sub groups (called strata) like : (gender, age, range, job role)

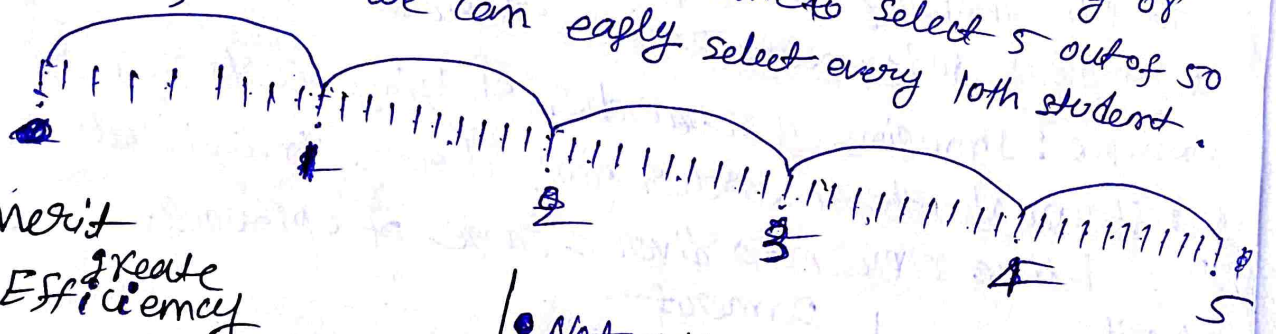
Example:
for example you have to select 50 workers for industry so you can make strata and select randomly from each group like 25 from gender group, 25 from age group.

Merit
- (i) More Accuracy
- (ii) Increased precision

Demerit
- Complexity
- Need for information.

③ Systemetic random sampling: In this sampling population is arranged in a particular order either in ascending or a descending.

Example: If we want to select 5 students from 50 the we systemeticly arrange this in either ascending or descending order and since I have to select 5 out of 50 so $\frac{50}{5} = 10$ we can eagly select every 10th student.



Merit	
• Efficiency (create)	• Not suitable for small population
• Representativeness	• if order is not maintained may be biased.

④ Cluster (area) random sampling: In this sampling population is devided into subgroups based upon area like (city blocks)

Example: Consider A city devided into neighbourhoods; you randomly choose a few neighbourhoods in your sample.

Merit	Demerit
Efficiency	• Analysis complexity
practicality	• logistical challenges.

① Convenieme sampling: In this sampling researcher selects the ~~most~~ accessible population members

Example: For example your company in ludhiana so you will select labourers in ludhiana itself.

Merit	Demerit
• More convenient sampling	• Not representative
• cost-effective	• Not accurately when large population
• saving time and resources	

② ~~Ju~~ Judgement sampling: ~~The researcher selects population where good~~.

In this sampling selecting sample based on the researcher's judgement or expertise.

Example: Imagine a researcher studying customer in a restaurant, the researcher may choose individuals they believe represent a diverse range of opinions.

Merit	Demerit
• Efficiency	• Bias risk
• Expertise	• Difficulty

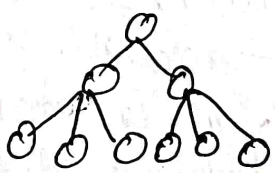
③ Quota sampling: In this sampling when researchers select participants based on certain quota.

Example: Consider survey on smartphone in a city you might set quotas for age groups.

Merit	Demerit
• Cost-effective	• Bias Risk
• Quick implementation	• Difficulty in Quota Setting.

④ Snowball sampling: In this sampling researcher collect data from a participant and asked for other participants and again collect data from refered ~~participant~~ participants like chain.

Merit
- Cost-effective
- More Accurate



Demerit
(i) Bias
(ii) lack of control.

(Pyq) diff b/w Sampling error and non sampling error

Sampling error	non-sampling error
① Sampling error due to chance variations in the sample.	non-sampling error due to random variation in the sample.
② It cause → Random variation in the sample.	cause → Bias in the data collection.
③ due to wrong method of sampling	③ due to human factor.
④ It can be controlled by ~~reduced~~ increasing sample size	④ It can be controlled by careful study.
⑤ Ex ⇒ The difference between the average height in the sample and the true average height of all students is the sampling error	⑤ Ex ⇒ Imagine a survey conducted over the phone where some individuals do not have access to phone. This introduces a non-sampling error.

* Population Mean → μ	Sample Mean → $\bar{x}$
* Standard Deviation → σ	Sample SD → s
* Variance → σ^2	Sample variance → s^2
* population propotion → P	Sample population propotion → p

Merit

* Test of hypothesis: we have to make decision about population on the basis of sample

Ex ⇒ A drug chemist has to decide whether a new drug is effective or not curing a disease or not

Such decisions are known as ~~statistical~~ statistical decision.

