Population, Sample and Data Condensation

Mathematics + Sample → Statistics

100% Accuracy — 100% ≥ Accuracy

Statistics is the science which deals with the collection, presentation, analysis and interpretation of numerical data.

Accuracy depends upon :

- Size of Sample
- Sampling method

sample have three constraints :

- Population (if P ↑ then accuracy ↑)
- money (if M↑ then accuracy ↑)
- Time (if T↑ then accuracy ↑)

Intelligent → More thinking power (IQ)

Information
↓
Knowledge
↓ Processing
Artificial Intelligence

Data- Data is defined as "raw facts".

Information- Information is defined as "processed data".

Data $\xrightarrow{\text{processing}}$ Information
(Raw fact)

Types of Data-

- Primary data (also called <u>first hand data</u>)
- Secondary data (also called <u>second hand data</u>)

Primary data -

It is collected by researcher themselves. Since, it is first hand data it is always-
- Upto date
- Accurate
- In desired format (relevant)

The major drawback of this data is that it requires a large amount of resources (money, time, effort)

Unit of effort - man hour If 1 person work 1 hour then it is 1 man hour • No. of person × No. of hour

Secondary data-

It is not collected by researcher directly but collected from some authentic publication agency. This suffers from obvious disadvantage of non-relevance but it has many advantages like easy to attain/easy to gain.

→ Primary data is mainly used in cases where no previous data is available or resources are limited and population size is small.

Ex-• If nature of study is dynamic (changes with time rapidly) then primary data is used.

• If you want high degree of accuracy.

•

→ In other cases secondary data is used

Ex- • If population/sample size is much bigger

• If nature of study is static (not much changes).

Methods of collecting Primary data -

- Direct Personal Interview
- Indirect personal Investigation
- Data Collected through correspondents
- Through Questionnaires
- Data Collected through Enumerators

① Direct Personal Interview:-

In this method the data is collected from the source itself in a one-to-one dialogue manner. The interviewer ask the question and record the answer. This method is usually adopted in cases when the source is aware about the topic.

The major advantage of this ~~data~~ method is that it collect accurate data and it suffers from a disadvantage that it requires a lot of time and patience.

② Indirect Personal Investigation :-

This method is used in case when the source is not available and still you have to collect the data. In this case we collect the data from third person. Hence, it is not always accurate and not in detail.

③ Data collected through Correspondents :-

This method is used in case where population is very large and we don't have enough resources to collect the data from all. In this case we have to be dependent on those persons who can reach to the source on our behalf and ~~collect~~ collect data.

The major advantage is it covers a huge population but the reliability of data is always doubtful.

④ Through Questionaires :-

This is the most widely used method of data collection. In this method a group of questions are formulated and sent to the source through post or ~~email~~ electronic medium. The major advantage is it covers a huge population all together and the major disadvantage is non-response.

⑤ Data collected through Enumerator :-

If the data to be collected is complex in nature and source may not understand the questions directly. we need enumerators. Enumerators are person who takes your question to the source make them understand to the user and record the answers.

The major advantage is very less non-response and major disadvantage is you need a lot of trained enumerator which require huge money.

Sampling Method-

~~Broadly~~ Selecting a few from a large population is called a Sample.

To select a sample which is according to the nature of the study we are having different methods. These methods can be classified into two categories broadly-

(A) Random Sampling (~~Probabilistic~~) (Probabilistic)

(B) Non- Random Sampling (Non-probabilistic)

(A) Random Sampling-

Random sampling is defined as method in which chance of selection of every element is same.

In random sampling we have different methods-

- Simple Random Sampling
- Systematic Sampling
- Stratified Sampling
- Clustered Sampling

• Simple Random Sampling-

In this method a selection is made completely randomly without thinking any of its property. The major advantage of this method is it is the simple most but it may not represent the whole population.

• Systematic Sampling-

In this Sampling samples are selected randomly but according to a rule (system).

This is again a simple method but the selection of rule may invoke controversy.

- **Stratified Sampling-**

In this method the whole population is divided into group known as strata (strata is a greek word). And then in every strata random sampling is done. Advantage is that since we are making strata every Segment of population is selected.

The major disadvange is defining the strata.

- **Clustered Sampling-**

If the groups are made on the basis of physical location it is known as clustered sampling. The major advantage is that it covers all the geographical locations and the major disadvantage is that the groups are dynamic.

