STATISTICS METHODS

AND

[redacted] OPERATIONS

(ALLIED PAPER II)

STATISTICS AND OPERATION RESEARCH

Unit 1 : ~~(Chap 4 in that book)~~ [chap. 4 in text book]

Measure of Central tendancy.

Arithmatic mean, Median, Mode

Geometric Mean, Median, Mode

Arithmetic Mean, combined mean,

Weighted Mean

Unit 2 :

Measure of despertion : [chap 5 in text book]

* Range
* Quartile deviation.
* Mean deviation
* standard deviation.
* Combined standard deviation.
* Co-efficent of Variation
measure of skewness.
* Karl Pearson's Method,
* Bauley's coefficient of skewness.

Unit 3 [Chap 7 in text book]

Correlation :

Karl ~~prson~~ person's method

Spearmon's Method

Regulation

Lines of regulation

Unit 4

Sample space : [chap. 13 in text book]

→ Events

→ Probability

→ Add & multiplication theorem

→ Conditional Probability

→ Bayer's Theorem.

Unit 5 [chapt. 9 in text book]

Time Series

- → Meaning and definition.
- → Uses
- → Components of time series
- → Method of measuring trend.

Textbook:

Business Statistics and Operation Research.
Dr. S.P. Rajagopalan, R. Sattanathan.

Scanned by CamScanner

UNIT 4

MEASURMENT OF CENTRAL TENDANCY.

~~Definition of statistics:~~ MEAN

1. Arithmatic Mean * $A.M = \Sigma x/n$.

   * $A.M = A + \frac{\Sigma d}{n}$.

   * $A.M = \frac{\Sigma fx}{\Sigma f}$

   * $A.M = A + \frac{\Sigma fx}{\Sigma f}$.

A → Assumed Mean.

f → frequency. x → data

* If ~~the~~ n values are given, say $x_1, x_2, \ldots$ etc.,
  then $A.M = \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\Sigma x}{n}$.

* If the values of observation are large, then you can use assumed mean method than arithmatic mean.

* If the frequency distribution are given for the corresponding observation then arithmatic mean is

$$A.M = \frac{\Sigma fx}{\Sigma f}$$

or $= A + \frac{\Sigma fd}{\Sigma f}$.

PROPERTIES OF A GOOD AVERAGE :

* It should be easy to compute and understand.
* It should be based on all items.
* It should not be affected by external observations.
* It should be rigidly defined.
* It should be capable of future mathematical treatement.

Step deviation Method :

This method is used in case of continuous distribution. The formula for calculating the arithmatic mean is given by. ,

$$\bar{x} = A + \frac{\Sigma fd}{\Sigma f} \times i.$$

A → Assumed mean
i → length of order.
x → midvalue of C.I.
d =

where $d = \frac{x - A}{i}$

Combinational Arithmatic Mean :

Combinational mean of two groups is given by

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

where $\bar{x}_1$ = mean of first group
$\bar{x}_2$ = mean of second group
n_1 = no. of observations in group 1
n_2 = no. of obrervation of grp

For many groups,

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + \dots + n_n\bar{x}_n}{n_1 + n_2 + n_3 + \dots + n}$$

~~PROBL~~

PROBLEM:

(1). Find the arithmatic mean of the following data. If 3 is added to each data, then find the new A.M.

12, 15, 10, 9, 11, 14, 6.

$\bar{x}_1 = \frac{\sum x}{n}$ $\sum x \Rightarrow$ Summation of all data.

$n \Rightarrow$ no. of observation.

$= \frac{12+15+10+9+11+14+6}{7}$ $\bar{x} = \frac{\sum x}{n} =$

$\sum x = 77$ $n = 7$

$= \frac{77}{7} = 11.$ $\bar{x} = 11+3$

$= 14.$

If 3 is added to each data, then $\bar{x} = \bar{x}_1 + \frac{3n}{n}$

$\cancel{\bar{x} = 13 \frac{\sum x}{n}} = \cancel{\frac{3.77}{7}} \cancel{3.11} = 11 + \frac{3.\cancel{7}}{\cancel{7}}$

$\cancel{\bar{x} = 3n + \frac{\sum x}{n} = (3 \times 7) + \frac{77}{7}} = 11+3 = 14.$

(2) The following table given the marks obtained by 10 students in a class. Calculate the arithmatic mean.

Roll. no.	1	2	3	4.	5	6	7	8	9	10
Marks.	40	50	30	60	70	80	40	80	60	90

$\bar{x} = \frac{\sum x}{n} = \frac{40+50+30+60+70+80+40+80+60+90}{10}$

$= \frac{570}{10} = 57.$

③. From the following table, find the mean height. (5)

Height.	60	61	62	63	64
No. of children	2	3	5	8	7

let $x \to$ heights no $f \to$ no. of children

x	f	xf
60	2	120
61	3	183
62	5	310
63	8	504
64	7	448
Σ	25	1565

$\Sigma xf = 1545$ $\Sigma f = 25$.

Then $\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{1565}{25} = 62.6$.

④. From the following data of marks obtained by 60 students of class calculate the arithmetic mean.

Marks (x).	20	30	40	50	60	70.
No. of student (f)	8	12	20	10	6	4

x	f	d $\boxed{d = x - A}$	fd.
20	8	−20	−160
30	12	−10	−120
A ← 40	20	0	0
50	10	10	100
60	6	20	120
70	4	30	120.

$\boxed{A = 40.}$

$\boxed{\Sigma f = 60}$ $\boxed{\Sigma fd = 60}$

Arithmatic Mean

$$\bar{x} = A + \frac{\Sigma fd}{\Sigma f} = 40 + \frac{60}{60} = 40 + 1 = 41,$$

⑤. The following is the age distribution of 100 people in a street. Calculate the arithmetic mean.

Age group. (x)	0-10	10-20	20-30	30-40	40-50	50-60
No. of person.(f)	5	10	25	30	20	10

x	f	mid x	d	fd.
0-10	5	5	-20	-100
10-20	10	15	-10	-100
20-30	25	$\boxed{25}$ (A)	0	0
30-40	30	35	10	300
40-50	20	45	20	400
50-60	10	55	30	300
	$\boxed{100}$ $\boxed{\Sigma f}$			$\boxed{800}$ $\boxed{\Sigma fd}$

Arithmetic mean $\bar{x} = A + \frac{\Sigma fd}{\Sigma f}$

$= 25 + \frac{800}{100} = 25+8 = 33.$

$\boxed{\bar{x}=33}$

⑥. Find the missing frequency for the following distribution if the mean $\bar{x} = 12.9$.

C.I	0-5	5-10	10-15	15-20	20-25
frequency.	3	F	8	5	4.

C.I	f	mid value x	d $\boxed{x-A}$	fd.
0-5	3	2.5	-10	-30
5-10	F.	7.5	-5	-5F
10-15	8	12.5	0	0
15-20	5	17.5	5	25
20-25	4	22.5	10	40
	$\boxed{20+F}$ $\boxed{\Sigma f}$			$\boxed{5(7-F)}$ $\boxed{\Sigma fd}$

10
25
35-5F
= 5(7-F).

The mean given $\bar{x} = 12.9$.

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

$$= 12.5 + \frac{5(7-F)}{20-F}$$

$$= \frac{(12.5)(20-F) + 35 - 5F}{(20-F)}$$

$$12.9 = \frac{25 - 12.5F + 35 - 5F}{20-F} = \frac{60 - 17.5F}{20-F}$$

$$(12.9)(20-F) = 60 - 17.5F$$

$$258 - 12.9F = 60 - 17.5F$$

$$258 - 60 = 12.9F - 17.5F$$

$$198 = -4.4F$$

$$F = \frac{198}{4.4} = \frac{9}{0.2}$$

$$12.9 = 12.5 + \frac{5(7-F)}{20-F}$$

$$0.4 = \frac{5(7-F)}{20-F}$$

$$(8 + 0.4F) = 35 - 5F$$

$$8 - 35 = -0.4F - 5F$$

$$\Rightarrow -27 = -5.4F$$

$$27 = 5.4F \Rightarrow F = \frac{27}{5.4} = \frac{270}{54} = 5$$

$$F = \frac{27}{5.4}$$

$$\boxed{F = 5}$$

Ans
3) The mean height of 25 male workers in a factory is 61cm and the mean of height of ~~65~~ 35 female workers in the same factory is 58 cm. Find the combined mean height of 60 workers in the factory.

Ans

From the given data,

$n_1 = 25$ $n_2 = 35$

$\bar{x}_1 = 61$ $\bar{x}_2 = 58$.

Then the total average is,

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

$$= \frac{(25\times61) + (35\times58)}{25+35} = \frac{1525 + 2030}{60}$$

$$= \frac{3555}{60} = 59.25$$

~~4. The mean height of 25 male workers in a factory is 61cm and~~

4. In a class of 50 students, 10 have failed and their average marks is 25. The total marks secured by the entire class is 2810. Find the average marks of those who have passed.

$n_0 = 50$ students. $n_1 = 10$ $\bar{x}_1 = 25$.

$\Sigma x = 2810$. $n_2 = 50-10 = 40$. $\bar{x}_2 = ?$

$\bar{x}_1 \rightarrow$ mean of students failed.

$\bar{x}_2 \rightarrow$ Mean of students passed.

We know that

$$\bar{x} = \frac{\Sigma x}{n} = \frac{2810}{50} = \frac{281}{5} = 56.2$$

$$\boxed{\bar{x} = 56.2}$$

In other way,

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2} = \frac{(10\times 25) + (40\times\bar{x}_2)}{10+40}$$

$$\frac{281}{\cancel{5}} = \frac{250 + 40\bar{x}_2}{\cancel{50}10}$$

$$2810 = 250 + 40\bar{x}_2$$

$$2810 - 250 = 40\bar{x}_2$$

$$2560 = 40\bar{x}_2$$

$$\frac{\cancel{2560}64}{\cancel{40}} = \bar{x}_2$$

$$\boxed{\bar{x}_2 = 64}$$

2810
-250
2560

(8). The mean marks of 100 students are found to be 40. Later on it was discovered that a score of 53 was misread as 83. Find the correct mean corresponding to the correct score.

From the data,

$n_1 = 100$ $\bar{x}_1 = 40$.

Correct value = 53

Wrong value = 83

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

$$40 = \frac{\Sigma x}{100}$$

$$\boxed{4000 = \Sigma x}$$

Then the ~~correction to~~ mean sum of wrong data

$\Sigma x = 4000$

Now, the correct ~~over~~ data summation.

$CRT \ \Sigma x = 4000 + 53 - 83.$

$= 4000 - 30 = 3970.$

$\boxed{CRT \ \Sigma x = 3970}$

Then correct mean.

$CRT \ \overline{X} = \frac{CRT \ \Sigma x}{n}.$

$= \frac{\cancel{3970}^{39.7}}{\cancel{100}} = 39.7.$

$\boxed{CRT \ \overline{X} = 39.7}$

⑩. The average rainfall of a city from monday to saturday is 0.3 inch. Due to heavy rainfall on sunday, the average rainfall for the week increased with 0.5 inch. Find out the rainfall on ~~saturday~~ sunday.

Given;

Average of city from Mon. to Sat. (or $\overline{X}_1$) $\frac{\sum_{i=1}^{6} x_i}{n} = 0.3$ inch..

Average of city from Mon to Sun (or $\overline{X}_2$) $\frac{\sum_{i=1}^{7} x_i}{n} = 0.5$ inch.

$\overline{X}_1 = \frac{\sum_{i=1}^{6} x_i}{n} \Rightarrow 0.3 = \frac{\sum_{i=1}^{6} x_i}{6} \Rightarrow \boxed{1.8 = \sum_{i=1}^{6} x_i}$

$\overline{X}_2 = \frac{\sum_{i=1}^{7} x_i}{n} \Rightarrow 0.5 = \frac{\sum_{i=1}^{7} x_i}{7} \Rightarrow \boxed{3.5 = \sum_{i=1}^{7} x_i}$

Then the rainfall on sunday $= \sum_{i=1}^{7} x_i - \sum_{i=1}^{6} x_i$

$= 3.5 - 1.8 = 1.7$ inch.

$\sum_{i=1}^{6} x_i \Rightarrow$ mon-Saturday rainfall

$\sum_{i=1}^{7} x_i \Rightarrow$ mon-Sunday rainfall

(11). The mean of 200 items is 50. Later on it was discovered that 192 was wrongly taken as 92. Find the correct mean.

$n = 200$ $\quad$ $\bar{x} = 50$.

correct value $= 192$ $\quad$ Wrongly value $= 92$.

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

$$50 = \frac{\Sigma x}{200}.$$

$\boxed{10,000 = \Sigma x}$ → Summation of 200 items.

Correct summation.

CRT Σx = Summation − Wrong value + correct value.

$= 10000 - 92 + 192$.

$= 10,000 + 100 = 10,100$.

Then the correct mean

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

$$= \frac{10100}{200} = 50.5.$$

$\boxed{\text{CRT } \bar{x} = 50.5}$

Scanned by CamScanner

Merits of Arithmatic Mean

* It is easily to understand and easy to calculate
* It is based upon all the observations.
* It is rigidly defined.
* It is capable of future algebric treatement.
* It is least affected by fluctuation of sampling.

Demerits of Arithmatic Mean :

* It is affected very much by extream values.
* It cannot be accurately determined even if one of the values is not known.
* It cannot be calculated for distributions with open end classes.
* It cannot be located graphically.

Home work :

1). The contents of 50 boxes of matches were counted, by given the following results.

No. of Matches.	41	42	43	44	45	46
No. of Boxes.	5	8	13	12	7	5

Calculate the mean.

Given $\boxed{n = 50}$.

x	f	xf
41	5	205
42	8	336
43	13	559
44	12	528
45	7	315
46	5	230.
	$\boxed{\Sigma f = 50}$	$\boxed{\Sigma xf = 2173}$

Then mean

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

$$= \frac{2173}{50}$$

$$= 43.46$$

② . Find out the missing frequency from the following data whose arithmetic mean is 67.45.

x	59.5–62.5	62.5–65.5	65.5–68.5	68.5–71.5	71.5–74.5
f	5	18	F	27	8

Given $\boxed{\text{Mean} = 67.45}$

C.I	f	middle value x	d=A−x	Σfd
59.5–62.5	5	61	−6	−30
62.5–65.5	18	64	−3	−54
65.5–68.5	F	$\boxed{67}$ (A)	0	0
68.5–71.5	27	70	3	81
71.5–74.5	8	73	6	48
	$\boxed{\Sigma f = 58+F}$			$\boxed{\Sigma fd = 45}$

Given mean = 67.45

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

$$67.45 = \frac{45}{58+F} + \cancel{67}$$

$$\cancel{3912.1 + 67.45\,F = 45}$$

$$\cancel{67.45\,F = 3912.1 - 45}$$

$$\cancel{67.45\,F = 3867.1}$$

$$\cancel{\boxed{F = 57.332}}$$

$$0.45 = \frac{45}{58+F}$$

$$\frac{\cancel{45}}{100} = \frac{\cancel{45}}{58+F}$$

$$58+F = 100$$

$$F = 100 - 58$$

$$\boxed{F = 42}$$

Scanned by CamScanner

MEDIAN

Median is the value which divides the distribution into two halves. In simple words, median is the middle value of the given distribution.

If $x_1, x_2, \dots, x_n$ are given.,

- Arrangement in either ascending / descending order.
- If n is odd.

  The median is $\left(\frac{n+1}{2}\right)^{th}$ item/entity.

- If n is even.

  The median is $\left(\frac{n+1}{2}\right)^{th}$ entity that leads 2 items.

  The mean of those items gives you the accurate median.

When frequencies are given.,

- Find the cumulative frequency (C.F).
- Then median = size of $\left(\frac{N+1}{2}\right)^{th}$ item.

If the data is in continuous series,

- Determine the median by

$$\text{Median} = L + \frac{N/2 + c.f^{*}}{2} \times i.$$

$L \to$ Lowest limit of middle class*

$N \Rightarrow \Sigma x$ *. ~~C.F → Preceeding~~

* Middle class : The entity and its corresponding values are grouply called as middle class.

*Note! In this case, C.F is taken from the preceeding data.

data ~~to~~

1. Find the median of the following data

84, 91, 72, 68, 87, 78

$\Rightarrow$ 68, 72, 78, 84, 87, 91.

Here $n = 6$. (Even)

Median $= \left(\frac{n+1}{2}\right)^{th}$ entity.

$= \frac{6+1}{2} = \frac{7}{2} = 3.5$ (lies between 3 & 4 th entity)

= between 78 and 84.

i.e. Median

$= \frac{78+84}{2} = \frac{162}{2} = 81.$

$\boxed{\text{Median} = 81}$.