Hypothesis: It is a claim that need to be tested. An attempt to arive at a decision about the population on the basis of sample information

It is necessary to make assumption about the parameter first. Such of assumption are called statistical Hypothesis which may or may not be true.

* There are two types of Hypothesis.

(i) Null Hypothesis (ii) Alternate Hypothesis

(i) Null Hypothesis: It says that there is no diff b/w sample stats ($\bar{x}$) and population parameter (μ). [$\bar{x} = \mu$]

What are the absorbed diff due to fluctuation ~~fluction~~ in sampling from same population.

Ex = $\bar{x} = \mu$ $\mu = 50$

Alternate Hy: Any Hypothesis is called alternet Hypothesis (H_a) different from null Hy

It is denoted by H_1 या H_a Ex → $\mu 1 \Rightarrow \mu \neq 50$

$\mu > 50$ या $\mu < 50$

Sample.

* Type 1 error and type 2 error difference b/w.

Type 1 error	Type 2 error
It is made when we reject the Null Hypothesis although it was true.	It is made when we accept the Null Hypothesis although it was false.
(ii) the probability of making this error is α: $P(E_1) \rightarrow \alpha$	(ii) The probability of making this error is β $P(E_2) \rightarrow \beta$
(iii) The probability of making Type 1 correction is $1-\alpha$	(iii) The probability of making Type 2 correction is $1-\beta$
(iv) Type 1 error is called a false positive.	(iv) A type 2 error is called a false negative.

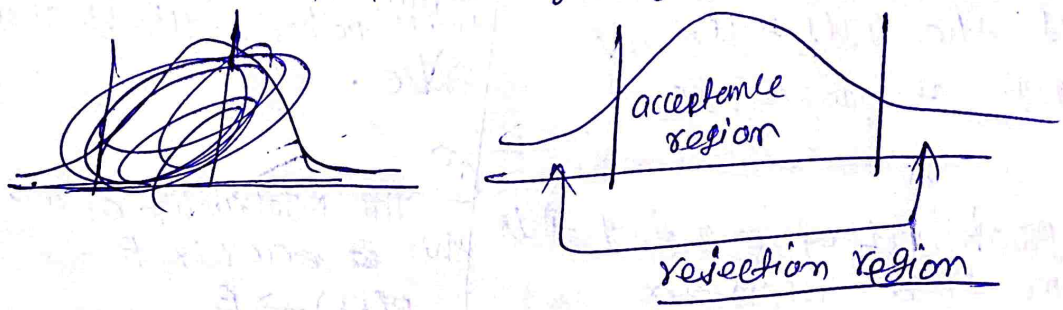
(2 marks)

* Level of significance: The degree of significance which we accept or reject Hypothesis is known as level of significance.

100% Accuracy is not possible in making a decision so we fix the confidence (error) at 5% for a greater precision that means there is 5% chance of error. If we fix confidence at 1% level the decision is correct to 99%. If no level of significance is given we always take level of confidence 0.05 or 5% (α).

* Critical region / rejection region :

The stats which leads to rejection of Null Hypothesis give us region is known as rejection region.



* Test of Hypothesis :

if $n \geq 30 \rightarrow$ large sample

(i) Test of Hypothesis about population mean.

(ii) " " " diff. b/w two population mean.

(iii) " " " Two population standard deviation

(iv) " " " population proportion.

(v) " " " about difference of two population proportion.

① Test of Hypothesis about population Mean (μ)

$$\boxed{Z = \frac{\bar{X} - \mu}{S.E_{\bar{X}}}}$$

Standard error of mean → $S.E_{\bar{X}}$

$$\boxed{S.E_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

Q the mean hight of a sample of 100 student, mean hight of random sample of 100 student is 64 inch and std deviation is is 3 inch. Test the statement the mean hight of the population is 67 inch. at 5% level of significance.

(1.64)

Ans → $\bar{X} = 64''$ · $S = 3''$ · $\mu = 67''$ $Z = 1.96$

$n = 100$

1 tail test of 5% $Z \rightarrow 1.64$

1% $Z \rightarrow 1.33$

① Null Hypothesis → there is no significant diff b/w mean hight and mean hight of population

$H_0 \rightarrow \bar{X} = \mu$

$\bar{X} = 67$

Assume $H_a = \bar{X} \neq 67$

$\boxed{\begin{array}{l} \bar{X} > 67 \\ \cancel{\bar{X} < 67} \end{array}}$ — one tail

$$SE_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{S}{\sqrt{100}} = \frac{3}{\sqrt{100}} = 0.3$$

at 5% of

Q Difference b/w primary and secondary data

primary data	secondary data
(1) primary data is a type of data researchers directly collect from main source	(1) secondary data is a type of data researchers collect already existing data produced by previous reachers.
(ii) Researcher has full control over data.	(2) Researcher has limited control
(3) It is more accurate	(3) It depends upon source and collection method.
(4) It is related to real-time data.	(4) It is related to the past.
(5) More Time Consuming	(5) less time consuming.