(B) Non-Random Sampling-

Not-probabilistic sampling is the sampling in which the chance of selection of items are not always same rather they are dependent on some characteristics. Broadly non-random sampling is divided into 3 categories-

- Convenience
- Quota
- Decision /Judgement

- **Convenience Sampling -**

As the name suggest this type of sampling is always dependent on the convenience of the researcher. This is one of the easiest sampling method is dependent on the availability and willingness of the participants.

The major disadvantage is that it may be biased and does not represent the whole scenario.

• <u>Quota Sampling</u>-

This method of sampling is often used by researcher, in this method there are some rules followed ahead of collecting the data. The whole population is divided and some representative figures are collected from every group which represent a quota.

The major advantage of this method is that it covers every segment of population.

The major drawback is defining and redefining the quota.

• <u>Decision / Judgement sampling</u>-

It is also known as selective sampling. This technique mainly ~~realise~~ relys on the judgement of researcher when selecting who to ask to participate.

The major advantage of this sampling technique is that the sample is selected in a way that most accurate and purposeful data is collected~~ted~~.

The major disadvantage is ~~you~~ it suffers from biasness.

Data Classification and Tabulation-

After collecting the data its time to represent the data.

For the representation purpose the following 3 steps are performed.-

(1) Classification

① Classification-

Classification means dividing the data into different groups on the basis of the value of their property. The classification range is always decided on the basis of study.

Ex- Marks of 10 Students

⇒ 7, 5, 3, 6, 7, 9, 4, 2, 0, 8

Marks	No. of Student
0-2	1
2-4	2
4-6	2
6-8	3
8-10	2

when range is small i.e. No. of classes are more

- Accuracy ↑
- Precision ↑

Marks	No. of st	Mid pt (Avg)
0-5	4	2.5
5-10	6	7.5

when range is Large i.e. less no. of classes.

- Accuracy ↓
- Precision ↓

Number of classes depends on the nature of accuracy required normally more classes means more accuracy.

② Tabulation-

After classification the data is represented in the form of table where every row represent one class and the table normally have two columns- first column represent range and second column represent frequency.

③ Diagramatical Representation-

After tabulation the third step is representing the data in form of diagram or graph.

A diagram is a pictorial representation of data. The data can be quantitative and qualitative.

→ Diagrams are not up to scale

Graphs are specific diagram and it follows two rules -

- It is drawn always upto scale.
- Always involves axis, normally represented by x and y.

Diagram -

① Pie Chart - A piechart is a diagram which represents the data divided into various section. Each section represents a certain percent of the various component.

Steps involved in creating pie chart -

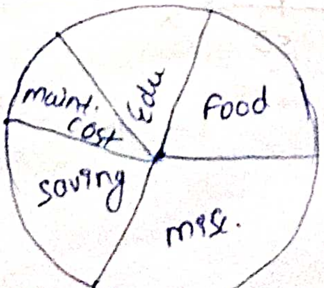
- Express each of the component value as a percentage of the total.
- Since 100% = 360° (in a circle), hence 1% = 3.6° therefore the percentage value will be multiplied with 3.6 of any angles.
- Draw a appropriate circle and then draw a random radius and then divide the circle into parts on the basis of angle calculated.

Ques. A person spends following amount on monthly basis acc^n to the given table. Represent it by pie diagram.

Head	Food	education	Maint. Cost	saving	miscellaneous
Expenses	10,000	10,000	5000	10000	15 000

Sol^n -

Head	Expenses	%	Angle
Food	10000	$\frac{10000}{50000} \times 100 = 20\%$	20×3.6 = 72
Education	10000	20%	20×3.6 = 72
Maint. cost	5000	10%	10×3.6 = 36
Saving	10000	20%	20×3.6 = 72
Misc.	15000	30%	30×3.6 = 108
Total	50000		

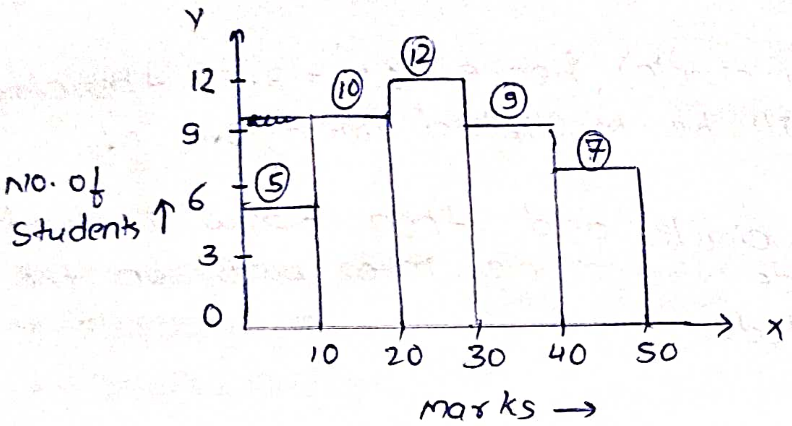


Graphs –

- Histogram
- Frequency Polygan
- Frequency Curve
- Ogive(Cumulative Frequency Curve)