2. Calculate the median of the following distribution.

x	10	15	8	20	18.
f	24	6	30	16	26.

x	f	c. f.
8	30	30
$\boxed{10}$	24	54
15	6	60
18	26	86
20	16	$\boxed{102} \rightarrow \boxed{N = \Sigma f}$

M-C

$\frac{N+1}{2} = \frac{102+1}{2} = 51.5$

Middle class is at $\boxed{x = 10}$

$\therefore$ The median $= 10$

Scanned by CamScanner

⑤. Calculate the median for the following data,

Savings less than f	10	20	30	40	50	60	70	80
c.f	15	35	64	84	96	120	192	256

~~x~~	~~f~~	~~c.f~~
~~0-10~~	~~15~~	~~15~~
~~10-20~~	~~35~~	~~50~~
~~20-30~~	~~64~~	~~114~~
~~30-40~~	~~84~~	~~198~~
~~40-50~~	~~96~~	~~294~~
~~50-60~~	~~120~~	
~~60-70~~	~~192~~	
~~70-80~~	~~256~~	

x	f	c.f
0-10	15	15
10-20	20	35
20-30	29	64
30-40	20	84
40-50	12	96
50-60	24	120
$\boxed{60-70}$	72	192
70-80	64	$\boxed{256}$ → N.

$$\frac{N+1}{2} = \frac{256+1}{2} = \frac{257}{2} = 128.5$$

∴ Middle class lies at interval $x = 60-70$.

lower limit $L = 60$.

$$\therefore \text{Median} = L + \frac{N/2 - c.f}{f} \times i$$

$\boxed{N/2 = 128}$

$$= 60 + \frac{128 - 120}{72} \times 10$$

$$= \cancel{60 + \frac{320}{72} \times 10} = 60 + \frac{8}{72} \times 10$$

$$= 60 + \frac{10}{9}$$

$$= \frac{540+10}{9} = \frac{550}{9}$$

$\boxed{\text{Median} = 61.11}$ //.

(4) Compute the mean and median of the following data :-

C.I	120-150	150-180	180-210	210-240	240-270	270-300	300-330	330-360
f	25	65	125	430	320	175	79	21

C.I $[x_1 - x_2]$	f	middle value $[x]$	$d = \frac{x-A}{i}$	fd	c.f
120-150	25	135	-3	-75	25
150-180	65	165	-2	-130	90
180-210	125	195	-1	-135	225
$\boxed{210-240}$	420	$\boxed{225}$ (A)	0	0	665 M.C
240-270	320	255	1	320	975
270-300	175	285	2	350	1150
300-330	79	315	3	237	1229
330-360	21	345	4	84	$\boxed{1250}$ → N
	$\boxed{\Sigma f = 1250}$			$\boxed{651}$ → Σfd	

Here $A = 225$.

Arithmatic Mean :

$$\text{Mean} = A + \frac{\Sigma fd}{\Sigma f} \times i$$

$$= 225 + \frac{651}{1250} \times 30$$

$$= 225 + 15.624$$

$$\boxed{\bar{X} = 240.624}$$

Arithmatic Median . [Here middle class $N/2 = \frac{1250}{2} = 625$]

$$\text{Median} = L + \frac{N/2 - c.f}{f} \times i$$

$$= 210 + \frac{625 - 225}{\cancel{430}\ 43} \times \cancel{30}\ 10$$

$$= 210 + \frac{400 \times 10}{43} = 210 + 93.023$$

$$= 303.023$$

Compute the median and arithmetic mean

Middle value	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

As the middle value is given, the length given between first and second entity is 10. i.e length $i = 10$.

x	f	$d = \frac{x-A}{i}$	fd	c.f
115	6	-4	-24	6
125	25	-3	-75	31
135	48	-2	-96	79
145	72	-1	-72	151
$\boxed{155}$	116	0	0	267 → M.C.
165	60	1	60	327
175	38	2	76	365
185	22	3	66	387
195	3	4	12	$\boxed{389}$ → N

$$\begin{array}{r} -267 \\ 214 \\ \hline -53 \end{array}$$

$\Sigma f = 390$ $\quad \Sigma fd = -53$

Arithmetic Mean:

$\boxed{\frac{N}{2} = 194.5}$

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

$$= 155 - \frac{53}{39\cancel{0}} \times 1\cancel{0}$$

$$= \frac{6045 - 53}{39} = \frac{5992}{39} = 153.64$$

$$\boxed{\bar{x} = 153.64}$$

Median: [Middle class in at N/2 = 194.5 that lies at 267]

$$\text{Median} = L + \frac{N/2 - c.f}{f} \times i$$

$$= 150 + \frac{194.5 - 151}{116} \times 10$$

$$= 150 + 3.75 = 153.75$$

$$\boxed{\text{Median} = 153.75}$$

Merits

Merits of Median:

* It is easy to understand and easy to calculate
* It is not affected by extreame value.
* In some cases, it can be located by inspection.
* It can be calculate for distributions with open end cases.
* It can be determined graphically.

Demerits of Median:

* It is not capable for ~~arith~~ algebric treatement.
* It is affected more by sampling fluctuation as compared to the value of mean.
* It is necessary ~~the~~ to arrange the data to calculate the median.

MODE

Mode is that which occurs most often ~~[illegible]~~ occured in the data that is with the highest frequency.

* For individual series, mode is the value which is repeated maximum number of times.
* For discreate series, the mode is obtained by just inspection. That is the value which corresponds to maximum frequency.

The inspection method is fail when the difference between the maximum frequency and the preceding frequency is very small, in such cases it is disirable to prepare grouping and analysis table to find the mode.

Problems

① Find the mode of the following data.

~~3, 2, 2, 5, 6~~ 2, 2, 3, 5, 6, 8, 5.

Mode = 5 (it occurs 3 times)

②. Find the mode of the following data

x	3	5	7	9	11	13	15	17
f	2	5	7	8	15	7	5	1

Highest frequency is 15, which corresponds to x values is 11.

∴ Mode = 11.

③. Calculate the Mode

x	25	30	35	40	45	50	55
f	7	11	17	15	14	10	11

Highest frequency : 17

Preceding frequency : 11.

Difference = 6 (small)

∴ As the difference is small, inspection method is fail.

GROUPING TABLE :

x	F col 1	(2) col 2	1-(2) col 3	(3) col 4	1-(3) col 5	2-(3) col 6
25	7	18		35		
30	11		28		43	
35	17	32				46
40	15		29	39		
45	14	24			35	
50	10		21			
55	11					

Analysis table

x	I	II	III	IV	V	VI	Total
25							
30					|		1
35	|	|			|	|	4
40		|	|	|	|	|	5
45			|	|		|	3
50				|			1
55							

The maximum tally occurs at $x = 40$

Hence $\boxed{\text{mode} = 40}$.

NOTE:

The grouping table has 6 columns:

col 1: The frequencies.

col 2: Group of data in two's

col 3: leaving the first frequency, group the data in twos.

col 4: Group of data in three's

col 5: Leaving the first frequency, group the data in threes

col 6: Leaving the 1st and second frequency, group the data in ~~bf~~ tree's.

Finally, ensure the highest frequency in each column.

Analytical table:

Tally box are drawn to variable x having max. value for each of six columns.

PROBLEM. ASSIGNMENT:

calculate the mean and median!

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-10
Frequency.	5	15	40	70	90	100	80	35	10	5

C.I	f	Midvalue (x)	fx	C.F
0-10	5	5	25	5
10-20	15	15	225	20
20-30	40	25	1000	60
30-40	70	35	2450	130
40-50	90	45	4050	$\boxed{220}$
$\boxed{50-60}$	$\boxed{100}$	55	5500	320 → M.C.
60-70	80	65	4900	400
70-80	35	75	2625	435
80-90	10	85	850	445
90-100	5	95	475	$\boxed{450}$ → N.

$\boxed{\Sigma f = 450}$. $\boxed{\Sigma fx = 22{,}400}$

Arithmatic Mean:

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

$$= \frac{\overset{448}{\cancel{22,400}}}{\underset{9}{\cancel{450}}} = 49.77.$$

Median:

- Median $= l + \dfrac{N/2 - c.f}{f} \times i$.

$N/2 = \dfrac{450}{2} = 225$, lower class is at ~~225~~ $\boxed{x=55}$

$c.f \le$ ~~445~~ $\boxed{c.f = 220}$ $\boxed{f = 100}$ $\boxed{i = 10}$

$$\text{Median} = 50 + \frac{225-220}{10\cancel{0}} \times 1\cancel{0}$$

$$= 50 + \frac{5}{10}$$

$$= \cancel{50+1/2}\ 50 + 0.5$$

$$= \cancel{52}\ 50.05$$

Calculate the mode :

x	10	20	30	40	50	60
f	6	14	29	31	36	38

From the table,

highest frequency is 88.

preceding frequency is 36.

The difference is small

So the inspection method ~~fails~~ fails.

x	COL 1	COL 2	COL 3	COL 4	COL 5	COL 6
10	6	20				
20	14.		~~43~~	49		
30	29.	60	43		74	
40	31					96 ~~[illegible]~~
50	36	74	67 ~~105~~	65		
60	38.					

Analysis Table

x	I	II	III	IV	V	VI	Total
10							
20					1		1
30					1	1	2
40			1	1	1	1	4
50		1	1	1		1	4
60	1	1	1	1			3

Since we have 2 highest frequency,

this method too fails.

Using Emperical Formula :,

Mode = 3 Median - 2 Mean

x	f	xf	c.F
10	6	60	6
20	14	280	20
30	29	870	49
40	31	1240	80
50	36	1800	116
60	38	2280	$\boxed{154} \to N$.

$\Sigma xf = 6530$.

$\Sigma f = 154$.

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{6530}{154} = 42.40.$$

$$\text{Median} = \frac{N+1}{2} = \frac{154+1}{2} = \frac{155}{2} = 77.5 \simeq 80$$

$\Rightarrow$ lies between $x = 30$ & $x = 40$,

$\therefore \boxed{\text{Median} = 40}$

$\therefore$ Mode = 3 Median - 2 Mean

$= 3(40) - 2(42.40)$

$= 120 - 84.80$

$= 35.20$

$\boxed{\text{Mode} = 35.20}$

For a continuous series, Mode $= L + \dfrac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$

Calculate the mode:

x	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
f	3	6	10	20	15	5	4	2

Modal class by inspection ⇒ x = 20-25 ; f = 20

$f_1 = 20$ $f_0 = 10$ $f_2 = 15$ $L = 20$

$$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 20 + \frac{20-10}{40-10-15} \times 5$$

$$= 20 + \frac{10}{\cancel{15}_3} \times \cancel{5} =$$

$$= 20 + 3.33 = 23.33$$

Calculate the mode:

C.I	90-100	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-
f	4	2	18	22	21	19	10	3	2

~~Modal class~~ Highest frequency by inspection = 22.
Second higher = 21.

Difference = 1 (small)

So Inspection method fails.

Grouping table:

x	col 1	col 2	col 3	col 4	col 5	col 6
90-100	4					
100-110	2	6				
110-120	18		20	24		
120-130	22	40			44	61
130-140	21		43			
140-150	19	40		62	50	
150-160	10		29			32
160-170	3	13	5	15		
170-180	2					

Analysis table

x	I	II	III	IV	V	VI	Total
90-100							
100-110							
110-120		1				1	2
120-130	1	1	1	1		1	5
130-140		1	1	1	1	1	5
140-150		1		1	1		3
150-160					1		1
160-170							
170-180							

We got two higher frequencies. Thus this method also fails.

Arithmetic Mean And Mode

C.I	f	Mi. Value x	$d=\frac{x-A}{i}$	fd	c.f
90-100	4	95	-4	-16	4
100-110	2	105	-3	-6	6
110-120	18	115	-2	-36	24
120-130	22	125	-1	-22	46
$\boxed{130-140}$	21	$\boxed{135}$ (A)	0	0	67
140-150	19	145	1	19	86
150-160	10	155	2	20	96
160-170	8	165	3	9	99
170-180	2	175	4	8	$\boxed{101}$ = N

$\boxed{\Sigma f = 101}$ $\boxed{\Sigma fd = -24}$

Let $\boxed{A = 135}$

$$\text{Mean} = A + \frac{\Sigma fd}{\Sigma f} \times i$$

$$= 135 + \frac{-24}{101} \times 10$$

$$= 135 - 2.37 = 132.624$$

$$\text{Median} = L + \frac{N/2 - c.f}{f} \times i$$

$$= 130 + \frac{50.5 - 46}{21} \times 10$$

$$= 130 + \frac{\cancel{4.5}\ 1.5}{\cancel{21}\ 7} \times 10$$

$$= 130 + \frac{15}{7} = 130 + 2.14 = 132.14$$

Empirical formula of Mode = 3 Median − 2 Mean

$$= 3(132.14) - 2(132.624)$$

$$= 396.42 - 265.24 = 131.78$$

$\boxed{\text{Mode} = 131.78}$

Calculate the ~~mode~~ Median.

Marks less than	5	10	15	20	25	30	35	40	45	50
Frequencies	~~29~~	~~224~~	~~465~~	~~582~~	~~624~~	~~644~~				
	29	224	465	582	624	644	650	653	655	

C.I	cf	Middle value	f	$d = \frac{x-a}{i}$	fd	c.f
0-5	~~29~~	2.5	29	-4	-116	29
5-10	~~224~~	7.5	224	-3	-672	253
10-15	~~465~~	12.5	465	-2	-930	718
15-20		17.5	582	-1	-582	1300
20-25		$\boxed{22.5}$ a	624	0	0	1934
$\boxed{25-30}$		27.5	644	1	644	2578
30-35		32.5	650	2	1300	3228
35-40		37.5	653	3	1959	3881
40-45		42.5	655	4	2620	$\boxed{4536}$ N

$\boxed{\sum fd = 4293}$

2268

$$\frac{N}{2} = \frac{4536}{2} = 2268.$$

∴ Middle class is at C.I ⇒ 25-30.

∴ $L = 25$.

$$\therefore \text{Median} = L + \frac{\frac{N}{2} - c.f}{f} \times i$$

$$= 25 + \frac{2268 - 1934}{644} \times 5$$

$$= 25 + 2.593 \cancel{= 64.827}$$

$$= 27.593$$

Merits of ~~Median~~ Mode:

- It is easy to calculate and in some cases it can be located by inspection.
- It is not affected by external values.
- It can be located for distribution with open-end classes.
- It can be determined graphically.

Demerits of Mode:

- The value of mode cannot always be determined. In some cases, we may have bimodal or multimodal series.
- It is not capable of further algebric treatement.
- The value of the mode is not based on each and every item of the series.
- It is affected to greater extent by sampling fluctuations as compared to the value of the mean.

Homework:

Calculate the median and mode.

Mark less than	5	10	15	20	25	30	35	40	45
Frequency	29	224	465	582	634	644	650	653	655.

C.I	C.F	F	Mid. Val. x	$d = \frac{x-A}{i}$	fd.
0-5	29	29	2.5	-4	-116
5-10	224	195	7.5	-3	-585
$\boxed{10\text{-}15}$	465	241	12.5	-2	-482 M.C.
15-20	582	117	17.5	-1	-117
20-25	634	52	$\boxed{22.5}$	0	0
25-30	644	10	27.5	1	10
30-35	650	6	32.5	2	12
35-40	653	3	37.5	3	9
40-45	655	2	42.5	4	8
		$\boxed{\Sigma f = 655}$			$\Sigma fd = -1261.$

$$\text{Mean} = A + \frac{\Sigma fd}{\Sigma f} \times i$$

$$= 22.5 - \frac{1261}{655} \times 5$$

$$\begin{array}{r} 22.500 \\ 9.63 \\ \hline 12.87 \end{array}$$

$$= 22.5 - 9.630$$

$$= 12.870$$

$$\text{Median} = L + \frac{\frac{N}{2} - c.f}{f} \times i$$

$$N/2 = \frac{655}{2} = 347.5 \quad c.f = 224 \quad f = 241.$$

$$\text{Median} = 10 + \frac{347.5 - 224}{241} \times 5$$

$$= 10 + \frac{123.5}{241} \times 5$$

$$= 10 + 2.562$$

$$= 12.562$$

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}.$$

$$= 3(12.56) - 2(12.87)$$

$$= 37.68 - 25.54$$

$$= 12.14$$

GEOMETRIC MEAN

* For ~~continuous~~ set of values and simple observations

i.e. $x_1, x_2, \ldots, x_n$.

$$G.M = (x_1 \cdot x_2 \cdot \ldots \cdot x_n)^{1/n}.$$

* For larger set of values;

$$G.M = \text{Antilog}\left[\frac{1}{n}\sum_{i=1}^{n}\log x_i\right]$$

* For discreate series

$$G.M = \text{Antilog}\left[\frac{\sum_{i=1}^{n} f\log x_i}{\sum f}\right].$$

* For continuous series

$$G.M = \text{Anti log}\left[\frac{\sum_{i=1}^{n} f\log m_i}{\sum f}\right]$$, $m \Rightarrow$ middle value.

PROBLEMS:

① Compute the ~~mean~~ geometric mean,

2, 8, 32, 36, 6.

$n = 5$.

$$G.M = (x_1 \cdot x_2 \cdots x_n)^{1/n}$$

$$= (2 \cdot 8 \cdot 32 \cdot 36 \cdot 6)^{1/5}$$

$$= 10.203$$

② Find the geometric mean,

82, 93, 50, 54, 72.

$n = 5$

$$G.M = (82 \cdot 93 \cdot 50 \cdot 54 \cdot 72)^{1/5} = 68.2650.$$

Method 2.

$(x_1 \cdot x_2 \cdots x_5)^{1/n}$

$\frac{1}{n} \log(x_1 \cdot x_2 \cdots x_5)$

$= \frac{1}{n}(\log x_1 + \log x_2 + \log x_3$

$\log(ab)$

$\log a + \log b$

x	log x
82	1.9138
93	1.9685.
50	1.6990.
54	1.7324.
72	1.8573
	9.1710

$$G.M = \text{Antilog}\left(\frac{1}{n}\sum_{i=1}^{5} \log x_i\right)$$

$$= \text{Antilog}\left(\frac{1}{5}(9.1710)\right)$$

$$= \text{Antilog}(1.8342)$$

$$= 68.26$$

③ Compute the geometric mean.