Q what is statistics and its importance.

Ans ⇒ statistics is the branch of mathematics for collecting, analysing and interpreting data. statistics can be used to predict the future, It determine the ~~probabith~~ probability that a specific event will happen or help answer questions about a survey.

Importance of statistics

(1) Data Analysis: It helps in analyzing and interpreting complex data.

(2) Research: It helps to researchers to draw conclusions and make decisions.

(3) Business Decision-making: Business use statistics to make informed decisions, understand market trends.

(4) Quality control: In manufacturing and industry statistics helps to ensure quality of products

Q write properties of mean.

ans ⇒ ① Sensitivity to changes: The mean is sensitive to extreme value.

② Balancing property: The sum of the deviations of individual values from the mean is always zero.

③ Affected by skewness ④ Arithmatic Mean formula

⑤ Sample vs. Population mean ⑥ unique value

⑦ Useful in Normal Distributions.

Q prove that $\Sigma(x-\bar{x})=0$

proof →

$\Rightarrow x_1-\bar{x}+x_2-\bar{x}+x_3-\bar{x}+x_4-\bar{x}+ \text{-------} +x_n-\bar{x}$ —— ①

$\because \bar{x} = \frac{\text{Sum of all observation}}{\text{no of items}}$ $\bar{x} = \frac{x_1+x_2+x_3+x_4+\text{---}x_n}{n}$

$\boxed{\bar{x}n = x_1+x_2+x_3+x_4+\text{----}x_n}$ —— ②

from eq^n ①

$x_1+x_2+x_3+x_4+\text{----}x_n - n\bar{x}$

from eq^n ②

$\cancel{n\bar{x}} - \cancel{n\bar{x}} = 0$

hence proved that $\boxed{\Sigma(x-\bar{x})=0}$

Q Explain different types of correlation

① positive correlation: when two variables move in the same direction so it is called as positive correlation. As one variable increases, the other also increases and vice versa.

Ex → As the hours of study increases, exam scores also increases

(ii) Negative correlation: Negative correlation when two variables move in opposite directions, As increase in one variable is associated with a decrese in the other and vice versa.

Ex ⇒ As the temperature decreases the number of ice cream sales decreases.

(3) Zero correlation: Zero correlation indicates no linear relationship between two variables. changes in one variable are not associated with the other.

Ex ⇒ The amount of rainfall and the number of books sold in a bookstore.

(4) (Q) difference b/w population and sample

population	sample
(1) A population is the entire collection of subjects.	A sample is subset of population.
(2) Measurement taken from a whole population are called parameters	Measurements taken from a sample are called statistics.
(3) It provides complete and accurate information	It provides an estimated result of the population.
(4) It always greater than sample.	It always smaller than population.

karl → 4
median → 4, 8,

(8) one tailed and two tailed test

one tailed test	two tailed test
(1) A one-tailed test is a statistical hypothesis test in which the alternative hypothesis specifies the direction of a difference b/w two population. $\mu = 50$	A two-taild test is a statistical hypothesis test in which the alternative hypothesis does not specify the direction of a difference between two populations. $\mu \neq 50$ $\mu < 50$ or $\mu > 50$
(2) when the alternative hypothesis specifies a direction then we use a one-tailed test.	(2) if no direction is given then we will use a two-tailed test.
(3) it is used to check wheather the one mean is different from another mean or not	(3) It is used to check whether the two mean different from one another or not
(4) It is a uni-directional hypothesis	(4) It a non directional hypothesis

Chi square test

It is an impotant test among serveral ~~test~~ test of significance, It is used for testing the significance of population variance of ~~known~~ Non parameteric test. No assumption is taken about the population, It can be used as the test of goodness of ~~fit~~ & as test of independence of Attribute.

(i) Goodness of fit
(ii) Test of independence of Attr

(1) Test of goodness of fit: We try to find out how the absorbed value of given fenomina are significantly different from the exepted value.

(2) chi square test determine how well theoritical distribution such as binomial poisson. It is always applied if sample size is greater than 50.

(Eg) the following table give the no of accident that take place in industry during ~~[illegible]~~ various days of the week. test if the accident are uniformly given in week.