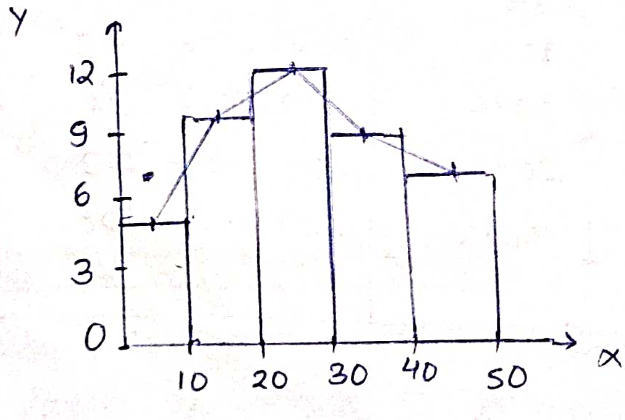
- Histogram–

Marks	0-10	10-20	20-30	30-40	40-50
No. of std.	5	10	12	9	7



- Frequency Polygan–

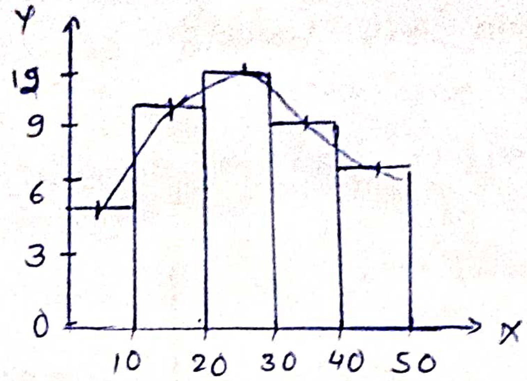
After making the histogram the ceiling of every bar is pointed out at the middle and these points are joined with straight lines. The sequence of all such lines is known as frequency polygon.



• <u>Frequency Curve</u>-

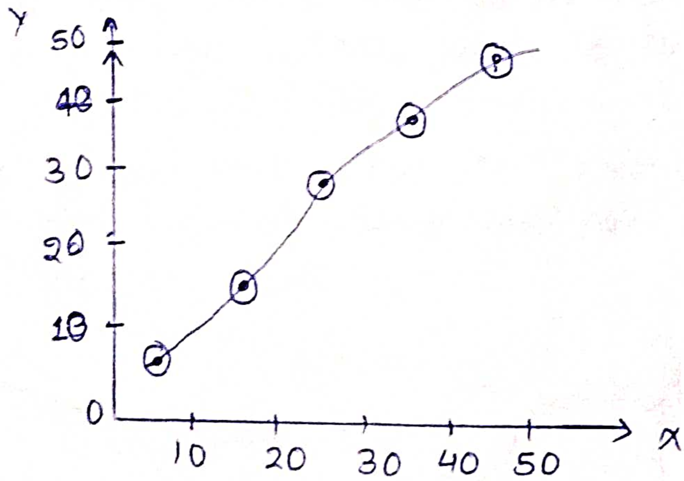
Since frequency polygon is a sequence of multiple straight lines hence at points where two straight lines meet they form an angle.

By smoothening the corners will convert frequency polygon into frequency curve.



• <u>Ogive (Cumulative Frequency Curve)</u>-

Marks	0-10	10-20	20-30	30-40	40-50
No. of Std	5	10	12	9	7
Cumulative Frequency	5	15	27	36	43



• <u>Less than ogive.</u>

Ogive also known as cumulative frequency curve is a unidirectional curve.

It also tells about relative rate of change.

Ogive is of two type-

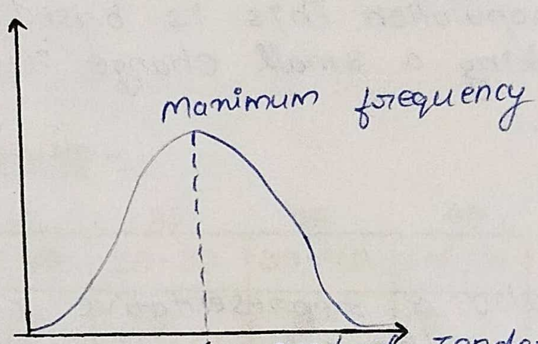
• Less than Ogive • more than ogive

Unit-2

Measures of Central Tendency

Whenever, a study is performed and data is collected it is plotted, the normal behaviour of such data will be like the normal curve. Normal curve is always in bell shape.