Category	I	II	III	IV	V	VI	VII	VIII
Monthly inc. x.	500	3750	3000	750	600	400	300	200
No. of Employees f	2	4	6	8	6	100	10	50

x	f	log x	f(log x)
500	2	2.6990	5.3980
3750	4	3.5740	14.2960
3000	6	3.4771	20.8626
750	8	2.8750	23.000
600	6	2.7781	16.6686
400	100	2.6020	260.2
300	10	2.4771	24.771
200	50	2.3010	115.05
	186		480.2412.

$$G.M = \text{Antilog}\ \frac{\sum f \log x}{\sum f}$$

$$= \text{Antilog}\left[\frac{480.2462}{186}\right]$$

$$= 381.92$$

7) Compute the ~~following~~ G.M using the given data.

class	0-10	10-20	20-30	30-40	40-50.
frequency	5	7	15	25	8.

C.I	Mid.Value m	f	$\log m$	$f \log m$
0-10	5	5	0.699	3.495
10-20	15	7	1.176	8.233
20-30	25	15	1.398	20.97
30-40	35	25	1.544	38.6
40-50	45	8	1.653	13.224
~~50-60~~		60		84.524

$\sum f = 60$ $\quad \sum f \log m = 84.524.$

$$G.M = \text{Antilog}\left[\frac{\sum f \log x}{\sum f}\right]$$

$$= \text{Antilog}\left[\frac{84.524}{60}\right]$$

$$= \text{Antilog}\,[1.4087]$$

$$= 25.63$$

5) Compute the geometric mean of the following data

x	0-10	10-20	20-30	30-40	40-50	~~50-60~~
f	5	7	15	25	8	

C.I	f	x	$\log x$	$f \log x$.
0-10	5	5		
10-20	7	15		
20-30	15	25		
30-40	25	35		
40-50	8	45		

Harmonic mean:

Harmonic mean is defined as a reciprocal of the arithmatic mean of the reciprocal of the items of the series.

$$H.M = \frac{1}{A.M} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}}$$

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

For discreate series,

$$H.M = \frac{\sum f}{\sum f/x}.$$

For continuous series,

$$H.M = \frac{\sum f}{\sum f/m_i}$$ $m_i \rightarrow$ middle value.

Scanned by CamScanner

Problems:

) Find the harmonic mean of following.

$\frac{1}{10} = \frac{10}{100}$

1, 0.5, 10, 45, 175, 0.01, 4, 11.2.

x	x	1/x
1	1	1.
0.5	1/2	2
10	10	~~0.0~~0.1
45	45	~~0.0~~0.02
175	175	~~5.714×10^-3~~ = 0.0057
0.01	$\frac{1}{100}$	100
4	4	0.25
11.2	11.2	0.089
		~~104.143~~ 103.516

$\Sigma\, 1/x = 103.173.$

$n = 8.$

Harmonic mean

$$HM = \frac{n}{\sum_{i=1}^{n} 1/x}$$

$$= \frac{8}{\sum_{i=1}^{8} 1/x}$$

$$= \frac{8}{103.173} = 0.0773.$$

③ Find the geometric mean of the following.

x	10	20	25	40	50
f	20	30	50	15	5

x	f	1/x	f 1/x
10	20	0.1	2
20	30	0.05	1.5
25	50	0.04	2.0
40	15	0.025	0.375
50	5	0.02	0.1
	$\Sigma f = 120$		$\Sigma f/x = 5.975$

$$\text{Harmonic mean} = \frac{\Sigma f}{\Sigma f/x} = \frac{120}{5.975}$$

$\simeq 20.08.$

③ Find Harmonic Mean.

x	10-20	20-30	30-40	40-50	50-60
f	4	6	10	7	3

C.I	f	x	1/x	f/x
10-20	4	15	0.067	0.268
20-30	6	25	0.040	0.240
30-40	10	35	0.028	0.288
40-50	7	45	0.022	0.156
50-60	3	55	0.018	0.055
	30			1.005

$\Sigma f = 30$ $\quad \Sigma f/x = 1.005$

Harmonic Mean

$$= \frac{\Sigma f}{\Sigma f/x}$$

$$= \frac{30}{1.005}$$

$$= 29.850$$

UNIT 2

Q.no:

MEASUREMENT OF DISPERSION

RANGE:

Range is defined as the difference between largest and smallest value of the distribution:

Range = L - S L → larger value, S → Smaller value.

Relative measure corresponding to range is coefficient of Range and it is given by,

$$\text{Coefficient of range} = \frac{L-S}{L+S}$$

PROBLEMS:

Calculate the range and coefficient of the following:

(1) 27, 32, 16, 15, 10, 30, 15, 29, 19, 35:

L = 35 S = 10.

∴ Range = L - S
= 35 - 10 = 25

$$\text{co-efficient} = \frac{L-S}{L+S} = \frac{25}{35} = \frac{5}{7} = 0.714.$$

(2)

x	4	6	8	10	12
f	15	25	12	36	30.

L = 12 S = 4.

Range = L - S = 12 - 4 = 8

$$\text{coefficient} = \frac{L-S}{L+S} = \frac{8}{16} = \frac{1}{2} = 0.5$$

③

Marks	10-20	20-30	30-40	40-50
No. of students	5	8	10	7

$L = 50$ $S = 10.$

$\therefore$ Range $= L - S = 50 - 10$

$= 40.$

Coefficient $= \frac{L-S}{L+S} = \frac{40}{60} = \frac{2}{3} = 0.667.$

Quartiles of Distribution:

In the series of distribution, the distributions are divide into four parts are called quartiles.,

$Q_k =$ size of $k\left(\frac{N+1}{4}\right)^{th}$ item $= L + \frac{kN/4 - cf}{f} \times i$

For individual ~~series~~ and discrete series)

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

$Q_2 =$ size of $2\left(\frac{N+1}{4}\right)^{th}$ item $=$ median.

$Q_3 =$ size of $3\left(\frac{N+1}{4}\right)^{th}$ item.

For continuous series;

$$Q_1 = L + \frac{N/4 - c.f}{f} \times i$$

$$Q_2 = L + \frac{2N/4 - cf}{f} \times i$$

$$Q_3 = L + \frac{3N/4 - cf}{f} \times i$$

Quartile Deviation:

Quartile deviation (a.k.a. semi-interquartile range)

$$Q.D = \frac{Q_3 - Q_1}{2}.$$

$$\text{Coefficient of } Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Problems:

From the given data, calculate the Quartile deviation ~~distribution~~ and its coefficient.

1490, 692, 777, 335, 582, 488, 753, 384, 407, 672, 522

$\Rightarrow$ 335, 384, 407, 488, 522, 582, 672, 692, 753, 777, 1490.

$N = 11$.

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

= size of $\frac{12}{4}^{th}$ item = size of 3rd item

= 407.

Q_2 = size of $2\left(\frac{N+1}{4}\right)^{th}$ item

= size of $\frac{24}{4}^{th}$ item = size of 6th item

= 582. $\Rightarrow$ Median.

Q_3 = size of $3\left(\frac{N+1}{4}\right)^{th}$ item.

= size of $3\left(\frac{12}{4}\right)^{th}$ item = size of 9th item

= 753

Quartile distribution;

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{753 - 407}{2} = \frac{346}{2} = 173.$$

coefficient of Q.D

$$\text{co-efficient of } Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{346}{1160} = 0.298$$

(3). Find the Quartile distributions and its coefficient of Q.D of the data :,

Roll.no	1	2	3	4	5	6	7
Marks	20	28	40	12	30	15	50.

Here $N = 7$. Marks $\Rightarrow$ 12, 15, 20, 28, 30, 40, 50.

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

$=$ size of $\left(\frac{7+1}{4}\right)^{th}$ Item $=$ size of $\frac{8}{4}^{th}$ Item.

$=$ size of 2^{th} Item

$= \cancel{12}\ \cancel{28}.\ 15$

$Q_2 =$ size of $2\left(\frac{N+1}{4}\right)^{th}$ Item $=$ size of $\frac{16}{4}$th Item

$=$ size of 4th Item

$= \cancel{12}.\ 28$

$Q_3 =$ size of $3\left(\frac{N+1}{4}\right)^{th}$ Item $=$ size of $3\left(\frac{8}{4}\right)^{th}$ Item

$=$ size of 6th Item

$= \cancel{15}\ 40$

$$Q.D = \frac{Q_3 - Q_1}{2} = \cancel{15-28}\ \frac{40-15}{2} = \frac{25}{2}$$

$$\boxed{Q.D = 12.5}$$

$$\text{Coefficient} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{40-15}{40+15} = \frac{25}{55} = \frac{5}{11} = 0.455$$

$$\boxed{\text{Co efficient} = 0.4545}$$

(3). Compute Quartile deviation and its co eff.

Weight	60	61	62	63	65	80	75	70
No. of workers	1	3	5	7	10	1	3	1

$x \rightarrow$ weights $f \rightarrow$ No. of workers.

x	60	61	62	63	65	70	75	80
f	1	3	5	7	10	1	3	1

x	f	c.f	
60	1	1	
61	3	4	
62	5	9	$\rightarrow Q_1$
63	7	16	
65	10	26	$\rightarrow Q_3$
70	1	27	
75	3	30	
80	1	31	$\rightarrow N.$

$\boxed{N=31}$.

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

$=$ size of $\frac{31+1}{4}^{th}$ item $=$ size of $\frac{32}{4}^{th}$ item

$=$ size of 8th item $= 62$.

$Q_3 =$ size of $3\left(\frac{N+1}{4}\right)^{th}$ item.

$=$ size of $3\left(\frac{32}{4}\right)^{th}$ item $=$ size of $3(8)^{th}$ item

$=$ size of 24^{th} item $= 65$.

Quartile deviation $= \frac{Q_3-Q_1}{2} = \frac{65-62}{2} = \frac{3}{2} = 1.5$

Coefficient $= \frac{Q_3-Q_1}{Q_3+Q_1} = \frac{3}{65+62} = \frac{3}{127}$

$= 0.0236$

Scanned by CamScanner

(14) Find the Quartile deviation for the following.

Marks	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	8	20	25	30	12	5

Marks x	Frequency f	c. f.
0-10	8	8
10-20	20	28 → Q_1
20-30	25	53
30-40	30	83 → Q_3
40-50	12	95
50-60	5	100 → N

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

= size of $\frac{100}{4}$ th item = size of 25 th item

Q_1 class = 10-20.

$L = 10$ $\quad C.f = 8$ $\quad N/4 = 25$ $\quad i = 10.$

$$Q_1 = L + \frac{N/4 - c.f}{f} \times i = 10 + \frac{25-8}{20} \times 10$$

$$= 10 + \frac{17}{2} = 10 + 8.5$$

$$\boxed{Q_1 = 18.5}$$

~~Q_3 = size of~~

Q_3 ~~class~~ = size of $3\frac{N}{4}$ th item = size of $\frac{300}{4}$ th item

= size of 75 th item

Q_3 class = 30-40.

$L = 30$ $\quad C.F = 53$ $\quad 3N/4 = 75$ $\quad i = 10.$ $\quad f = 30$

$$Q_3 = L + \frac{3N/4 - c.f}{f} \times i$$

$$= 30 + \frac{75-53}{30} \times 10 = 30 + \frac{22}{3} = 30 + 7.33$$

$$= 37.33.$$

Quadile deviation,

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{37.33 - 18.5}{2} = \frac{18.83}{2} = 9.415$$

$\boxed{Q.D = 9.415}$

$$\text{co.eff. of } Q.D = \frac{9.415}{55.83} = 0.1686.$$

Mean deviation:

i) Mean deviation above mean or Median.

For individual series;

$$M.D = \frac{\Sigma|D|}{n} \qquad |D| = |x - \text{Mean}| \text{ or } |x - \text{Median}|$$

$$\text{coefficient of } M.D = \frac{M.D}{\text{Mean}} \text{ or } \frac{M.D}{\text{Median}}.$$

For continuous Series;

$$M.D = \frac{\Sigma f|D|}{\Sigma f} = \frac{\Sigma f|D|}{N}.$$

PROBLEMS :

① Calculate the mean deviation above its median and its coefficient form.

4000, 3000, 4200, 5800, 4800.

⇒ 3000, 4000, 4200, 4800, 5800.

$$\text{Median} = \text{size of } \frac{N+1}{2} \text{ th item.}$$

$$= \text{size of } \frac{6}{2} \text{ th item}$$

$= \text{size of 3rd item} = \cancel{4200}$.

$\boxed{\tilde{X} = 4200}$

x	$\lvert x - \text{Median}\rvert$
3000	1200
4000	200
4200	0
4800	600
5800	1600
	$\Sigma\lvert D\rvert = 3600$

Mean deviation $MD = \frac{\Sigma \lvert D\rvert}{n}$

$= \frac{3600}{5} = 720.$

Coefficient $= \frac{M.D}{\text{Median}} = \frac{\cancel{720}^{6}}{\cancel{4200}_{35}} = 0.1714.$

(5). Find the mean deviation:

10, 20, 30, 25, 40, 45, 50, 70, 80, 90.

Mean $= \frac{\Sigma x}{n} = \frac{10+20+30+25+40+45+50+70+80+90}{10}$

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

x	$\lvert x - \text{Mean}\rvert$
10	36
20	26
25	21
30	16
40	6
45	1
50	4
70	24
80	34
90	44
	212

$\Sigma \lvert D\rvert = 212.$

Mean deviation

$M.D = \frac{\Sigma \lvert D\rvert}{n} = \frac{212}{10}$

$= 21.2.$

Scanned by CamScanner

(3) Calculate the mean deviation ~~and~~ about its median for the following

x	10	11	13	14	12
f	3	12	12	3	18

x	f	cf	|D| = |x - Median| Median = 12	f|D|
10	3	3	-2	+6
11	12	15	-1	12
12	18	33	0	0
13	12	45	1	12
14	3	48	2	6
	$\boxed{\Sigma f = 48}$			$\boxed{\Sigma fD = 36}$

From the ~~respective~~ value of N = 48,

Middle class (24) is at x = 12.

∴ Median = 12.

Mean deviation abt. its mean = $\frac{\Sigma f|D|}{\Sigma f} = \frac{36}{48} = \frac{3}{4} = 0.75$

(4) Calculate the mean deviation about the mean for the following

No. of calls	2	3	4	5	6	7
f	1	5	8	4	2	1

4 - 4.2

40 - 4.2

x	f	fx	D = |x - Mean| $\bar{x}$ = 4.2	f|D|
2	1	2	~~1.8~~ 2.2	2.2
3	5	15	~~0.8~~ 1.2	6.0
4	8	32	~~0.8~~ 0.2	1.6
5	4	20	0.8	3.2
6	2	12	1.8	3.6
7	1	7	2.8	2.8
	$\boxed{\Sigma f = 21}$	$\boxed{\Sigma fx = 88}$		$\Sigma fD = 19.4$

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{88}{21} = 4.190 \approx 4.2$$

$$\text{Mean deviation} = \frac{\Sigma f|D|}{\Sigma f}$$

$$= \frac{19.4}{21} = 0.924$$

$$\text{coefficient of MD} = \frac{MD}{Mean} = \frac{0.924}{4.2} = 0.22$$

⑤ Calculate the mean deviation and its coefficient for the following.

Size	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	40
frequency	6	10	18	30	15	12	10	6	3

Size x	frequency f	cf	Mid. x	$\lvert D\rvert = \lvert x - \text{Median}\rvert$ $\boxed{\tilde{X} = 18.8}$	$f\lvert D\rvert$
4 - 8	6	6	6	12.8	76.8
8 - 12	10	16	10	8.8	88
12 - 16	18	34	14	4.8	86.4
$\boxed{16\text{-}20}$	$\boxed{30}$	64	18	0.8	24
20-24	15	79	22	3.2	48
24-28	12	91	26	7.2	86.4
28-32	10	101	30	11.2	112
32-36	6	107	34	15.2	91.2
36-40	3	$\boxed{110}$	38	19.2	57.6
	$\boxed{N = 110}$				$\boxed{670.4}$

$$\text{Median} = L + \frac{N/2 - c.f}{f} \times i$$

$$= 16 + \frac{55-34}{30_{15}} \times \overset{2}{\cancel{4}}$$

$$= 16 + \frac{21^{7}}{15_{5}} \times 2 = 16 + \frac{14}{5} = \frac{80+14}{5} = \frac{94}{5} = 18.8$$

Mean deviation = $\frac{\Sigma f |D|}{\Sigma f}$

$= \frac{670.4}{110} = \frac{67.04}{11} = 6.095$

Coefficient of Mean deviation.

$= \frac{M.D}{Median} = \frac{6.095}{18.8} = 0.324$

STANDARD DEVIATION :-

Standard deviation is the square root of the mean of the squared deviation from their arithmatic mean. It is denoted by σ.

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

COEFFICIENT OF VARIATION :-

Co-efficient of variation is used in such problems when we want to compare the variability in 2 series. It is defined by.,

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

$\sigma \rightarrow S.D$
$\bar{x} \rightarrow Mean.$

S.D from actual Mean :-

$$\sigma = \begin{cases} \sqrt{\frac{\Sigma (x-\bar{x})^2}{n}}, & \text{for individual series.} \\ \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}}, & \text{for frequency distributions.} \end{cases}$$

S.D from Assumed Mean:

$$\sigma = \begin{cases} \sqrt{\dfrac{\Sigma d^2}{n} - \left(\dfrac{\Sigma d}{n}\right)^2}, & \text{for individual obs.} \\ \sqrt{\dfrac{\Sigma f d^2}{N} - \left(\dfrac{\Sigma f d}{N}\right)^2}. & \text{for frequency distribution.} \end{cases}$$

Here, $\boxed{d = x - A}$

STEP DEVIATION METHOD:

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

Here $\boxed{d = \dfrac{x - A}{i}}$.