Days	Mon	Tue	wed	Thus	fri	sat
No of Accident	14	18	12	11	15	14

degree of freedom → 2M → 9

H_0 : Accident are not uniformly given in week.

H_a : Accident are ~~given~~ uniformly given in week.

O	E	O − E	$(O-E)^2$	$(O-E)^2/E$
14	14	0	0	0
18	14	4	16	1.14
12	14	2	4	0.28
11	14	−3	9	0.64
15	14	1	1	0.07
14	14	0	0	0

$\frac{14+18+12+11+15+14}{6} = 14$

$\boxed{\Sigma\frac{(O-E)^2}{E} = 2.14}$

$$\boxed{\chi^2 = \frac{\Sigma(O-E)^2}{E}}$$

$\chi^2 = \frac{30}{14}$ $\boxed{\chi^2 = 2.14}$

$V = n-1 = 6-1$ $\boxed{V=5}$

Tabulated val = 11.07

at 0.05 tabulated value is = 11.07

∴ Calculated value is less than tabulated value so we accept ~~select~~ Null hypothesis hence Accident are not uniformly given in week.

Q 2. A sample analysis of examination result of 5000 students ~~were~~ were made. It was found that 220 had faild. 170 had ~~second~~ third class, 90 second class, 20 got a first class, are they figure commensurate with the ratio of 4:3:2:1

Faild → 200
Th → 170
Sec → 90
Fir → 20

O	E	O−E	$(O-E)^2$	$(O-E)^2/E$
200	200	20	400	2
170	150	20	400	2.66
90	100	−10	100	1
60	50	−30	900	18

~~200 = 5000 × 4/10~~

Faild $500 \times \frac{4}{10} = 200$

thir $500 \times \frac{3}{10} = 150$

Sec $500 \times \frac{2}{10} = 100$

fir $500 \times \frac{1}{10} = 50$

$$\chi^2 = \frac{(O-E)^2}{E} =$$

$$\chi^2 = 2 + 2.66 + 1 + 18$$

$$\chi^2 = 4.66 + 9$$

$$\boxed{\chi^2 = 13.66}$$

~~Tabulated value = 7.81~~

degree of freedom $\nu = n-1 = 4-1 = 3$

☆ tabulated value of χ^2 at 5% level of significance 3 = 7.81

since calculated value is greater than tabulated value so we reject the Ho.

Q3. Test of goodness of fit for binomial distribution,

---

(2)

Test of χ^2 of independence of attribute.

Chi square test enable us to examine wheather or not two attributes are associated or independent of of one another for example we may be interesting in knowing wheather a medicine is controlling and a fever or not.

Q A sample of 200 person with a particular disease was selected. out of them 100 were given drug and ther were not. The result were observed as follows

	Drug (c_1)	No-drug (c_2)	total (c_3)	
Guard	55	65	120	R_1
No Guard	45	35	80	R_2
Total	100	100	200	

H_0 → drug has not been effective in curing the disease

H_a → drug has been effective in curing the disease.

Let the Null hypothesis =

O	E	(O−E)	$(O-E)^2$	$(O-E)^2/E$
55	60	−5	25	0.416
65	60	5	25	0.625
45	40	5	25	0.416
35	40	−5	25	0.625

$\Sigma(O-E)^2/E = 2.082$

$$\chi^2 = \frac{(O-E)^2}{E} = 2.082$$

degree of freedom $(r-1)(c-1) \Rightarrow (2-1)(2-1)$
$= 1 \times 1 = 1$

The tabulated value of χ^2 at 5% level of frequency 1 = 3.84

Since, the calculated value $\chi^2 = 2.082$ is less than tabulated value 3.84 hence we accept the Null hypothesis and conclude that the drug has not been effective in curing the disease.

Q2. Can vaccination be regarded as preventive measure of small pox: as evidence of following data of 1482 people exposed to small pox localities 368 were attacked of those 1482 personal and 343 were vaccinated and those only 35 are attacked.

38.03 | −0.249

χ^2 obs = 48.22

Table are not given in question.

		Vaccinated	N-Vaccinated	Total
A1	Attacted	35	333	365
A2	Non-Attacted	308	806	1114
	Total	343	1139	1482

O	E	O−E	$(O-E)^2$	$(O-E^2)/E$
35	85.17	−50.17	2517.02	29.55
308	257.82	50.18	2518.03	9.76
333	282.82	50.18	2518.03	8.90
806	856.17	−50.17	2517.02	2.93
				51.14

$\chi^2 = (O-E)^2/E$ $\boxed{\chi^2 = 51.4}$

degree of freedom $= (r-1)(c-1) = (2-1)(2-1) = 1$

$V = 1$

the tabulated value of χ^2 at 5% level of frequency $1 = 3.8$

Since, the calculated value $\therefore \chi^2 = 51.4$ is greater than tabulated value hence the null hypothesis is rejected.