Central Tendency → midpoint



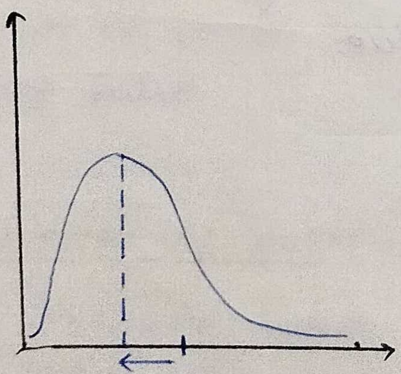
↳ central Tendency / value with maximum frequency

• Normal Curve •

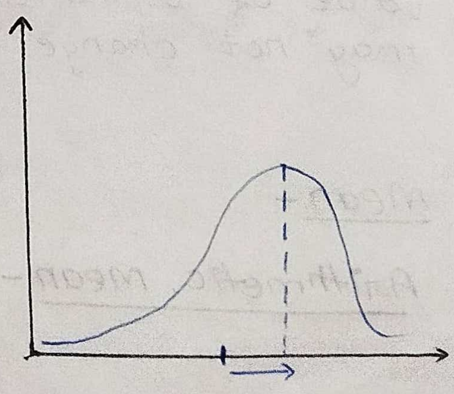
Bell shape -

We can clearly see that the shape of this curve is similar to a bell, where the point of maximum frequency lies almost in the middle and it is called the central tendency.

It is not guaranteed that in every case we are going to get normal curve we can get skewed curve which is of two types-



-vely skewed curve



+vely skewed curve

To measure the central tendency value different methods are proposed with different basis of working These are defined as -

- mean (Arithmetic, Geometric, Harmonic)
- mode
- Median

1) Mean -

mean is a representative value of central tendency for a given population. This is based on every observation and hence making a small change in population changes the value of mean.

2) Mode -

A mode is also a representative central tendency value. It is defined as the value with maximum frequency. This is not dependent on every value of population, hence a slight change in population may not change the mode value.

3) Median -

Another representative value is median. For getting the median we need to first sort the value either in increasing or in decreasing order and the value lying at the central spot is called median. Here also we can say that since median is not dependent on every value of population, making a slight change in population. may not change the median value.

Mean -

(a) Arithmetic Mean -

- For individual observation-

Ex- 5, 3, 9, 8, -7, 3, 0, 2

$\Rightarrow \boxed{\bar{x} = \frac{\Sigma x}{n}}$ where, Σx = add and n = no. of observation

{Πx = multiply}

- For discrete series-

Value(x)	3	7	9	11	13	15
Frequency(f)	5	10	12	13	10	7

shortcut Method

$\boxed{\bar{x} = A + \frac{\Sigma fd}{n}}$ where, A = Assume mean
d = deviation from value ($d = x - A$)

$\boxed{\bar{X} = \frac{\Sigma fx}{\Sigma f}}$ where, x = value and f = frequency corresponding the value

or $\boxed{\bar{x} = \frac{\Sigma fx}{n}}$

Direct Method

- For continuous series-

	5	15	25	35	45
Range	0-10	10-20	20-30	30-40	40-50
Frequency	5	10	15	9	6

mid value of range
$\Rightarrow \frac{\text{upperlimit} + \text{lowerlimit}}{2}$

$\boxed{\bar{X} = \frac{\Sigma fm}{\Sigma f}}$ where, m = mid value of range/class
f = frequency corresponding the class

Benefits of arithmetic mean-

- It can be easily calculated.
- Its calculation is based on all the observation.
- It is easy to understand.
- It is rigidly defined by mathematical formula.
- When it comes to comparing the two series it is the best value.

Demerits of arithmetic mean-

- The extreme value have a greater effect on mean.
- It can't be calculated if all the value are not known.
- It can't be defined on qualitative data.
- Mean may lead to wrong Conclusion if data is not accurate.

Ques- The mean of 68 numbers is 18, If each number is divided by 6 find the new mean.