COMBINED STANDARD DEVIATION:

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

$d_1 = \bar{X}_1 - \bar{X}_{12}$

$d_2 = \bar{X}_2 - \bar{X}_{12}$

for n steps,

$$\sigma_{12\cdots n} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + \cdots + N_n\sigma_n^2 + N_1 d_1^2 + N_2 d_2^2 + \cdots + N_n d_n^2}{N_1 + N_2 + \cdots + N_n}}$$

PROBLEMS:

① Compute the standard deviation and coefficient of variation for the following data:

1, 5, 4, 2, 3, 8, 6, 2, 8.

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

$$\frac{\Sigma x}{n} = \frac{1+5+4+2+3+8+6+2+8}{9} = \frac{39}{9} = 4.\dot{3}3.$$

$\Sigma x^2 = 1521.$

1) Compute the standard deviation and its C.V for the following:

1, 5, 4, 2, 3, 8, 6, 2, 8.

$$\frac{\Sigma x}{n} = \frac{1+5+4+2+3+8+6+2+8}{9} = \frac{39}{9} = 4.33$$

$$\frac{\Sigma x^2}{n} = \frac{1+25+16+4+9+64+36+4+64}{9} = \frac{223}{9} = 24.78$$

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

$$= \sqrt{24.78 - (4.33)^2}$$

$$= \sqrt{24.78 - 18.75}$$

$$= \sqrt{6.03} = 2.449$$

$$C.V = \frac{\sigma}{\bar{x}} \times 100 = \frac{2.449}{4.33} \times 100 = 56.57$$

5. Calculate the S.D from the following data.

9, 27, 18, 54, 45, 72, 36, 63, 81.

x	$d = x - A$	d^2
9	-36	1296
18	-27	729
27	-18	324
36	-9	81
$\boxed{45}$	0	0
54	9	81
63	18	324
72	27	729
81	36	1296
	$\Sigma d = 0$	$\Sigma d^2 = 4860$

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

$$= \sqrt{\frac{4860}{9}}$$

$$= \sqrt{540}$$

$$= 23.238$$

Scanned by CamScanner

(3) A consignment of 80 articles is classified according to the size of the article as below. Find the SD and its coefficient.

Measurement	80↑	70↑	60↑	50↑	40↑	30↑	20↑	10↑	0↑
No. of Articles	5	14	34	65	110	150	170	176	180,

C.I	c.f	f	x	$d=\frac{x-A}{i}$	d^2	fd	fd^2
0-10	180	4	5	-4	16	-16	64
10-20	176	6	15	-3	9	-18	54
20-30	170	20	25	-2	4	-40	80
30-40	150	40	35	-1	1	-40	40
40-50	110	45	$\boxed{45}$	0	0	0	0
50-60	150	31	55	1	1	31	31
60-70	170	20	65	2	4	40	80
70-80	176	9	75	3	9	27	81
80-90	180	5	85	4	16	20	80
		180				4	510

$\Sigma f=180$ $\Sigma fd=4$ $\Sigma fd^2=510$ $A=45$ $i=10$.

$$\sigma=\sqrt{\frac{\Sigma fd^2}{\Sigma f}-\left(\frac{\Sigma fd}{\Sigma f}\right)^2}\times i=\sqrt{\frac{510}{180}-\left(\frac{4}{180}\right)^2}\times 10$$

$$=\sqrt{2.833-(0.022)^2}\times 10$$

$$=\sqrt{2.832516}\times 10$$

$\boxed{\sigma=16.830}$

$$=1.6830\times 10=16.830$$

$$\bar{x}=A+\frac{\Sigma fd}{\Sigma f}\times i=45+\frac{\cancel{4}^{2}}{\cancel{180}_{9}}\times 1\cancel{0}=45+\frac{2}{9}=\frac{407}{9}=45.22.$$

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

$$=\frac{16.820}{45.220}\times 100=0.3721\times 100$$

$$=37.21$$

$\boxed{C.V=37.21}$. $\boxed{\sigma=16.830}$.

Find the coefficient
(5) Coefficient variation of two different distributions are 58% and 69%. Their standard deviations are 21.2 and 15.6 respectively. What are their arithmetic mean?

$C.V_1 = 58\%$ $\qquad$ $C.V_2 = 69\%$

$\sigma_1 = 21.2$ $\qquad$ $\sigma_2 = 15.6$.

$C.V_1 = \frac{\sigma_1}{\bar{x}_1} \times 100$ $\qquad$ $C.V_2 = \frac{\sigma_2}{\bar{x}_2} \times 100$

$58 = \frac{21.2}{\bar{x}_1} \times 100$. $\qquad$ $69 = \frac{15.6}{\bar{x}_2} \times 100$

$\bar{x}_1 = \frac{21.2}{58} \times 100$ $\qquad$ $\bar{x}_2 = \frac{15.6}{69} \times 100$

$= 36.55$ $\qquad$ $= 22.6$.

(6) Following data life of two models of refridgerators A and B.

Life (in months)	A	B
500 - 700	5	4
700 - 900	11	30
900 - 1100	26	12
1100 - 1300	10	8
1300 - 1500	8	6.

~~Find the~~ $\bar{x}$ $\frac{\Sigma f x}{\Sigma f} \times i$

$\sigma = \sqrt{f}$

Find the average life of each model. Which model has greater uniformity?

Let A ⟷ x & B ⟷ y.

C.I	f	x	$d = \frac{x-A}{i}$	d^2	fd	fd^2
500-700	5	600	-2	4	-10	20
700-900	11	800	-1	1	-11	11
900-1100	26	(A) $\boxed{1000}$	0	0	0	0
1100-1300	10	1200	1	1	10	10
1300-1500	8	1400	2	4	16	32
	$\Sigma f = 60$				$\Sigma fd = 5$	$\Sigma fd^2 = 73$

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

$$= 1000 + \frac{5}{\cancel{60}_{3}} \times \cancel{200}^{10}$$

$$= 1000 + \frac{50}{3} = \frac{3000 + 50}{3} = \frac{3050}{3} = 1016.67$$

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

$$= \sqrt{\frac{73}{60} - \left(\frac{\cancel{5}}{\cancel{60}}\right)^2} \times 200$$

$$= \sqrt{1.217 - \cancel{0.00069}\ 0.6944} \times 200$$

$$= \sqrt{0.5122\ \cancel{1.21631}} \times 200$$

$$= 0.792 \times 200$$

$$= \cancel{220.4}\ 144.$$

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{220.4}{1016.67} \times 100$$

$$\boxed{C.V_1 = 21.679}$$

CI	g	x	$d = \frac{x-A}{i}$	d^2	gd	gd^2
500-700	4	600	-2	4	-8	16
700-900	30	800	-1	1	-30	30
900-1100	12	$\boxed{1000}$	0	0	0	0
1100-1300	8	1200	1	1	8	~~64~~ 8
1300-1500	6	1400	2	4	12	~~144~~ 24
~~1500-1700~~						
	$\Sigma g = 60.$				$\Sigma gd = 18$	$\Sigma gd^2 = 182.$

$$\bar{X} = A + \frac{\sum gd}{\sum g} \times i$$

$$= 1000 + \frac{18}{60} \times \overset{10}{20\cancel{0}}$$

$$\frac{\sum gd}{\sum g} = \frac{18}{60} = \frac{3}{10}$$

$$= 0.3$$

$$= 1000 + 60$$

$$\boxed{\bar{X} = 1060}$$

$$\sigma = \sqrt{\frac{\sum gd^2}{\sum g} - \left(\frac{\sum gd}{\sum g}\right)^2} \times i$$

$$= \sqrt{\frac{\overset{60.03}{182}}{60} - \left(\frac{3}{10}\right)^2} \times i$$

$$= \sqrt{60.033 - 0.15} \times 200$$

$$= \sqrt{59.883} \times 200$$

$$= 7.737 \times 200$$

$$\boxed{\sigma = 1547.7.}$$

$$C.V = \frac{\sigma}{\bar{X}} \times 100.$$

$$= \frac{1547.7}{106\cancel{0}} \times 10\cancel{0}$$

$$= 14.60 \times 10$$

$$= 146.$$

(9) Two samples of size 40 and 50 have the same mean 53. But different standard deviation are 19 and 18 respectively. Find the standard deviation of a combined sample of size 90.

$n_1 = 40$ $n_2 = 50$ $\bar{x}_1 = 53$ $\bar{x}_2 = 53$.

$\sigma_1 = 19$ $\sigma_2 = 18$

~~a+b+a.c~~

~~=a(b+c)~~

$$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2} \cancel{= 40}$$

$$= \frac{(40.53)+(50.53)}{40+50} = 53.$$

$3\sqrt{10}$.

$d_1 = \bar{x}_1 - \bar{x}_{12} = 53-53 = 0$

$d_2 = \bar{x}_1 - \bar{x}_{12} = 53-53 = 0.$

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1+N_2}}$$

$$= \sqrt{\frac{(40.361)+(50.324)}{90}} = \sqrt{\frac{14440+16200}{90}}$$

$$= \sqrt{\frac{30640}{90}} \cancel{= \sqrt{228.7}}$$

$\cancel{= 28}$.

$$= \sqrt{340.44} = 18.45.$$

(8) The mean and standard deviation of 200 items are found to be 60 and 20 respectively. If at the time of calculation, 2 items were wrongly taken as 3 and ~~64~~ 67 instead of 30 and 17. Find the correct mean and standard deviation. Also find the correct coefficient of variation.

$n = 200$; $\bar{x} = 60$; $\sigma = 20$

Wrong values = 3 & 67

Correct values = 13 and 17.

Scanned by CamScanner

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

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

$\Sigma x = 60 \times 200 = 12000$

$\boxed{\text{: } \cancel{\text{from}}\ \Sigma x = 12000}$

~~CRT Σx = Total Σx − Wrong Σx + tor~~

CRT $\Sigma x = \Sigma x -$ Wrn. value Σx + correct value.

$= 12000 - (3+67) + 30+17$

$= 12000 - 70 + 47$

$= 12000 - 23 \quad = 11977$

$\boxed{\text{CRT } \Sigma x = 11977}$

$\sigma = 20$

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

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

20
×20
400

$n(\sigma^2 + \bar{x}^2) = \Sigma x^2$.

$\frac{60 \times 60}{3600}$

$200(400\cancel{0} + 360\cancel{0}) = \Sigma x^2$

$200(\cancel{7600}\ 4000) = \Sigma x^2$ $\quad \boxed{\Sigma x^2 = 800000}$

$\boxed{\cancel{1520000 = \Sigma x^2}}$

crt $\Sigma x^2 = \Sigma x^2 -$ (wrong value^2) + (crt value^2).^{2}

$= \cancel{15200000}\ 800000 - (3^2 + 67^2) + (13^2 + 17^2)$

$= \cancel{15200000}\ 800000 - (9 + 4489) + (169 + 289)$

$= \cancel{15200000}\ 800000 - 4498 + 458$.

$= \cancel{15200000}\ 800000 - 4040 = \cancel{147960}\ 795{,}960$//.

~~$= 1515960$~~

~~CRT//CRT~~

$\text{CRT } \sigma = \sqrt{\frac{\text{crt } \Sigma x^2}{n} - \text{crt}\left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{795960}{200} - \left(\frac{11977}{200}\right)^2}$

$= 19.84$

(4) From the following data, find ~~the out~~ which product is more stable in prices.

						Mean
Product A	20	22	19	23	16	20
Product B	10	20	18	12	15	15

x → product A   y → product B.

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
20	0	0	10	-5	25
22	2	4	20	5	25
19	-1	1	18	3	9
23	3	9	12	-3	9
16	-4	16	15	0	0
		30			68

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{(30)}{5}} = \sqrt{6} = 2.449$$

$$\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}} = \sqrt{\frac{68}{5}} = \sqrt{13.6} = 3.688$$

$$C.V_x = \frac{\sigma_x}{\bar{x}} \times 100 = \frac{2.449}{20} \times 100 = 12.245$$

$$C.V_y = \frac{\sigma_y}{\bar{y}} \times 100 = \frac{3.688}{15} \times 100 = 24.59$$

$CV_x < CV_y$

∴ Product A is more stable in prices.

(10). Following are marks obtained by student A & B in 10 set of Examination.

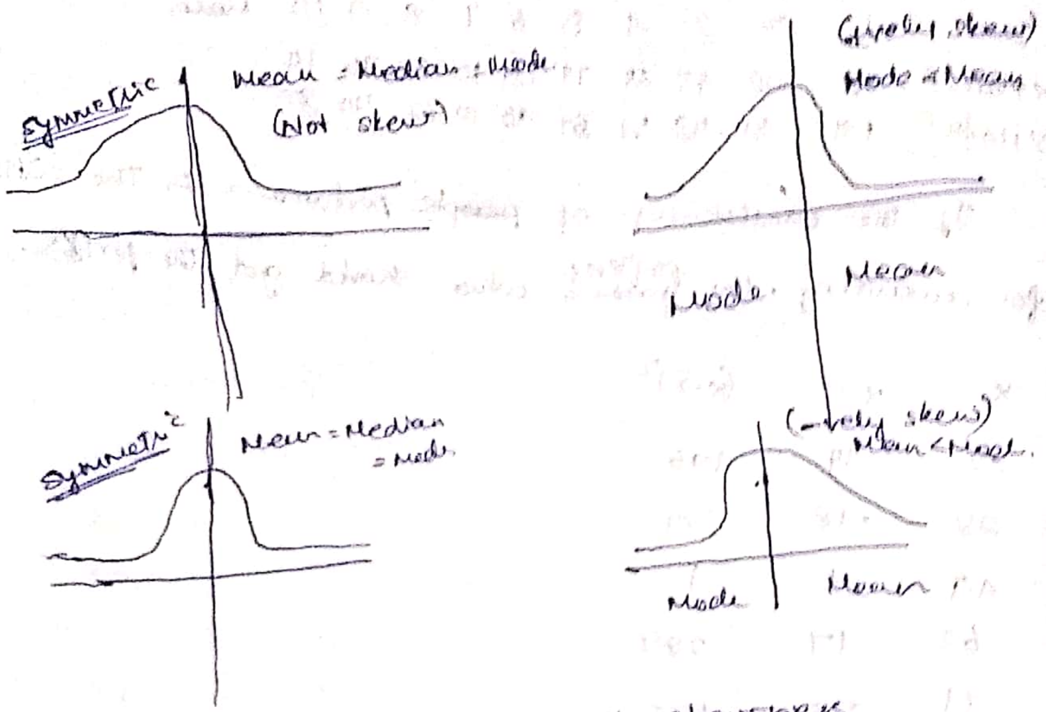
Sets	1	2	3	4	5	6	7	8	9	10	Mean
x Marks A	32	28	47	63	71	39	10	60	96	14	
y Marks B	19	31	48	53	67	90	10	62	40	80	

If the consistency of ~~people~~ performance is the criterion for awarding the prizes who should get the prize.

x	$x-\bar{x}$	$(x-\bar{x})^2$
32	-14	196
28	-18	324
47	1	1
63	17	289
71	25	625
39	-7	49
10	-39	1521
60	14	196
96	50	2500
14	-32	
460		

MEASUREMENT OF SKEWNESS

SKEWNESS



Karl pearsons Coefficient of skewness

$$S_k = \frac{\text{Mean} - \text{Mode}}{S.D.}$$

1. Measure of skewness ~~describes~~ describes the shape of the distribution
2. If the distribution is symmetric, Mean = Mode = Median
3. If the frequency curve has the long tail to the right or left, we say that it is positively or negatively skewed respectively.