Q In a survey of 200 boys, of which 75 were intelligent, 40 had educated father while 85 of the unintelligent boys had uneducated fathers. Do these figures support the hypothesis that educated fathers have intelligent boys use χ^2 test.

Sol^n:

	intelligent boy	unintelligent boy	Total
Educated father	40	40	80
uneducated father	35	85	120
Total	75	125	200

H_0 = Educated fathers do not have intelligent boys

H_a: Educated fathers do have intelligent boys.

O	E	$(O-E)$	$(O-E)^2$	$(O-E)^2/E$
40	30	10	100	3.33
40	50	-10	100	2
35	45	-10	100	2.22
85	75	10	100	1.33

$\chi^2 = 8.88$

degree of freedom $v = (r-1)(c-1) = 1$

the tabulated value at 5% level of significance $t = 3.84$

since the calculated value $\chi^2 = 8.88$ is greater than tabulated value (3.84) so we reject the Null hypothesis and conclude that Educated fathers do have intelligent boys.

38.03 | -0.249

Bowley's Coeficient

$$B.C = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

Quartile for individual, ~~discrete~~ discrete and continue series

Individual Series	Discrete	Continue Series
Q_1 = size of $\left(\frac{N+1}{4}\right)$th item	Q_1 = size of $\left(\frac{N+1}{4}\right)$th item	$Q_1 = l + \frac{\left(\frac{N}{4} - cf\right)}{f} \times i$
Q_3 = size of $3\left(\frac{N+1}{4}\right)$th item	Q_3 = size of $3\left(\frac{N+1}{4}\right)$th item	$Q_3 = l + \frac{\left(\frac{3N}{4} - cf\right)}{f} \times i$

Q_1 = c of $\left(\frac{N}{4}\text{th}\right)$

Q_3 = c of $3\left(\frac{N}{4}\text{th}\right)$

No of children (x)	No of families (f)	c.f
0	7	7
1	10	17
2	16	33
3 M	25	58
4	18	76
5	11	87
6	8	95

$N = \Sigma f = 95$

Q_1 = size of $\left(\frac{N+1}{4}\right)$th item

$Q_1 = \left(\frac{95+1}{4}\right) = \frac{96}{4} = 24$

> 24

$Q_1 = 2$

$Q_3 = 3\left(\frac{95+1}{4}\right)$

$Q_3 = 3\left(\frac{96}{4}\right) = 72$

$Q_3 = 4$

Median = $\left(\frac{N+1}{2}\right)$th item $= \left(\frac{95+1}{2}\right) = \left(\frac{96}{2}\right) = 48$th item

$M = 3$, $Q_1 = 2$, $Q_3 = 4$

$$B.C = \frac{4+2-2\times 3}{4-2} = \frac{4+2-6}{2} = \frac{6-6}{2} = 0$$

X	f	cf
0 – 10	10	10
10 – 20	25	35
20 – 30	20	55
30 – 40	15	70
40 – 50	10	80
50 – 60	35	115
60 – 70	25	140
70 – 80	10	150
	N = 150	

$Q_1 = \text{size of } \left(\frac{N}{4}\right)\text{th item}$

$Q_1 = \frac{150}{4} = 37.5$

$Q_1 = 20 + \frac{\left(\frac{150}{4} - 35\right)}{20} \times 10$

M

$Q_1 = 21.25$

$Q_3 = 3\left(\frac{N}{4}\right)\text{th item} \Rightarrow Q_3 = 3 \times \left(\frac{150}{4}\right) = Q_3 = 112.5$

$Q_3 = 50 + \frac{\left(3\left(\frac{150}{4}\right) - 80\right)}{35} \times 10$

$Q_3 = 50 + \left(\frac{112.5 - 80}{35}\right) \times 10$

$Q_3 = 50 + 9.28$

$Q = 59.28$

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

$M = \text{size of } \left(\frac{N}{2}\right)\text{th item} = \frac{150}{2} = 75$

$M = 40 + \left(\frac{75 - 70}{10}\right) \times 10 = 40 + \frac{5}{10} \times 10$

$M = 45$

$B.C = \frac{Q_1 + Q_3 - 2M}{Q_3 - Q_1}$

$B.C = \frac{21.25 + 59.28 - 2 \times 45}{59.28 - 21.25}$

$B.C = \frac{-9.47}{38.03}$

$B.C = -0.249$