Sol^n- Given,

~~$\Sigma x = 18 \times 6$~~

$\bar{A} = 18$, $n = 68$, $\Sigma x = ?$

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

$18 = \frac{\Sigma x}{68}$

$\Sigma x = 18 \times 68 = 1224$

New $\Sigma x = \frac{1}{6}(1224)$ or $\frac{1}{6}(18 \times 68)$

Now, new mean $= \frac{\Sigma x}{n} = \frac{18 \times \cancel{68}}{6 \times \cancel{68}} = 3$ Ans

Q- while calculating the average of 15 times it come out to be to be 20. Later it was realise that by mistake we took 18 in place of 28 and 25 in place of 20 while calculating. What is the correct average?

Sol^n- Given, $n = 15$, $\bar{x} = 20$, $\Sigma x = ?$

$20 = \frac{\Sigma x}{15}$

$\therefore \Sigma x = 20 \times 15 = 300$

wrong value	Correct Value
18	28
25	20

New Σx = old Σx - (wrong item) + correct item

$= 300 - 18 - 25 + 28 + 20$

$= 300 - 43 + 48$

$= 305$

New $\bar{x}$ i.e. correct average $= \frac{\text{new } \Sigma x}{n} = \frac{305}{15} = 20.3$ Ans

Q- The average of 10 number was found to be 13 which number should be added to this group so that the average become 15.

Sol^n- Given,

$n = 10$

$\bar{x} = 13$

$\Sigma x = ?$

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

$\Sigma x = 10 \times 13 = 130$

New $n = 11$

New, $\bar{x} = 15$

New $\Sigma x = ?$

new $\Sigma x = n \times$ new $\bar{x}$

$= 11 \times 15$

$= \cancel{150}\ 165$

$\therefore$ New number added = New Σx - Old Σx

$= 165 - 130$

$= 35$

we have to add 35 so that the average become 15.

Ans

Combined mean of two different series-

Series 1:-

n_1 observation

$\bar{x}_1$ is the mean

Series 2:-

n_2 observation

$\bar{x}_2$ is the mean

$$\boxed{\bar{x}_{12} = \frac{n_1 \times \bar{x}_1 + n_2 \times \bar{x}_2}{n_1 + n_2}}$$

Q- A class runs in two sections, sec A and sec B. In sec A 70 students have passed and their mean pass per% is 65% where as from sec B 60 students have passed with their mean pass per% as 60%. Find out the mean pass % of whole class.

Sol^n Given,

$n_1 = 70$ $\quad n_2 = 60$

$\bar{x}_1 = 65$ $\quad \bar{x}_2 = 60$

$$\bar{x}_{12} = \frac{n_1 \times \bar{x}_1 + n_2 \times \bar{x}_2}{n_1 + n_2} = \frac{70 \times 65 + 60 \times 60}{70 + 60} = \frac{4550 + 3600}{130} = \frac{8150}{130}$$

$= 62.69\%$

Ans

Q- In a factory there are some workers which can be divided into two categories skilled and unskilled. The average daily wages of skilled employees is Rs 1000 and the average daily wages of unskilled employees is Rs 700. If the average daily wages of all the employees is Rs 820. Find the per % of skille labour.

Soln Given,

$n_1 = ?$ $\quad n_2 = ?$ $\quad \bar{x}_{12} = 820$

$x_1 = 1000$ $\quad \bar{x}_2 = 700$

Let the per % of skilled labours be x and per % of unskilled labours be $100-x$.

i.e. $n_1 = x$, $n_2 = 100-x$

Now,

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

$$820 = \frac{1000x + 700(100-x)}{\cancel{x}+100-\cancel{x}}$$

$$82000 = 1000x + 70000 - 700x$$

$$12000 = 300x$$

$$x = \frac{12000}{300}$$

$$\boxed{x = 40\%}$$

Percentage of skilled labours = 40% Ans

Q- Find out the mean of the following data which represents the income of following data.

14780
15760
26690
27750
24840
24920
16100
17810
27050
26950

Soln $\text{Mean} = \frac{222650}{10}$

$= 22265$ Rs

Q- Find the mean of the marks obtained by the students given in the following table.

Marks	0	1	2	3	4	5
No. of std.	5	10	20	25	25	15

Solⁿ.