Karl's Pearson's coefficient of skewness:

$$S_k = \frac{\text{Mean} - \text{Mode}}{S.D}$$

① Find the ~~Karl~~ coefficient of skewness for the following.

x	1	2	3	4	5	6	7
f	10	18	30	25	12	3	2

x	f	fx	fx^2
1	10	10	10
2	18	36	~~72~~
3	30	90	270
4	25	100	400
5	12	60	300
6	3	18	108
7	2	14	98
	100	328	1258

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{328}{100} = 3.28$$

Mode = 3 (highest frequency)

$$S.D\ \sigma = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{1258}{100} - \left(\frac{328}{100}\right)^2}$$

$$= \sqrt{\frac{1258}{100} - \frac{(328)^2}{10000}}$$

$$= \sqrt{\frac{1258}{100} - \frac{107584}{100000}}$$

$$= \sqrt{12.58 - 10.7584}$$

$$= \sqrt{1.8216}$$

$$= 1.35$$

$$S.k = \frac{\text{Mean} - \text{Mode}}{S.D}$$

$$= \frac{3.28 - \cancel{1.35}\ 3}{1.35}$$

$$= \frac{0.28}{1.35}$$

$$= 0.207$$

A. J. Thomas

② Calculate coefficient of skewness from, ~~more than~~

Marks More than	0↑	10↑	20↑	30↑	40↑	50↑	60↑	70↑	80↑
No. of students	150	140	100	80	80	70	30	14	0

C.I	C.f	f	x	$d=\frac{x-A}{i}$	d^2	fd	fd^2
0-10	150	10	5	-4	16	-40	160
10-20	140	40	15	-3	9	-120	360
20-30	100	20	25	-2	4	-40	80
$\boxed{30}$-40	80	0	35	-1	1	0	0
40-50	80	10	$\boxed{45}$	0	0	0	0
50-60	70	40	55	1	1	40	40
60-70	30	16	65	2	4	32	64
70-80	14	14	75	3	9	42	126
80-90	0	0	85	4	16	0	~~0~~
		$\Sigma f = 150$				-86	830

$\Sigma f = 150$ $\quad \Sigma fd = -86$ $\quad \Sigma fd^2 = 830.$

$$\text{Mean} = A \pm \frac{\Sigma fd}{\Sigma f} \times i \qquad \text{Median} =$$

$$= 45 + \frac{-86}{150} \times 10$$

$$= 45 - (0.5733 \times 10)$$

$$= 45 - 5.733$$

$$= 39.267$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} \times 10$$

$$= \sqrt{\frac{830}{150} - \left(\frac{-86}{150}\right)^2} = \sqrt{5.533 - 0.333}$$

$$= \sqrt{5.200}$$

$$= 2.280 = 22.8036$$

Scanned by CamScanner

$N/2 = $ $N = 150$

$N/2 = 75$ $L = 30$ $c.f = 80$ $f = 10$.

$$\text{Median} = L + \frac{N/2 - c.f}{f} \times i$$

$$= 30 + \frac{75 - 80}{10} \times 10$$

$$= 30 - 5 = 25$$

It has 2 highest frequency, so Median is used.

i

Mode = 3 Median − 2Mean.

$$S_k = \frac{\text{Mean} - \text{Mode}}{S.D.}$$

$$= \frac{\text{Mean} - 3\text{Median} + 2\text{Mean}}{S.D}$$

$$= \frac{3\text{Mean} - 3\text{Median}}{S.D}$$

$$= \frac{3(\text{Mean} - \text{Median})}{S.D}$$

$$= \frac{3(39.267 - 25)}{22.8136}$$

$$= 1.8761$$

A.J. Thomas

Bowley's Coefficient of Skewness

Q_1 = Size of $\frac{N}{4}^{th}$ = $L + \frac{N/4 - c.f}{f} \times i$

Q_2 = Median = $L + \frac{N/2 - c.f}{f} \times i$

Q_3 = Size of $3N/4^{th}$ = $L + \frac{3N/4 - c.f}{f} \times i$

Then, $BS_k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

Problems

① Find the Bowley's Coefficient of skewness for

x	0	1	2	3	4	5	6
f	7	10	16	25	18	11	8

x	f	c.f	
0	7	7	
1	10	17	
2	16	33	→ Q_1
3	25	58	→ Q_2
4	18	76	→ Q_3.
5	11	87	
6	8	$\boxed{95}$ – N	

$\frac{N}{4} = 23.75$

$\frac{N}{2} = 47.5$

$\frac{3N}{4} = 71.25$

$\therefore BS_k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} = \frac{4+2-2(3)}{4-2}$

$= \cancel{\frac{76+33-2(58)}{76-33}}$ $= \frac{6-6}{2} = 0.$

$= \cancel{\frac{109-116}{43}} = \cancel{\frac{-7}{43}}$

Skewness is zero.

∴ The distribution is symmetrical

alculate Bowley's coefficient of skewness

x	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f	10	25	20	15	10	35	25	10.

C.I	f	c.f
0-10	10	10
10-20	25	35
20-30 →	20 →	55 → Q_1 class
30-40	15	70
40-50 →	10 →	80 → Q_2 class
50-60 →	35 →	115 → Q_3 class
60-70	25	140
70-80	10	$\boxed{150}$ → N.

$\frac{N}{4} = \frac{150}{4} = \cancel{40}\ 37.5$

$\frac{N}{2} = \frac{150}{2} = \cancel{80}\ 75$

$3\frac{N}{4} = \frac{3(150)}{4} = \cancel{120}\ \cancel{115}\ 3(37.5)$

$= .5$

Q_1 class.

C.I = 20-30 $\quad N/4 = 37.5$

f = 20 $\quad$ L = 20

c.f = 35

$Q_1 = L + \frac{N/4 - c.f}{f} \times i$

$= 20 + \frac{\cancel{40}\ 37.5 - 35}{20} \times 10 = 20 + \frac{2.5}{2} = \cancel{20+2.5 = 22.5}\ 20+1.25$

$= 21.25$

Q_2 class

C.I = 40-50 $\quad$ L = 40

$N/2 = 80$ $\quad$ ~~L=40~~, f = 10 $\quad$ c.f = 70.

$Q_2 = L + \frac{N/2 - c.f}{f} \times i = 40 + \frac{\cancel{80}\ 75 - 70}{10} \times 10$

$= 40 + \cancel{10}\ 5 = \cancel{50}\ 45$

Q_3 class

C.I = 50-60 $\quad$ L = 50

$3N/4 = \cancel{117.5}\ 112.5$ $\quad$ f = 35 $\quad$ c.f = 80.

$Q_3 = L + \frac{3N/4 - c.f}{f} \times i = 50 + \frac{\cancel{117.5}\ 112.5 - 80}{35} \times i$

$= 50 + \frac{32.5}{35} \times 10 = 50 + \frac{325}{35}$

$= 59.285$

Scanned by CamScanner

Bowley's Coefficient

$$BS_k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$= \frac{59.285 + 21.25 - 2(45)}{59.285 - 21.25}$$

$$= \frac{-94.65}{38.04} = -2.488$$

The frequency of the distribution shows the following measurement.

Mean = 45, Median = 48, coefficient of skewness 0.4

Estimate the standard deviation.

Karl pearson's skewness coefficient

$$S_k = \frac{3(\text{Mean} - \text{Mode})}{S.D}$$

$$\cancel{0.4} = S_k = \frac{\text{Mean} - (3\,\text{median} - 2\,\text{mean})}{S.D}$$

$$= \frac{\text{Mean} - 3\,\text{Median} + 2\,\text{Mean}}{S.D}$$

$$= \frac{3(\text{Mean} - \text{Median})}{S.D}$$

$$0.4 = \frac{3(45 - 48)}{S.D}$$

$$S.D = \frac{3(-3)}{0.4}$$

$$= \frac{-9}{0.4} = \frac{-9}{4 \times 10^{-1}} = \frac{-9}{4} \times 10$$

$$= \frac{-90}{4} = -22.5$$

(1) In the frequency distribution, the coefficient of skewness based on Quartiles is 0.6. If the sum of the upper and lower quarter is 100, and the median is 38. Find the value of the upper and lower quartile.

Given

$BS_k = 0.6$ $Q_3 + Q_1 = 100$ $Q_2 = 38$

$$\boxed{Q_1 = 100 - Q_3}$$

$$BS_k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$= \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - (100 - Q_3)} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 + Q_3 - 100}$$

$$\boxed{BS_k = \frac{(Q_3 + Q_1) - 2Q_2}{2Q_3 - 100}}$$

$$0.6 = \frac{100 - 2(38)}{2Q_3 - 100}$$

$$2Q_3 - 100 = \frac{100 - 2(38)}{\cancel{0.}6 \times 10^{-1}}$$

$$2Q_3 - 100 = \frac{\cancel{24}^{4} \times 10}{\cancel{6}}$$

$$2Q_3 - 100 = 40$$

$$2Q_3 = 140$$

$$\boxed{Q_3 = 70}$$

$Q_1 + Q_3 = 100$

$Q_1 + 70 = 100$

$Q_1 = 100 - 70$

$$\boxed{Q_1 = 30}$$

$\therefore$ upper Quartile $= 30$

lower Quartile $= 70$

~~Unit~~ 3

UNIT 3

Correlations

~~Co~~ Correlation is one which studies the relationship between two or more variables. The degree of relationship between the variables is measured by ~~co~~ correlation technique.

Properties of Correlation Coefficient :

It measures the direction and degrees of relationship between ~~th~~ the relation.

① It is independant of the choice of ~~both~~ origin and the scale of observation.

② It is independant of the unity of ~~the~~ measurement.

③ It lies between −1 and +1 ; (or $-1 \le r \le 1$)

Karl Pearson's ~~coefficient~~ Method :

$$r = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$

cov(x,y) = ~~Coefficient of variation~~ Covariance of x and y.

$$cov(x,y) = \frac{\Sigma XY}{N}$$

$$\therefore \boxed{r = \frac{\Sigma XY}{N\sigma_x \sigma_y}}$$

$$\cancel{\sigma_x} \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}$$

$$r = \frac{\Sigma xy}{N\sigma_x \sigma_y} \qquad x = X - \bar{X}, \quad y = Y - \bar{Y}$$

$$r = \frac{N\Sigma dxdy - \Sigma dx\, dy}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

Problem

① Calculate the correlation coefficient between x and y:

x	9	10	11	12	13	14	15	16	17	18
y	18	17	19	21	20	23	22	25	27	26

x	$dx = x-\bar{x}$	dx^2	y	$dy = y-\bar{y}$	dy^2	$dx\,dy$
9	-4	16	18	-3	9	12
10	-3	9	17	-4	16	12
11	-2	4	19	-2	4	4
12	-1	1	21	0	0	0
13	0	0	20	-1	1	0
14	1	1	23	2	4	2
15	2	4	22	1	1	2
16	3	9	25	4	16	12
17	4	16	27	6	36	24
18	5	25	26	5	25	25
135	5	85	218	8	112	93

$\frac{\Sigma x}{n} = \bar{x}$ $\quad \Sigma dx^2 = 85$

$\boxed{\bar{x} = 13}$ $\quad N = 9$ $\quad \bar{Y} = 21$

$$r = \frac{N\Sigma dx\,dy - \Sigma dx\,\Sigma dy}{\sqrt{N\Sigma dx^2 - \Sigma(dx)^2}\sqrt{N\Sigma dy^2 - \Sigma(dx)^2}}$$

$$= \frac{10(93) - (5)(8)}{\sqrt{10(85) - 25}\sqrt{10(112) - 64}}$$

$$= \frac{930 - 40}{\sqrt{850 - 25}\sqrt{1120 - 64}} = \frac{890}{(28.72)(32.496)}$$

$$= 0.9535.$$

(2) The Marks of Commerce and Economics obtain by given students is given below. find the correlation coefficient.

Marks in Commerce	50	60	58	47	49	33	65	43	46	68
Marks in Eco	18	17	19	21	20	23	22	25	27	26

x	$dx = X-\bar{X}$	dx^2	y	$dy = y-\bar{y}$	dy^2	$dx\,dy$
$\boxed{50}$	0	0	18	-3	9	0
60	10	100	17	-4	16	-40
58	8	64	19	-2	4	-16
47	-3	9	$\boxed{21}$	0	0	0
49	-1	1	20	-1	1	1
33	17	289	23	2	4	34
65	15	225	22	1	1	15
43	-7	49	25	4	16	-28
46	-4	16	27	6	36	-24
68	-2	4	26	5	25	-10
519	32	757	218	8	112	-68

$\bar{X} = 50$. $\boxed{\Sigma dx\,dy = -68}$ $\boxed{\bar{Y}\ 21}$ $\Sigma dx = 33$ $\Sigma dy = 8$

$\boxed{N = 10}$ $\boxed{\Sigma dx^2 = 757}$ $\boxed{\Sigma dy^2 = 112}$ $\boxed{\Sigma dx\,\Sigma dy = 264}$

$$r = \frac{N\Sigma dx\,dy - \Sigma dx\,\Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} - \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{10(-68) - 264}{\sqrt{7570 - 1089} - \sqrt{1120 - 64}}$$

$$= \frac{-680 - 264}{\sqrt{6481} \times \sqrt{1056}}$$

$$= \frac{-944}{(80.50)(32.49)} = \frac{-944}{2615.4}$$

$$\boxed{r = -0.360}$$

$$\begin{array}{r} 99 \\ 99 \\ \hline 1089 \end{array}$$

$$\begin{array}{r} 7570 \\ 1089 \\ \hline 6481 \end{array}$$

$$\begin{array}{r} 1120 \\ 64 \\ \hline 1056 \end{array}$$

(3). If the covariance between the variable x and y is 10 and the variances of x and Y are 16 and 9 respectively. Find the correlation coefficient.

Given:

$$\text{Covariance of } x \text{ and } Y = \frac{\Sigma xY}{N} = 10.$$

Variance of $x = 16$ variance of $Y = 9$.

$\sigma_x = 4$ $\sigma_y = 3$

$$r = \frac{\Sigma xY}{N\sigma_x\sigma_y} = \frac{10}{4\times3} = \frac{10}{12} = \frac{5}{6} = 0.833$$

(4). Find the ~~correlation~~ Pearson's correlation coefficient for the following data.

demand	85	93	95	105	120	130	150	160
Prices	15	18	20	24	30	35	40	50.

x	$dx = x-\bar{x}$	dx^2	Y	$dy = y-\bar{y}$	dy^2	$dx\,dy$
85	-20	400	15	-15	225	300
93	-12	144	18	-12	144	144
95	10	100	20	-10	100	100
$\boxed{105}$.	0	0	24	-6	36	0
120	15	225	$\boxed{30}$	0	0	0
130	15	225	35	5	25	75
150	45	2025	40	10	100	450
160	55	3025	50	20	400	1100
$\bar{x} = 105$	98	6544	$\bar{Y} = 30$.	-8	1030	2169

$N = 8$ $\bar{x} = 105$

$\bar{Y} = 30$

$\Sigma dx = 98$ $\Sigma dx\,dy = 2169$.

$\Sigma dy = -8$

Karl Pearson's coefficient of coefficient

$$r = \frac{N\sum dx\,dy - \sum dx \sum dy}{\sqrt{N\sum dx^2 - (\sum dx)^2}\,\sqrt{N\sum dy^2 - (\sum dy)^2}}$$

$$= \frac{8(2169) - (98)(8)}{\sqrt{8(6549) - (98)^2}\sqrt{8(1030)(8)^2}} = \frac{16568}{\sqrt{52392 - 9604}\sqrt{8240 - 64}}$$

$$= \frac{16568}{\sqrt{42788}\,\sqrt{8176}}$$

$$= \frac{16568}{18695.1}$$

$$= 0.86221$$

$$\boxed{r = 0.86221}$$

Spearman's Rank correlation

• Ranks given • Ranks not given • Ranks repeated

This method is based on rank. In calculating rank correlation coefficient, Problems may be of three type.

- When ranks are given.
- When ranks are not given
- When repeated values are given.

In Type I :

Rank correlation co-efficient

$$\rho = 1 - \frac{6\sum d^2}{N(N^2-1)}$$; d = difference between pair of rank i.e (X−Y).

In Type II :

The values must be assigned Ranks, with rank 1 to the highest value, and rank 2 to the next highest and so on., followed by similar formulas as in Type I.

In Type III :

• Repeated Ranks

$$\rho = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}\sum M_i(M_i^2-1)\right]}{N(N^2-1)}$$

$M_i \rightarrow$ Average of repeated ranks.

$N \rightarrow$ No. of values given

PROBLEMS

① Calculate the rank correlation between the rank for given X and Y.

X	10	8	1	2	4	0	9	3	5	4	7
Y	6	10	5	4	8	3	1	2	9	8	7

X	Y	d = x − y	d^2
10	6	4	16
8	10	−2	4
1	5	−4	16
2	4	−2	4
0	3	−3	9
9	1	8	64
3	2	1	1
5	9	−4	16
4	8	−4	16
7	7	0	0
			146

$$\rho = 1 - \frac{6 \sum d^2}{N(N^2-1)}$$

$$\rho = 1 - \frac{6(146)}{10(99)}$$

$$= 1 - \frac{876}{990}$$

$$= 1 - 0.885$$

$$= 0.1152$$

(8) 10 competitors in a beauty are ranked by three judges.

Judge 1	2	7	1	5	3	4	8	6	10	9
Judge 2	10	6	3	8	7	2	9	5	4	1
Judge 3	2	5	6	9	1	3	7	4	8	10

Use rank correlation coefficient to determine which pair of judges has the nearest approach to common taste

R_1	R_2	R_3	(R_1-R_2) d_1	d_1^2	(R_2-R_3) d_2	d_2^2	(R_3-R_1) d_3	d_3^2
2	10	2	−8	64	8	64	0	0
7	6	5	·1	1	1	1	2	4
1	3	6	−4	16	−3	9	−5	25
5	8	9	−3	9	1	1	−4	16
3	7	1	−4	16	6	36	2	4
4	2	3	2	4	+1	1	1	1
8	9	7	1	1	−2	4	1	1
6	5	4	1	1	−1	1	2	4
10	4	8	6	36	−4	16	2	4
9	1	10	8	64	−9	81	−1	1
				212		214		60

$$\rho_{12} = 1 - \frac{6 \sum d_1^2}{N(N^2-1)}$$

$$= 1 - \frac{6(200)}{10(99)} = 1 - \frac{1200}{990}$$

$$= 1 - 1.212 = -0.212$$

$$\rho_{23} = 1 - \frac{6 \sum d_2^2}{N(N^2-1)}$$

$$= 1 - \frac{6(214)}{10(99)} = 1 - \frac{1284}{990}$$

$$= 1 - 1.297$$

$$= -0.297$$

$$\rho_{13} = 1 - \frac{6 \sum d_3^2}{N(N^2-1)}$$

$$= 1 - \frac{6(60)}{10(10^2-1)} = 1 - \frac{360}{990} = 1 - 0.364$$

$$= 0.6363$$

Among 3 judge,

1st and 3rd judge have nearest approach to common test.

x 68 64 75 50 64 80 75 40 55 64

y 62 58 68 45 81 60 68 48 50 70

~~X~~	Y	~~d = X - Y~~	~~d^2~~
68	62	6	
64	58	6	
75	48	27	
50	45	5	
64	81	19	
80	60	20	
75	68		
40	45		
55	50	5	
64	70		

75
48
27
81
64
17

X	Y	R_1	R_2	R_2
68	62	~~84~~	~~4~~	4
64	58	6	5	6
75	68	2.5	7	3.5
50	45	8	8.5	
64	81	6	1	1
80	60	1	5	5
75	68	2.5	3	3.5
40	48	9	8.5	
55	50	7	6	
64	70	6	2	2

(3) Calculate the ~~rank~~ correlation coefficients by rank for the foll. distribution.

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

x	Y	R(x)	R(Y)	d = R(x) - R(Y)	d^2
68	62	4	6	-2	4
64	58	6	8	-2	4
75	68	2.5	4.5	-2	4
50	45	9	2	7	49
64	81	6	1	5	25
80	60	1	7	-6	36
75	68	2.5	4.5	-2	4
40	48	10	10	0	0
55	50	8	9	-1	1
64	70	6	3	3	9
				0	136

R(X) has 2 tie of length 2 (2.5) 3 (6)

R(Y) has 1 tie of length 2 (4.5)

$$\frac{1}{12}\sum M_i(M_i^2-1) = \frac{1}{12}\left[2(4-1)+2(4-1)+3(9-1)\right]$$

$$= \frac{1}{12}\left[2(3)+2(3)+3(8)\right] = \frac{1}{12}\left[6+6+24\right]$$

$$= \frac{1}{12}\left[36\right] = 3.$$

∴ The Rank correlation coefficient.

$$\rho = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}\left[\sum m(m^2-1)\right]\right]}{N(N^2-1)}$$

$$= 1 - \frac{6\left[136+3\right]}{10(100-1)} = \frac{990 - 6(408)}{990} = \frac{990-2298}{990}$$

$$= \frac{1458}{990}$$

$$\boxed{\rho = -1.472}$$

1) The coefficient of Rank correlation of the marks obtained by 10 students in statics and commerce was found to be 0.8. It was later discovered that the difference in ranks between in 2 subjects obtained by one of the students was wrongly taken as 7 instead of 9. Find the correct coefficient of correlation in ranks.

Given:

$N = 10$, $\rho_{a+b} = 0.8$, $d_{a-b} = 7$, crt $d = 9$.

$$\rho = 1 - \frac{6 \sum d^2}{N(N^2-1)}$$

$$0.8 = 1 - \frac{6 \sum d^2}{10(10^2-1)} = 1 - \frac{6 \sum d^2}{990}$$

$$0.8 - 1 = - \frac{\cancel{6}^{2} \sum d^2}{\cancel{990}\ 330}$$

$$1 - 0.8 = \frac{\cancel{2} \sum d^2}{\cancel{330}\ 165}$$

$$165 \times 0.2 = \sum d^2 \qquad \boxed{\sum d^2 = 33}$$

$$\text{CRT } \sum d^2 = \sum d^2 - \text{wrong } d^2 + \text{correct } d^2.$$

$$= 33 - 7^2 + 9^2 = 33 - 49 + 81$$

~~CRT $\sum d^2$ = 33+2 = 35~~ $= 33 + 32 = 65$

$$\text{CRT } \sum d^2 = 1 - \frac{6 \text{ crt} \sum d^2}{N(N^2-1)} = 1 - \frac{6(65)}{10(100-1)} = 1 - \frac{\cancel{6}(65)}{\cancel{990}\ 165}$$

$$= 1 - \frac{65}{165}$$

$$\text{CRT } \rho = 1 - \frac{6 \text{ CRT} \sum d^2}{N(N^2-1)}$$

$$= 1 - \frac{6(0.606)}{10(99)} = \frac{99 - 3.636}{990}$$

$$= \frac{95.364}{990} = 0.0963.$$

~~$\frac{165 - 1}{165} = \frac{166}{165}$~~

$$= \frac{165 - 65}{165}$$

~~$= 1.000006$.~~

$$= \frac{100}{165} = 0.606$$

(5) The co-efficient of rank correlation is 0.5. No. of students is 10. Difference in rank was wrongly taken as 3 instead of 7. Find the rank correlation coefficient.

Given:

$\rho = 0.5$ $N = 10$ wrong $d = 3$ crt $d = 7$.

$\rho = 1 - \frac{6\sum d^2}{N(N^2-1)}$

$0.5 = 1 - \frac{6\sum d^2}{990}$ (165)

$1 - 0.5 = \frac{\sum d^2}{165}$

$0.5 \times 165 = \sum d^2$

$\boxed{82.5 = \sum d^2}$

crt $\sum d^2 = \sum d^2 - (\text{wrong } d)^2 + (\text{right } d)^2$

$= 82.5 - 9 + 49$

$= 82.5 + 40 = 122.5$

crt $\rho = 1 - \frac{6(\text{crt} \sum d^2)}{N(N^2-1)} = 1 - \frac{6(122.5)}{990}$

$= 1 - \frac{122.5}{165}$

$= 1 - 0.7424$

$= 0.258$

Regression

Regression of Y on X

$$b_{yx} = \frac{N \sum dx dy - \sum dx \sum dy}{\sqrt{N \sum dx^2 - (\sum dx)^2}} = \frac{r\sigma_y}{\sigma_x} \quad y - \bar{y} = b_{yx}(x - \bar{x})$$

Regression of X on Y

$$b_{xy} = \frac{N \sum dx dy - \sum dy \sum dx}{\sqrt{N \sum dy^2 - (\sum dy)^2}} = \frac{r\sigma_x}{\sigma_y} \quad x - \bar{x} = b_{xy}(y - \bar{y})$$

$r = \sqrt{b_{xy} \cdot b_{yx}}$

Here, b_{xy} & b_{yx} = regression coefficient $\qquad$ correlation coeffion.

Regression is the statistical method with which we can estimate unknown value of one variable from the known value of another variable.

Regression equation of Y on X

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

Calculate regression equation

X	Y
6	9
2	11
10	5
4	8
8	7

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

$$b_{yx} = \frac{N\sum dx\,dy - \sum dx \sum dy}{\sqrt{N\sum dx^2 - (\sum dx)^2}}$$

$dx = X - 6$. $dy = Y - 8$

X	Y	$dx = X-\bar{X}$	$dy = Y-\bar{Y}$	$dx\,dy$	dx^2
					0
6	9	0	1	0	16
2	11	-4	3	-12	16
10	5	+4	-3	-12	4
4	8	-2	0	0	4
8	7	+2	-1	-2	40
$\sum x = 30$	$\sum y = 40$	0	0	-26	

-12-12

$\bar{X} = \frac{\sum x}{n} = \frac{30}{5} = 6$

$\bar{Y} = \frac{\sum y}{n} = \frac{40}{5} = 8$

$$b_{yx} = \frac{N\sum dx\,dy - \sum dx \sum dy}{\sqrt{N\sum dx^2 - (\sum dx)^2}}$$

$$b_{yx} = \frac{5(-26) - 0}{5(40) - 0^2}$$

$$= \frac{-130}{200}$$

$$= -0.65.$$

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 8 = -0.65(x - 6)$$

$$y - 8 = -0.65x + 3.9$$

$$y = -0.65x + 3.9 + 8.$$

$$\boxed{y = -0.65x + 11.9}.$$

X	Y	$dx = X - 18$ $dx = x - \bar{x}$	$dy = y - \bar{y}$	$dx\,dy$
10	5	-8	-7	56
12	6	-6	-6	36
13	7	-5	-5	25
17	9	-1	-3	3
18	13	0	1	0
20	15	2	3	6
24	20	6	8	48
30	21	12	9	108
$\Sigma x = 144$	$\Sigma y = 96$	0	0	$\Sigma dx dy = 282$

$\bar{X} = \frac{\Sigma x}{n}$ $\quad \bar{Y} = \frac{\Sigma y}{n}$

$\bar{X} = \frac{144}{8}$ $\quad \bar{Y} = \frac{96}{8}$

$\bar{X} = 18$ $\quad \bar{Y} = 12$

dx^2	dy^2
64	49
36	36
25	25
1	9
0	1
4	9
36	64
144	81
Σdx^2 310	$dy^2 = 274$

$$b_{xy} = \frac{N\Sigma dx dy - \Sigma dx \Sigma dy}{N\Sigma dy^2 - (\Sigma dy)^2}$$

$$= \frac{8(282) - 0}{8(274) - 0}$$

$$= \frac{2256}{2192} = 1.029$$

Scanned by CamScanner

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

$$X - 18 = 1.029(Y - 12)$$

$$X - 18 = 1.029Y - 12.348$$

$$X = 1.029Y - 12.348 + 18.$$

$$X = 1.029Y + \cancel{30.348}\ 5.652$$

Estimated value of x when y = 26.

$$X = 1.209(26) + 5.652 = 32.406.$$

$$\cancel{= 37.086}$$

Regression equation of Y on X.,

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

$$b_{yx} = \frac{\cancel{N}\ \Sigma dydx}{\cancel{N}\ \Sigma dx^2} \quad (\because \Sigma dx \Sigma dy \text{ and } \Sigma dx^2 = 0)$$

$$= \frac{282}{310} = 0.91.$$

$$Y = 0.91(x - \bar{x})$$

$$\boxed{Y = 0.91X - 16.38}$$

Scanned by CamScanner

Estimate x when Y = 26.

Calculate the coefficient of correlation also

X	Y
10	5
12	6
13	7
17	9
18	13
20	15
24	20
30	21

~~26475~~4

x	$dx = x-\bar{x}$	x^2	Y	$y = Y-\bar{Y}$	y^2	xy
10	-8	64	5	-7	49	56
12	-6	36	6	-6	64	36
13	-5	25	7	-5	25	25
17	-1	1	9	-3	9	3
18	0	0	13	1	1	0
20	2	4	15	3	9	~~12~~6
24	6	36	20	8	~~16~~64	48
30	12	144	21	9	81	108
$\boxed{\Sigma x = 144}$	0	310	$\boxed{\Sigma Y = 96}$	0	2~~54~~74	282

$\frac{\Sigma x}{n} = \frac{\cancel{144}}{8} = 18$ $\qquad$ $\frac{\Sigma Y}{n} = \frac{\cancel{96}}{8} = 12$

~~Regression Equation of~~

Regression Equation of X on ~~x~~Y,

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

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

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

$(\because dx = 0 \;\&\; dy = 0)$.

$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{282}{274} = 1.029$

$\therefore \boxed{X = 1.029(Y-12)+18}$ $\qquad$ $\therefore \boxed{X = 32.406, \text{ if } Y=26}$

If $Y = 26$, $X = 1.029(26-12)+18 = 1.029(14)+18 = 14.406+18$

Obtain the lines of regression from the following data.

X	4	5	6	8	11
Y	12	10	8	7	5

Verify that the coefficient of correlation is the G.M of regression coefficients.

x	$dx = x-\bar{x}$	dx^2	y	$dy = y-\bar{y}$	dy^2	$dxdy$
4	-2	4	12	4	16	-8
5	-1	1	10	2	4	-2
$\boxed{6}$	0	0	$\boxed{8}$	0	0	0
8	2	4	7	-1	1	-2
11	5	25	5	-3	9	-15
34	4	34	42	2	30	-27

$\Sigma x = 34$ $\qquad$ $\Sigma y = 42$

$\frac{\Sigma x}{n} = \bar{x} = \frac{34}{5} = 6.8$ $\qquad$ $\frac{\Sigma y}{n} = \frac{42}{5} = 8.4$

$$b_{xy} = \frac{N\Sigma dydx - \Sigma dx\,\Sigma dy}{N\Sigma dy^2 - (\Sigma dy)^2}$$

$$\cancel{= \frac{5(-27) - (2)(4)}{5(34) - (4)^2}} = \frac{-135-6}{\sqrt{170-16}} = \frac{-141}{\sqrt{154}}$$

$$= \frac{5(-27) - (2)(4)}{5(30) - (2)^2} = \frac{-135-6}{150-4} = \frac{-141}{146} = -0.979$$

$$b_{yx} = \frac{N\Sigma dxdy - \Sigma dy\,\Sigma dx}{N\Sigma dx^2 - (\Sigma dx)^2}$$

$$= \frac{5(-27) - (2)(4)}{5(34) - (4)^2} = \frac{-141}{154} = -0.929$$

Regression equation of Y on X ;

$Y - \bar{Y} = b_{yx}(X - \bar{X})$ $\qquad$ $\bar{x} = 6.$ $\quad$ $\bar{y} = 8.$

$Y - 8 = -0.929(X-6)$

$Y - 8 = -0.929X + 5.574$

$\cancel{Y = -0.929X - 8.574}$

$\Rightarrow Y = -0.929X + 13.574$

Regression of x on y

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 6 = -0.979(y - 8)$$

$$\Rightarrow x - 6 = -0.979y + 7.832$$

$$\boxed{x = -0.979y + 13.832}$$

$\boxed{b_{xy} = -0.979}$ // $\boxed{b_{yx} = -0.929}$

$$r = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{(-0.979)(-0.929)}$$

$$= -\sqrt{(0.979)(0.929)} = -0.954.$$

$$r = \frac{N\sum dx\,dy - \sum dx \sum dy}{\sqrt{N\sum dx^2 - (\sum dx)^2} \cdot \sqrt{N\sum dy^2 - (\sum dy)^2}}$$

$$= \frac{(5)(-27) - (4)(2)}{\sqrt{154} \cdot \sqrt{146}} = \frac{-141}{149.94} = -0.954$$

~~Both~~ The

∴ Verified that the G.M of regression coefficients ~~and~~ is the coefficient of correlation

(4). You are given below the following information about advertising and sales

Correlation coefficient = 0.8

	adv. Exp (X)	Sales (Y)
Mean	10	90
S.D	3	12.

(i) Obtain the two regression lines.

(ii) Find the likely sales when advertisement expenditure is Rs.15 lakhs.

(iii) What should be advertisement expenditure if the company wants to ~~attain~~ attain sales target of Rs. 120 lakhs?

Given;

$\bar{x} = 10$ $\bar{y} = 90$ $\sigma_x = 3$ $\sigma_y = 12.$ $r = 0.8.$

To find;

(i) Equations (ii) x if $y = 15$ (iii) y if $x = 120.$

solution

~~(i) Equation~~

$\cancel{x - \bar{x} = b_{xy}\, y - \bar{y}}$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = 0.8\,\frac{3}{12} = \left(\frac{0.2}{0.8}\right) \cancel{to} \left(\frac{2}{4}\right) = 0.2.$$

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

$$= (0.8)\,\frac{12}{3} = (0.8\times4) = 3.2.$$

$x - \bar{x} = b_{xy}(y - \bar{y})$ $y - \bar{y} = b_{yx}(x - \bar{x})$

$x - 10 = (0.2)\,y - (0.2)(90)$ $y - 90 = (3.2)(x - 10)$

$x - 10 = 0.2y - 18$ $y - 90 = 3.2x - 32$

(i) $\boxed{x = 0.2y - 8}$ $\boxed{y = 3.2x + 58}$

(ii) If adv. expenditure is 15 lakhs (or $x = 15$).

$y = 3.2(15) + 58$

$= 48 + 58 = 106.$ $\boxed{y = 106}$ Sales target will be 106 lakhs.

(iii) If sales target is 120 lakhs (or $y = 120$).

$x = 0.2(120) - 8.$

$= 24 - 8 = 16$ $\boxed{x = 16}$ The advertisement expenditure will be 16 lakhs

---

⑤ From 10 observations on price x and supply y of a comodity, the following figures were obtained.

$\Sigma x = 130$ $\Sigma y = 220$ $\Sigma x^2 = 2288$ $\Sigma y^2 = 5506$ $\Sigma xy = 3467.$

Compute the line of regression of y on x and estimate the supply when the price x is 16.

$\boxed{N = 10}$

$$b_{xy} = \frac{N\Sigma xy - \Sigma x\,\Sigma y}{N\Sigma y^2 - (\Sigma y)^2} = \frac{34670 - (130)(220)}{55060 - (220)^2}$$

$$= \frac{6070}{6660} = 0.9114$$

Given;

$\bar{x} = 10 \quad \bar{y} = 90 \quad \sigma_x = 3 \quad \sigma_y = 12 \quad r = 0.8$

To find;

(i) Equations (ii) x if Y = 15 (iii) Y if x = 120

Solution

~~(i) Equation~~

$\cancel{x - \bar{x} = b_{yx}\, y - \bar{y}}$

$b_{xy} = r \dfrac{\sigma_x}{\sigma_y} = 0.8 \dfrac{3}{12} = 0.2$

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

$= (0.8) \dfrac{12}{3} = (0.8 \times 4) = 3.2$

$x - \bar{x} = b_{xy}(y - \bar{y}) \qquad y - \bar{y} = b_{yx}(x - \bar{x})$

$x - 10 = (0.2)y - (0.2)(90) \qquad y - 90 = (3.2)(x - 10)$

$x - 10 = 0.2y - 18 \qquad y - 90 = 3.2x - 32$

(i) $\boxed{x = 0.2y - 8} \qquad \boxed{y = 3.2x + 58}$

(ii) If adv. expenditure is 15 lakhs (or x = 15).

$y = 3.2(15) + 58$

$= 48 + 58 = 106. \quad \boxed{y = 106}$ Sales target will be 106 lakhs

(iii) If sales target is 120 lakhs (or y = 120)

$x = 0.2(120) - 8$

$= 24 - 8 = 16 \quad \boxed{x = 16}$ The advertisement expenditure will be 16 lakhs

(5) From 10 observations on price x and supply y of a commodity, the following figures were obtained.