Marks (x)	No. of std. (f)	fx
0	5	0
1	10	10
2	20	40
3	25	75
4	25	100
5	15	75
	100	300

$\bar{x} = \frac{\Sigma fx}{n} = \frac{300}{100} = 3$

↓

Direct method

Short-cut method - $\boxed{\bar{x} = A + \frac{\Sigma fd}{n}}$

where, A ⇒ Assume mean

d → deviation from the value

Marks (x)	No. of std. (f)	$d = x - A$	$f \cdot d$
0	5	-2	-10
1	10	-1	-10
2	20	0	0
3	25	1	25
4	25	2	50
5	15	3	45
	100		100

Let assume mean (A) = 2

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

$= 2 + \frac{100}{100}$

$= 2 + 1$

$= 3$

Q- Find out the mean of the following frequency distribution

Class	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	5	10	15	25	25	20

Soln

Class	Frequency	mid point(m)	fm	$d = x - A$	fd
20-30	5	25	125	-30	-150
30-40	10	35	350	-20	-200
40-50	15	45	675	-10	-150
50-60	25	55	1375	0	0
60-70	25	65	1625	10	250
70-80	20	75	1500	20	400
	100		5650		150

Direct method -

$$\bar{x} = \frac{\Sigma fm}{n} = \frac{5650}{100} = 56.5$$

Shortcut method -

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

Let assume mean (A) be 55,

$$\bar{x} = A + \frac{\Sigma fd}{n} = 55 + \frac{150}{100} = 55 + 1.5 = 56.5$$

Q- Find the mean of the following frequency distribution by ~~both direct and~~ shortcut method.

Class	40-50	30-40	20-30	10-20	0-10
Frequency	10	11	13	12	4

→ Exclusive frequency distribution

Class	Frequency	mid point(m)	$d = x - A$	fd
0-10	4	5	-30	-120
10-20	12	15	-20	-240
20-30	13	25	-10	-130
30-40	11	35	0	0
40-50	10	45	10	100
	50			-390

Let A be 35,

$$\bar{x} = 35 + \frac{(-390)}{50} = 35 - 7.8 = 27.2$$

Q. From the following data of income distribution calculate the arithmetic mean. It is given that the ~~total~~ income of person is not less than ₹20.

Hourly Income	No. of person
Below 30	15
Below 40	36
Below 50	60
Below 60	77
Below 70	88
Below 80	100

Soln

Class	Frequency	mid point (m)	d=m-A	fd
20-30	15	25	-30	-450
30-40	21	35	-20	-420
40-50	24	45	-10	-240
50-60	17	55	0	0
60-70	11	65	10	110
70-80	12	75	20	240
	100			-760

Let A = 55,

$$\bar{x} = A + \frac{\Sigma fd}{n} = 55 + \left(\frac{-760}{100}\right) = 55 - 7.6 = 47.4$$

Q- Find out the missing frequency ~~with~~ if the mean of the given distribution is 19.

Class	5-10	10-15	15-20	20-25	25-30
Frequency	2	2	?	4	4

Solⁿ Given, $\bar{x} = 19$, Let missing frequency be f, Let $A = 12.5$

Class	Frequency	mid point	$d = x - A$	fd
5-10	2	7.5	-5	-10
10-15	2	12.5	0	0
15-20	f	17.5	5	$5f$
20-25	4	22.5	10	40
25-30	4	27.5	15	60
	$12+f$			$90+5f$

Now,

$$\bar{x} = A + \frac{\sum fd}{n}$$

$$19 = 12.5 + \frac{90+5f}{12+f}$$

$$19 - 12.5 = \frac{90+5f}{12+f}$$

$$6.5 = \frac{90+5f}{12+f}$$

$$6.5f + 78 = 90 + 5f$$

$$1.5f = 12$$

$$f = \frac{12}{1.5}$$

$$f = \frac{12}{15} \times 10$$

$$\boxed{f = 8}$$

Q. mean of a following frequency distribution is 50 but two frequencies are missing. And the missing frequency.

Class	0-20	20-40	40-60	60-80	80-100	
Frequency	17	f_1	32	f_2	19	Total=120

Solⁿ- Given,

$\bar{x} = 50$, Let missing frequency be f_1 and f_2.

Class	Frequency	mid point	fm
0-20	17	10	170
20-40	f_1	30	$30f_1$
40-60	32	50	1600
60-80	f_2	70	$70f_2$
80-100	19	90	1710
	$68+f_1+f_2$		$30f_1+70f_2+3480$

Now, Total frequency = 120

$68 + f_1 + f_2 = 120$

$f_1 + f_2 = 52$ — (i) or $f_1 = 52 - f_2$ — (i)

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

$50 = \frac{30f_1 + 70f_2 + 3480}{120}$

$6000 = 30f_1 + 70f_2 + 3480$

$600 = 3f_1 + 7f_2 + 348$

$3f_1 + 7f_2 = 600 - 348$

$3f_1 + 7f_2 = 252$ — (ii)

Putting value of f_1 in equⁿ (ii)