$\Sigma x = 130 \quad \Sigma y = 220 \quad \Sigma x^2 = 2288 \quad \Sigma y^2 = 5506 \quad \Sigma xy = 3467$

Compute the line of regression of y on x and estimate the supply when the price x is 16.

$\boxed{N = 10}$

$b_{xy} = \dfrac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma y^2 - (\Sigma y)^2} = \dfrac{34670 - (130)(220)}{55060 - (220)^2}$

$= \dfrac{6070}{6660} = 0.9114$

Regression Equation of Y on X

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

$$b_{yx} = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} = \frac{34670 - (130)(220)}{\cancel{55600}\ 22800 - (130)^2}$$

$$= \frac{34670 - 28600}{\cancel{55600}\ 22800 - 16900} = \frac{6070}{5980} = 1.015$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{220}{10} = 22.$$

$$\bar{X} = \frac{\sum X}{N} = \frac{130}{10} = 13.$$

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

$$\Rightarrow Y - 22 = 1.015(X - 13)$$

$$\Rightarrow Y = 1.015X - 13.2 + 22$$

$$\Rightarrow Y = 1.015X + 8.8$$

If $X = 16$,

$$Y = 1.015(16) + 8.8.$$

$$= 16.24 + 8.8$$

$$\boxed{Y = 25.04}$$

(6). The following gives the aptitute test scores and productivity indices of 10 workers selected at random

Aptitute scores (X)	60	62	65	70	72	48	53	73	65	82
Productivity index (Y)	68	60	62	80	85	40	52	62	60	81

Find the two lines of regression and estimate:

(i) the productivity index of a worker whose test score is 92

(ii) The test score of a worker whose productivity is 75.

~~UNIT - 4~~

~~PROBABILITY THEORY.~~

~~EVENT:~~

~~When an experiment is performed, the the collection of possible outcome is called an event.~~

The Given data is tabulated here:

X	$dx = X-\bar{X}$	dx^2	Y	$dy = Y-\bar{Y}$	dy^2	$dxdy$
60	-5	25	68	3	9	-15
62	-3	9	60	-5	25	15
$\boxed{65}$	0	0	62	-3	9	0
70	5	25	80	15	225	75
72	7	49	85	20	400	140
48	-17	289	40	-25	625	425
53	-12	144	52	-13	169	156
73	8	64	62	-3	9	-24
65	0	0	60	-5	25	0
82	17	289	81	16	256	272
650	0	894.	650	0	1752	1044.

$N=10$, $\bar{X}=65$, $\bar{Y}=65$, $\Sigma dx^2 = 894$, $\Sigma dxdy = 1044$.

$\Sigma dx = 0$, $\Sigma dy = 0$, $\Sigma dy^2 = 1752$.

~~Regression Equation of~~

$$b_{xy} = \frac{N\Sigma dxdy - \Sigma dx\Sigma dy}{N\Sigma dx^2 - (\Sigma dx)^2} \qquad b_{xy} = \frac{N\Sigma dxdy - \Sigma dx\Sigma dy}{N\Sigma dy^2 - (\Sigma dy)^2}$$

$$= \frac{10440}{8940} = 1.168 \qquad = \frac{10440}{17520} = 0.596.$$

Regression Equation of X on Y,

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

$$X - 65 = 0.596(Y-65)$$

$$X = 0.596Y - 38.74 + 65$$

$$\boxed{X = 0.596Y + 26.26}$$

Regression Equation of Y on X,

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

$\Rightarrow Y - 65 = 1.168(X - 65)$

$\Rightarrow Y - 65 = 1.168X - 75.92$

$\boxed{Y = 1.168X - 10.92}$

(i) To find Y, if $X = 92$.

$Y = 1.168(92) - 10.92$

$Y = 107.456 - 10.92$

$\boxed{Y = 96.536}$

$\boxed{Y \simeq 97}$

(ii) To find X, if $Y = 75$.

~~$X = 0.596Y$~~

$X = 0.596Y - 26.26$

$= 0.596(75) - 26.26$

$= 44.7 - 26.26$

$\boxed{X = 18.44}$.

$\boxed{X \simeq 18}$.

UNIT 4

PROBABILITY THEORY

Events :

An Event A is any collection of possible outcomes, when an experiment is performed.

Mutually Exclusive :

A and B are the mutually exclusive events when both cannot happened simultaniously in a single trial.

Then,
Probability of A (OR) B,

$$P(A\cup B) = P(A)+P(B). \quad \because P(A\cap B) = \phi$$

Independant Events :-

If the occurance of one event does not affect the occurance of another event, then those events are said to be independant events ;

That is ;
$$P(A\cap B) = P(A).P(B)$$

Probability :

An Event A can happened in m ways out of n possible outcomes ;

then probability of A,

$$P(A) = \frac{m}{n} \quad \text{or} \quad P(A) = \frac{n(A)}{n(S)}$$ where S-sample space

$$P(\bar{A}) = 1-P(A) = 1-\frac{m}{n}.$$

Properties :

- For any event A, $0 \leq P(A) \leq 1$
- $P(A)=0$, if and only if 'A' is an impossible event.
- $P(S)=1$, so known as total probability ;

Addition theorem:

If E_1 and E_2 are any two events, then probability of E_1 and E_2,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

Conditional Probability:

If E_1 and E_2 are events, the probability ~~that~~ E_2 occurs, given that E_1 has already occured, is denoted by $P(E_2/E_1)$. It is called conditional probability.

If these events were independant, then

$$P(E_2/E_1) = P(E_2).$$

Multiplication Theorem:

$$P(E_1 \cap E_2) = P(E_1) . P(E_2/E_1)$$
$$= P(E_2) . P(E_1/E_2).$$

If E_1 and E_2 Events were independant:

$$P(E_1 \cap E_2) = P(E_1) . P(E_2) \quad \left[\because P(E_2/E_1) = P(E_2)\right]$$

Baye's Theorem:

If $A_1, A_2, \ldots, A_n$ are arbitarily mutually exclusive, equally likely and exhausive events, then for any arbitary event B, which is the subset of $\bigcup_{k=1}^{n} A_k$, we have

$$P(A_k/B) = \frac{P(A_k) . P(B/A_k)}{\sum_{i=1}^{n} P(A_k) . P(B/A_k)} \;//.$$

Equally likely events: When one does not occurs ~~more~~ more often than others.

PROBLEMS

1) If a card is drawn from a pack containing 52 cards. Find the chance
(a) the card is spade. (b) the card is ace.

From the given data, $n(S) = 52$.

(a) The card is spade

no. of cards of spade = 13.

The propability P(the card is spade) = $\frac{13}{52} = \frac{1}{4}$

(b) The card is ace.

no. of cards of ace = 4

The propabality P(getting ace) = $\frac{4}{52} = \frac{1}{13}$

2) Find the propability of getting 3 white balls in a drawn of 3 balls from 5 ~~balls~~ white balls and 4 black balls.

Total number of balls = 5+4 = 9 no. of ways of getting 3 balls = $9C_3$

No. of ways of getting 3 ~~balls~~ white balls from 5 white balls = $5C_3$.

The propability of getting 3 white balls in a draw

$$P(A) = \frac{5C_3}{9C_3} = \frac{5\times4\times3}{9\times8\times7} = \frac{5}{42}$$

3) Five cards are drawn from the pack. Find the propability that
(a) 4 are aces; (b) 4 ace and 1 king; (c) 3 tens and 2 Js
(d) 9, 10, J, Q and K are obtained (e) At least 1 ace is obtained.

Total number of cards = 52.

No. of ways of getting 5 cards from the pack = $52C_5$.

(a) P(4 are aces)

$n(A)$ = no of ways getting 4 aces × no. of getting 1 card.

$= 4C_4 \times 48C_1$.

$$P(A) = \frac{n(A)}{n(S)} = \frac{4C_4 \times 48C_1}{52C_5} = \frac{48}{52\times51\times50\times49\times48} = \frac{1}{54145}$$

(b) 4ace and 1king.

$n(A)$ = no. of ways of getting 4ace × 1 king

$$p(A) = \frac{4C_4 \times 4C_1}{52C_5} = \frac{4}{52C_5} = \frac{1}{13\times51\times50\times49\times48}$$

(c) 3tens and 2 Jacks.

$n(A)$ = no. of ways of getting 3 10s × 2Jacks.

$$= 4C_3 \times 4C_2$$

$$P(A) = \frac{4C_3 \times 4C_2}{52C_5} = \frac{4\times6}{52C_5} = \frac{4\times6}{2598960} = \frac{1}{108290}$$

(d) 9, 10, J, Q and K are obtained.

No. of ways of getting those cards =

$$n(A) = 4C_1 \times 4C_1 \times 4C_1 \times 4C_1 \times 4C_1$$

$$~~DES~~\ P(A) = \frac{n(A)}{n(S)} = \frac{4C_1 \times 4C_1 \times 4C_1 \times 4C_1 \times 4C_1}{52C_5} = \frac{4\times4\times4\times4\times4}{2598960} = \frac{64}{162435}$$

(e) Atleast 1ace is obtained.

$$P(A) = 1 - P(\bar{A})$$

P(atleast 1 ace is obtained) = 1 − P(not getting Ace)

no. of ways of not getting ace = no. of ~~48C5~~ ways of getting 5 cards from 48 cards.

$$= 48C_5.$$

$$\therefore P(A) = 1 - P(\bar{A})$$

$$= 1 - \frac{48C_5}{52C_5} = \cancel{1 - \frac{1712804}{2598960}} = 1 - 0.$$

$$= 1 - \frac{48\times47\times46\times45\times44}{52\times51\times50\times49\times48} = 1 - \frac{47C_4}{52C_4} = 1 - \frac{178365}{54145}$$

$$= \frac{54145 - 178365}{54145}$$

$$= \frac{18472}{54145}$$

4) In a single throw of 2 dice, what is the probability of getting (a) a total of 8. (b) a total different from 8.

(a). No. of ways of getting total 8 from 2 dice :

$= \{(6,2), (2,6), (3,5), (5,3), (4,4)\}$

Total number of ways to 2 dice $= 6^2 = 36$ ways.

$\therefore$ P(getting a total of 8) $= \frac{n(A)}{n(S)} = \frac{5}{36}$
P(A)

(b) No. of ways of total different from 8

= no. of ways of not getting difference of 8. $= n(\bar{A})$

$\therefore \quad P(\bar{A}) = 1 - P(A)$

$= 1 - \frac{5}{36} = \frac{31}{36}$.

5) There are 4 hotels in a city. If 3 mens check into hotels in a day. What is the propability they check each into a different hotels?

Posibilities of 3 mens check into those hotels: $4 \times 4 \times 4 = 64$ ways.

No. of ways of ~~getti~~ 3 mens to get into those hotels in a day

$= 4 \times 3 \times 2 = 24$ ways.

$\therefore$ The propability of getting checked by each hotel

$P(A) = \frac{n(A)}{n(S)} = \frac{24}{64} = \frac{3}{8}$.

6) From 7 Indians and 4 Japanese, a comitte of 5 is to be formed. What is the propability of forming a comittee with atleast 2 Japanees.

Total no. of members $= 7 + 4 = 11$ members.

Out of them 5 is chosen and 2 can be from Japanese.

The posibilites are :

	Japaneese	Indians.
Case (a)	2	3
Case (b)	3	2
Case (c)	4	1

Case (a): 2 Japanese 3 Indian
no. of ways of selecting 5 members from above alternative is,

$$n(a) = 4C_2 \times 7C_3$$

Case (b): 3 Japanese 2 Indian.

no. of ways, $n(b) = 4C_3 \times 7C_2$.

Case (c): 4 Japanese 1 Indian.

no. of ways, $n(c) = 4C_4 \times 7C_1$

$\therefore$ Total no. of ways $n(A) = n(a) + n(b) + n(c)$

$= 4C_2 \times 7C_3 + 4C_3 \times 7C_2 + 4C_4 \times 7C_1$

$= 210 + 84 + 7 = 301$

$n(S) = 11C_5$.

$\therefore$ Propability $P(A) = \frac{n(A)}{n(S)} = \frac{301}{11C_5}$.

7) A can solve 75% of problems in a book and B can solve in 70%. What is the propability that either A or B can solve a problem chosen at random?

$P(A \text{ is solved})$, $P(A) = 0.75$

$P(B \text{ is solved})$, $P(B) = 0.70$

P_1 and P_2 are not mutual exclusive, but they are independent

$P(A \cap B) = P(A).P(B)$

$= 0.75 \times 0.70$

$= \frac{75 \times 70}{100 \times 100} = \frac{21}{40} = 0.525$

The propability that either A or B solve the problem,

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 0.75 + 0.70 - 0.525$

$= 1.45 - 0.525 = 0.925$

Three group of children consists of 2 girls and 1 boy, 2 girls and 2 boys, 1 girl and 3 boys. One child is chosen at random from each group. Find the chance of that 3 selected consists of 1 girl and 2 boys.

The possibilities are (with order to groups).

(A) 1 boy, 1 boy, 1 girl $\Rightarrow$ $1C_1$, $2C_1$, $1C_1$

(B) 1 girl, 1 boy, 1 boy $\Rightarrow$ $3C_1$, $2C_1$, $3C_1$

(C) 1 boy, 1 girl, 1 boy $\Rightarrow$ $1C_1$, $2C_1$, $3C_1$

Total no. of child in each group = 4.

Propability of case A

$$P(A) = \frac{1}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{\cancel{2}\ 1}{32}.$$

Propability of case B

$$P(B) = \frac{\cancel{3}}{4} \times \frac{\cancel{2}}{\cancel{4}_2} \times \frac{3}{4} = \frac{3}{32}.$$

Propability of case C

$$P(C) = \frac{3}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{9}{32}.$$

∴ The total propability of getting 1 girl and 2 boys

$P(A \cup B \cup C) = P(A) + P(B) + P(C)$ (∵ A, B, C are mutually exclusive)

$$= \frac{1}{32} + \frac{3}{32} + \frac{9}{32} = \frac{13}{32}$$ //.

Q) 20 students of a class appeared in Economics and statistics. Out of these, 8 passed in Economics, 7 passed in Statistics and 8 failed in both subjects. If one student is selected, find the propability that a student

(a) passed in both subjects. (b) failed in both subjects

(c) failed in economics or statistics.

From the data,

$n(S) = 20$

$n(A) = 8$ (economics) $n(B) = 7$ (statistics.

$$P(A) = \frac{n(A)}{n(S)} = \frac{8}{20} \qquad P(B) = \frac{7}{20}.$$

$\cancel{P(\overline{A \cap B}) = \frac{8}{20}}$ ∴ ~~n(A∩B) =~~

$n(\overline{A \cap B}) = 8$

Then $\boxed{n(A \cup B) = 12}$

$\cancel{P(\overline{A \cap B}) = 1 - P(\overline{A \cap B})}$

$\cancel{= 1 - \frac{8}{20} = \frac{20-8}{20} = \frac{12}{20} = \frac{3}{5}}$

$n(A \cup B) = n(A) + n(B) - n(A \cap B)$

$n(A \cap B) = n(A) + n(B) - n(A \cup B)$

$= 8 + 7 - 12$

$= 15 - 12 = 3$

$\boxed{n(A \cap B) = 3}$ ∴ no. of students passed in both is 3.

(a) $P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{3}{20}.$

(b) $P(\overline{A \cap B}) = \frac{n(\overline{A \cap B})}{n(S)} = \frac{8}{20}$

(c) $P(\overline{A \cup B}) = \cancel{1 - P(A \cup B)}$

$= \frac{n(\overline{A \cup B})}{n(S)}$ $\cancel{= 1 - \frac{12}{20} = 1 - \frac{3}{5} = \frac{2}{5}}$

$= \frac{n(\bar{A}) + n(\bar{B}) - n(\bar{A} \cap \bar{B})}{n(S)} = \frac{12 + 13 - 8}{20} = \frac{17}{20}.$

(b) What is the chance that the leap year is selected at random will contain 53 sundays?

A leap year has 366 days, i.e 52 weeks and 2 days. That 2 days may be of

n(A) = { Monday & Tuesday
Tuesday & wednesday
wednesday & Thursday
Thursday & friday
Friday & saturday
Saturday & Sunday
Sunday & Monday }

From the possible combinations, 2 of them are perfect that can fit 53 Sundays, i.e {Saturday & sunday, Sunday & monday}.

∴ The propability of getting 53 sundays in a leap year

$$P(A) = \frac{n(A)}{n(S)} = \frac{2}{7}.$$

~~ii) Problem in statistics is given to two students A and B. The odd in favour~~

ODD AGAINST = Q . P   Here P & Q are two events.

EVEN AGAINST = P . Q

P → Success event.

Q → Failure event

$P(\text{happening}) = P/q+p$ $q(\text{happening}) = q/p+q$

ii) A problem in statistics is given to 2 students A and B. The odd in favour of A solving the problem are 5 to 10 and against B solve the problem is 8 to 16. If A and B attempt, find the propability that the problem is solved.

ODD favour of A = 5 : 10

ODD against of B = 16 : 8

Propability of A solving problem (happening) $= \frac{5}{5+10} = \frac{5}{15} = \frac{1}{3}$.

Probability of B solving problem (not happening) $= \frac{16}{16+8} = \frac{16}{24} = \frac{2}{3}$

∴ P(the problem solved by B) $= 1 - \frac{2}{3} = \frac{1}{3}$.

∴ P(A & B is solved) $= \frac{1}{2} + \frac{1}{3} = \frac{2}{3}$.

(11) A problem in statistics is given to two students A and B. The odds in favour of A solving a problem are 5 to 10 and against B solving the problems 8 to 6. If A and B attempt, find the probability that the problem being solved. [27/35]

(12) A problem in mechanics is given to three students A, B, C whose chance of solving it are 1/2, 1/3, 1/4 respectively. Find the propability that are solved. [3/4]

(13) Three balls are drawn successfully from the box containing 6 red, 4 white and 5 blue balls. Find the propability that they are drawn in the order red, white and blue if each ball is (a) replaced. (8/225) (b) not replaced (4/91)

(14) Find the propability ~~that~~ of picking one black one white and one red from a set of carrom coins when 3 coins are taken out of it. (27/323)

(15) Find the odd against 8 to 4 in a single toss of a fair die. [2:1]

(16) There are 3 urns, Urn I contains 3 red and 7 green balls, Urn II contains 5 red and 3 green balls and Urn III contains 8 red and 4 green balls. One ball is drawn from one of the Urns and it is found to be red. What is the probability that it comes from

(a) Urn I &
(b) Urn IV } [36/191]

(17) In a bolt factory machines A, B, and C manufacture 25%, 35% and 40% of bolts respectively. Of the total of their outputs 5, 4 and 2% are defective bolts. A bolt is drawn at random from the product and it is found to be defective. What is the probability that it was manufactured by A? [0.362]

(18) A player tosses two fair coins. He wins ₹5.00 if 2 head occurs, ₹2.00 if one head occurs, and ₹1.00 if no head occurs. Find his expected winning.

(19) A bag contains 2W and 3B balls. Four persons A, B, C, D in the order drawn 1 ball and do not replace it. The person to draw first ball received ₹200. Determine their expectations.

(20) A man draws balls from a bag containing 3W and 5B balls. If he is to receive ₹4 for every W ball which he draws and ₹7 for each B ball, what is his expectation?

Scanned by CamScanner

UNIT 5

TIME SERIES

Definition:

Time series is the set of statistical observation arranged in chronological order. It consists of statistical data which are collected, observed over successive increments.

Uses of Time Series:

* It helps ~~the~~ in understanding past behaviour.
* It helps in planning future operations.
* It helps in evaluating current accomplishments.
* It facilitates comparison.

Components of time series:

It is ~~cus~~ customary to classify the fluctuation of a time series into four basic types of variations.

1. Secular trend
2. Seasonal variation
3. Cyclic variation
4. Irregular variation.

(i) Secular trend:

It refers to the general direction and moment of the time series considering a fairly long period to time.

(ii) Seasonal trend:

It refers to a recurrent pattern of change within the period. Then result of from the operation of forces connected with ~~climate~~ climate or custom of different times of period. It is measured using trends within a calander year, that may be hourly ~~monthly~~, daily, weekly, monthly, quatery or half-yearly.

(iii) Cyclic variation:

Like seasonal variation, they are oscillatory moment in the time series with ~~which~~ the period ~~factor~~ ~~Exon~~ of oscillation ~~2~~ year grater than one year.

(iv) Irregular Variation:

The moments are some episodic or accidently changes, which are purely random unpredictable.

example, ~~exa~~earthquake, revolution.

Various Methods of Trend setting

(i) Graphical (or) free hand method
(ii) Semi-avarage method
(iii) Moving avarage method
(iv) Method of least square.

~~Procedure~~

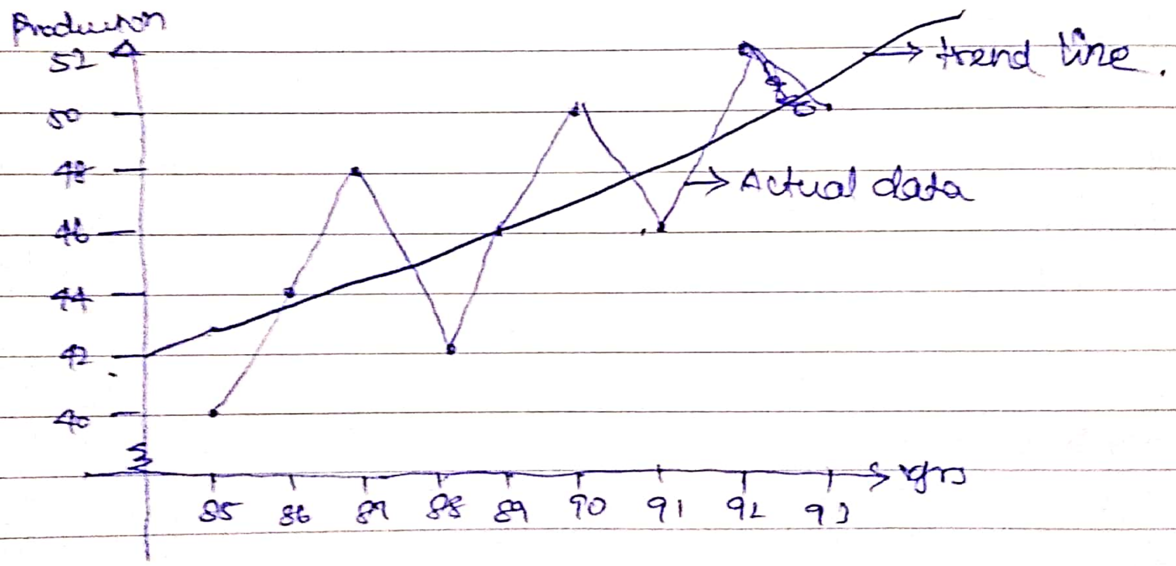
Procedure for obtaining a straight line trend is:,

(i) Plot the time series on a graph
(ii) Examine the direction of the trend based on the plotted information.
(iii) Draw a straight line which will best fit to the data according to personal judgement. It must be smooth straight line or combined of long grade curve.
(iv) The sum of the vertical deviations from the ~~tht~~trend of the annual observation above the trend equal to the sum of ~~observations~~ vertical observation from the trend of the observation below the trend.

STATISTICS

1. Fit a trend line to the following data by the free hand method.

Yrs	85'	86'	87'	88'	89'	90'	91'	92'	93'
Production	40	44	48	42	46	50	46	52	50



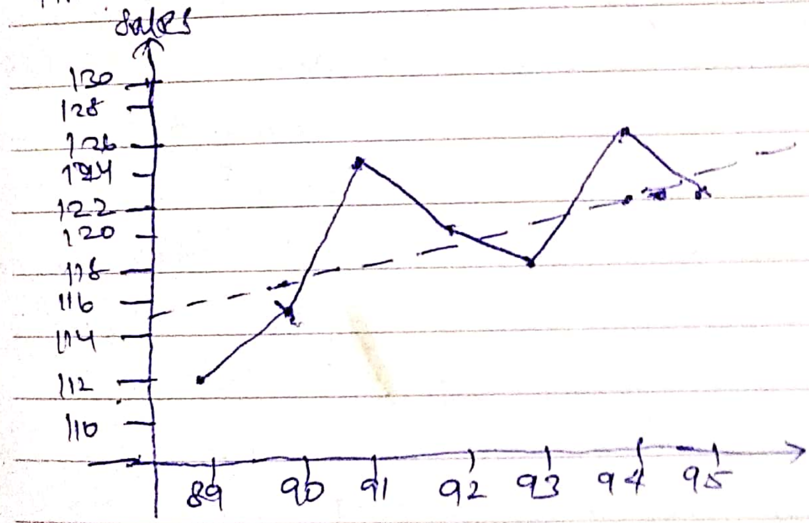
Method of semi averages :

Problem [Illustration]

2. Fit a trend line to the following data by the method of semi averages.

Yr.	89	90	91	92	93	94	95
Sales of firm	112	115	124	120	118	126	122.

$$\begin{array}{r} 227 \\ 124 \\ \hline 351 \end{array}$$



Avg [Yr = {89, 90, 91}]

$$A_1 = \frac{112+115+124}{3} = \frac{351}{3} = 117$$

Avg [Yr = {93, 94, 95}]

$$A_2 = \frac{118+126+122}{3} = \frac{366}{3} = 122$$

We got 2 points, 117 and 122 which shall be plotted corresponding to their ~~fn~~ middle years. Joining these two points, we get the trend line.

Priorities

Edu

Performance

3) Fit a trend line by the method of semi average to the data given below.

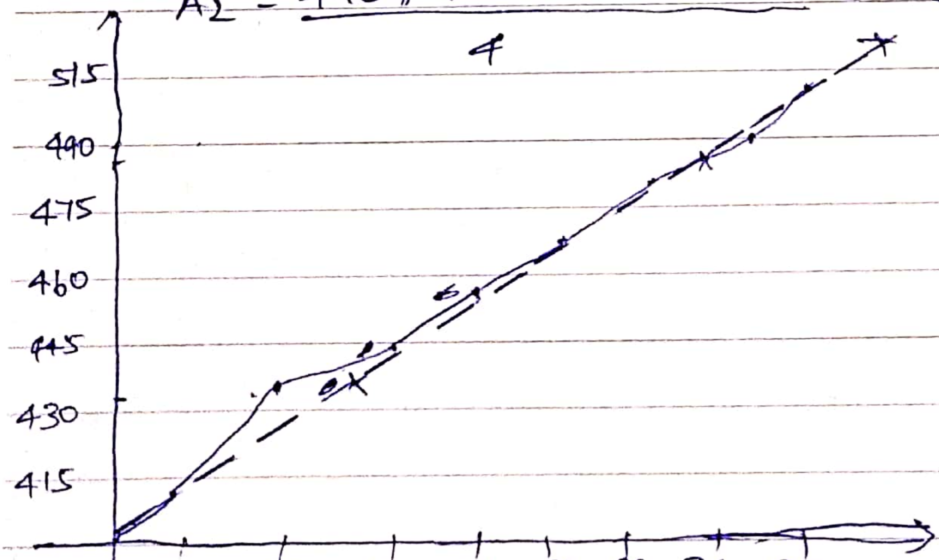
Year	88	89	90	91	92	93	94	95
Sales	412	438	444	454	470	482	490	500

Average 1 {Year = {88, 89, 90, 91}}

$$A_1 = \frac{412 + 438 + 444 + 454}{4} = 437.$$

Average 2 {Year = {92, 93, 94, 95}}

$$A_2 = \frac{470 + 482 + 490 + 500}{4} = 485.5$$



since there are two middle years, 437 correspond ~~to~~ 1st July 1989 and 485.5 corresp to 1st July 1993

From the graph, the ~~trend~~ ~~line~~ path moves ~~at~~ along with the trend line.

MOVING AVARAGE METHOD

using 3 year avarage, determine the trend short term fluctuation.

Year	68	69	70	71	72	73	74	75	76	77
Production	21	22	23	25	24	22	25	26	27	26

66

21+22+23 = 66

X	Y	3 yr moving	(t) Avarage	y-yt short term fluctuation
68	21		—	—
69	22	66	22	0
70	23	70	23.33	-0.33
71	25	72	24	1
72	24	71	23.66	0.67
73	22	71	23.67	-1.67
74	25	73	24.33	-0.67
75	26	78	26	0
76	27	79	26.33	1.33
77	26	—	—	—

Yr 71 72 73 74 75 76 77 78 79 80 81 82

Sugar Production 27.4 31.1 28.7 39.5 47.9 42.6 42.4 44.6 58.4 38.6 51.4 84.4

Year	y Production	4 year total	yt 4 year average	y-yt 4 year moving average centeral
71	27.4			
72	31.1	146.7	36.675	
73	38.7	157.2	39.3	
74	39.5	168.7	42.175	
75	47.9	178.4	44.6	
76	42.6	203.5	50.875	
77	42.6	214		
78	44.6	210		
79	58.4	~~215~~		
80	38.6	232.8		
81	51.4	~~74~~		
82	84.4			
8				

Method of Least Square

Fit a straight line:

General Equation: $Y = a + bX$ — ①

Summation → $\Sigma Y = Na + b\Sigma X$ — ②

① × ΣX → ~~$\Sigma XY = a\Sigma X + b\Sigma X^2$~~

$\Sigma XY = a\Sigma X + b\Sigma X^2$

Fit a straight line to the following series. Estimate the value for 1985. What is the monthly increase in production.

Yr	78'	79'	80'	81'	82'	83'	84'
Production	125	128.7	133	135	140	141	143.

Year	Y	X = Year - A	XY	X^2
78	125	-3	-375	9
79	128.0	-2	-256	4
80	133	-1	-133	1
$\boxed{81}$	135	0	0	0
82	140	1	140	1
83	141	2	282	4
84	143	3	429	9
	~~945.7~~ 945	0	87	28.

$\Sigma X = 0$

$\Sigma Y = 945$

$\Sigma X^2 = 28$

$\Sigma XY = 87$

Scanned by CamScanner

$\Sigma XY = a\Sigma X + b\Sigma X^2$

$87 = a(0) + b(28)$

$87 = b(28)$

$b = \frac{87}{28} = 3.107$ $\boxed{b = 3.107}$

$\Sigma Y = Na + b\Sigma X$

$945 = (7)(a) + (0)$

$\frac{945}{7} = a$ $\boxed{a = 135}$

$\therefore$ The Equation is $Y = 135 + 3.107X$.

In monthly production,

$\boxed{Y = 11.25 + 0.26X}$

Fit a straight line, trend for the following by ~~the~~ method of least square.

Yr	96'	97'	98'	99'	00'	01'
Prod.	7	9	12	15	18	23.

Year	Production	$X = \text{Year} - 98.5'$	x^2	xy
96	7	-2.5	6.25	43.75
97	9	-1.5	2.25	20.25
98	12	-0.5	0.25	3
99	15	0.5	0.25	3.75
00	18	1.5	2.25	40.5
01	23	2.5	6.25	143.75
	84	0	17.5	225

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$225 = a(0) + b(17.5)$

$225 = 17.5b$

$b = \frac{225}{17.5} = 12.86$

$\Sigma Y = Na + b\Sigma X$

$84 = 6a + 0$

$a = 14$

$\therefore$ The trend straight line is $Y = 14 + 12.86X$

$\boxed{Y = 14 + 12.86X}$

Scanned by CamScanner

Fit a straight line for the following data by method of least square.

Yr. 97' 98' 99' 00' 01'

70 ~~80~~ 74 80 ~~80~~ 86 90

Estimate the value for the year 2002.

Year	Y Production	X = Year − 99	X^2	XY
97	70	−2	4	−140
98	74	−1	1	−74
$\boxed{99}$	80	0	0	0
00	86	1	1	86
01	90	2	4	180
	400	0	10	52

14
16
0
0

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

$\cancel{40 = 10b}$ $\quad 52 = 10b$

$\cancel{\boxed{b = 4}}$ $\quad \boxed{b = 5.2}$

$\cancel{\Sigma Y = Na} + b\Sigma X$

40 $\quad \Sigma Y = Na + b\Sigma X$

$$40 = 5a + 0$$

$$\boxed{a = 8}$$

X = Year − 99

Given $\boxed{2002}$

$X = 2002 - 1999$

$\boxed{X = 3}$

∴ The Equation is

$$\boxed{Y = 8 + 5.2X}$$

If $X = 3$, $Y = 8 + 5.2(3)$

$= 8 + 15.6$

$$\boxed{Y = 23.6}$$

2017 JANUARY FEBRUARY MARCH APRIL MAY JUNE

Priorities

1) The production of a cement by a firm in year are given below

year	1	2	3	4	5	6	7	8	9
Production	4	5	5	6	7	8	9	8	10

Calculate the trend values by

(i) 3 year moving average

(ii) Least square method

(i) 3 YEAR MOVING AVARAGE

Year	Y Production	3 year total	Yt average.	Y - Yt Moving average
1	4	—	—	—
2	5	14	4.66	0.34
3	5	16	5.33	0.33
4	6	18	6	0
5	7	21	7	0
6	8	24	8	0
7	9	25	8.33	1.33
8	8	27	9	-1
9	10	—		—

(iii) (ii) LEAST SQUARE METHOD

Year - X Production - Y.

X	Y	XY	X^2
1	4	4	1
2	5	10	4
3	5	15	9
4	6	24	16
5	7	35	25
6	8	48	36
7	9	63	49
8	8	64	64
9	10	90	81
45	62	353	285

X	Y	X-A	XY	X^2
1	4	-4	-16	16
2	5	-3	-15	9
3	5	-2	-10	4
4	6	-1	-6	1
$\boxed{5}$	7	0	0	0
6	8	1	8	1
7	9	2	18	4
8	8	3	24	9
9	10	4	40	16
45	64	0	43	60

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

$$43 = a(0) + b(60)$$

$$b = \frac{43}{60} = 0.7167$$

$$\Sigma Y = Na + \Sigma X\, b$$

$$64 = 9a$$

$$a = \frac{64}{9} \quad 6.8889$$

Equation. $Y = 6.8889 + 0.7167(X-5)$

For 2 marks

Chapter 1:

- Merits & Demerits of Mean, Median, mode.
- Formula

Chapter 2:

- Defining coefficient of variation.
- Standard Deviation
- Combined standard deviation.
- Skewness definition.
- Karl pearson's skewness properties.

Chapter 3:

- Properties of correlation coefficient
- Spearman's Rank correlation.
- Regression Equation.

Chapter 4:

- Multiplication Theorem
- Bayes theorem..

Chapter 5:

- uses of time series
- components of timeseries
- ~~Parts of~~ Methods of trend setting.