$3(52 - f_2) + 7f_2 = 252$

$156 - 3f_2 + 7f_2 = 252$

$4f_2 = 96$

$f_2 = 96/4$

$\boxed{f_2 = 24}$

And $f_1 = 52 - 24$

$\boxed{f_1 = 28}$

Q- Find the mean of the following data

Class	50-59	40-49	30-39	20-29	10-19	0-9
Frequency	2	5	8	10	12	4

→ Inclusive frequency distribution

Let $A = 34.5$

Sol^n

Class	Frequency	mid point	$d = x - A$	fd
50-59	2	54.5	20	40
40-49	5	44.5	10	50
30-39	8	34.5	0	0
20-29	10	24.5	-10	-100
10-19	12	14.5	-20	-240
0-9	4	4.5	-30	-120
	41			-370

$$x = A + \frac{\Sigma fd}{n}$$

$$= 34.5 + \frac{(-370)}{41}$$

$$= 34.5 - 9.02$$

$$= 25.48$$

Q- The given numbers 2, 4, 6, 8, 10, have frequency $x+4$, $x+3$, $x+2$, $x+1$, x respectively. If the mean of this distribution is 5, find the value of x.

Sol^n

Number (x)	frequency (f)	fx
2	$x+4$	$2x+8$
4	$x+3$	$4x+12$
6	$x+2$	$6x+12$
8	$x+1$	$8x+8$
10	x	$10x$
	$5x+10$	$30x+40$

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

$5 = \frac{30x+40}{5x+10}$

$25x + 50 = 30x + 40$

$10 = 5x$

$[x = 2]$

Weighted Average-

When we calculate average for any frequency distribution, under normal circumstances all the values are given equal importance (weight), but in some cases the importance of values are different and in such cases the formula for calculating the average is also different and is written as

$$\boxed{W.A = \frac{x_1 * w_1 + x_2 * w_2 + \dots + x_n w_n}{w_1 + w_2 + \dots + w_n}}$$

where, $x_1, x_2 \dots x_n \rightarrow$ values

$w_1, w_2 \dots w_n \rightarrow$ Their weight

Median-

Unlike the mean where every value was considered, median is the positional value. Here instead of the values their position is important.

→ Individual values

→ Discrete Series

→ Continuous Series

- **Individual values-**

(a) When the no. of observations is odd

$$\boxed{\text{median} = \frac{n+1}{2} \text{ th value}}$$

Q- Find the median of 7, 11, 5, 3, 12, 9, 10

sorted → 3, 5, 7, 9, 10, 11, 12

median $= \frac{n+1}{2}$ th value $= \frac{7+1}{2} = \frac{8}{2} = 4$th value i.e. 9.

(b) when the no. of observations is even

$$\boxed{median = \frac{\frac{n}{2}th\ value + (\frac{n}{2}+1)th\ value}{2}}$$

Find the median of 7, 11, 5, 3, 12, 9, 10, 2.

Sorted → 2, 3, 5, 7, 9, 10, 11, 12

$$median = \frac{\frac{8}{2} + \frac{8}{2}+1}{2} = \frac{4th\ value + 5th\ value}{2} = \frac{7+9}{2} = \frac{16}{2} = 8$$

• Discrete Series -

For finding the ~~dis~~ median of discrete series value, we have to follow the following step-

1) If the given data is not in a sorted order, sort it first.
2) Find out cumulative frequency.
3) Find the middle item using $\left(\frac{n+1}{2}\right)$.
4) See the cumulative frequency which is equal or greater than $\left(\frac{n+1}{2}\right)$
5) The corresponding value of the variable gives median.

Q- Find the median of following

Wages	50	60	70	55	68	65
No. of workers	5	8	3	4	2	1

soln

wages	No. of workers	C.f
50	5	5
55	4	9
60	8	17 ←
65	1	18
68	2	20
70	3	23

Step-3 ⇒ $\frac{n+1}{2} = \frac{23+1}{2} = 12$ ∴ median wage is 60.

- Continuous Series -

The following steps are to be followed while getting the median of the data given as continuous series -

1) If the given data is not in a sorted order, sort it first.
2) Find cumulative frequency (c.f.).
3) Find $\left(\frac{n}{2}\right)$ and first c.f value which is just greater or equal to $(n/2)$.
4) The corresponding class is called median class which contains the median value.
5) Apply the formula for median

$$\boxed{\text{median} = L + \frac{i}{f}\left(\frac{N}{2} - c\right)}$$

where, $L \rightarrow$ the lower limit of median class
$i \rightarrow$ Class interval of median class
$f \rightarrow$ Frequency of median class
$N \rightarrow$ Total no. of observation
$C \rightarrow$ Cumulative frequency of the class preceding the median class.

Q- Find the median

Class Interval	10-20	20-30	30-40	40-50	50-60
Frequency	4	5	9	7	6

Soln

Class Interval	Frequency	Cf
10-20	4	4
20-30	5	9
30-40	9	18 ←
40-50	7	25
50-60	6	31

Step-3 ⇒ $n/2 = 31/2 = 15.5$

median class → 30-40

$L = 30,\ i = 10,\ f = 9,\ N = 31,\ C = 9$

$$\text{Median} = 30 + \frac{10}{9}\left(\frac{31}{2} - 9\right)$$
$$= 30 + \frac{\overset{5}{\cancel{10}}}{9} \times \frac{13}{\cancel{2}}$$
$$= 30 + \frac{65}{9} = 37.22$$

Q- Calculate the missing frequency from the following frequency distribution. The median is given as 25.625.

class	0-10	10-20	20-30	30-40	40-50
Frequency	4	7	16	10	?

Sol^n Let the missing frequency be x.

median = 25.625

median class = 20 - 30

$f = 16$

Class	Frequency	cf
0-10	4	4
10-20	7	11
20-30	16	27
30-40	10	37
40-50	x	$37+x$

& $L = 20$, $i = 10$, $N = 37+x$, $c = 11$

$$\text{median} = L + \frac{i}{f}\left(\frac{N}{2} - c\right)$$

$$25.625 = 20 + \frac{10}{16}\left(\frac{37+x}{2} - 11\right)$$

$$25.625 = 20 + \frac{\cancel{10}^{5}}{16}\left(\frac{37+x-22}{\cancel{2}}\right)$$

$$25.625 = 20 + \frac{5}{16}(x+15)$$

$$5.625 = \frac{5(x+15)}{16}$$

$$90 = 5(x+15)$$

$$18 = x+15$$

$$\boxed{x = 3}$$

Q- Find the median of following frequency distribution

Class	3-7	7-15	15-20	20-30	30-40	45-50
Frequency	8	10	12	9	6	5

Sol

Class	Frequency	cf
3-7	8	8
7-15	10	18
15-20	12	30 ←
20-30	9	39
30-40	6	45
45-50	5	50

Step 3 → $\frac{n}{2} = \frac{50}{2} = 25$

median class = 15-20

$L = 15, \quad i = 5, \quad f = 12, \quad C = 18$

$$\text{median} = L + \frac{i}{f}\left(\frac{N}{2} - C\right)$$

$$= 15 + \frac{5}{12}\left(\frac{50}{2} - 18\right)$$

$$= 15 + \frac{5}{12}\left(25 - 18\right)$$

$$= 15 + \frac{5}{12} \times 7$$

$$= 15 + \frac{35}{12} \approx 15 + 2.91$$

$$= 17.91$$

Q- Median value of the following distribution is 100.48 but two frequencies are missing. Determine the unknown frequencies.

Marks	55-64	65-74	75-84	85-94	95-104	105-114	115-124	125-134	135-144	Total
Frequency	2	19	78	?	301	?	92	15	4	900

Sol-

Marks	Frequency	CF
54.5-64.5	2	2
64.5-74.5	19	21
74.5-84.5	78	99
84.5-94.5	x	$99+x$
94.5-104.5	301	$400+x$
104.5-114.5	y	$400+x+y$
114.5-124.5	92	$492+x+y$
124.5-134.5	15	$507+x+y$
134.5-144.5	4	$511+x+y$
	900	

$$511+x+y=900$$
$$x+y=900-511$$
$$x+y=389 \quad -(i)$$

Now, $\text{median} = L + \frac{i}{f}\left(\frac{N}{2}-C\right)$

$$100.48 = 94.5 + \frac{10}{301}\left(\frac{900}{2}-(99+x)\right)$$

$$100.48-94.5 = \frac{10}{301}(450-99-x)$$

$$5.98 \times 301 = 10(351-x)$$

$$1799.98 = 3510 - 10x$$

$$10x = 3510 - 1799.98$$

$$10x = 1710.02$$

$$\boxed{x = 171.002}$$

Since, x is frequency, hence it has to be in whole no. so we will take $x=171$.

by eqn $\quad x+y=389$

$$171+y=389 \Rightarrow \boxed{y=218}$$

Mode-

As we are aware that mode is the value which is having the highest frequency. The class with the maximum frequency is called modal class.

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

where, L → Lower limit of modal class
f_1 → Frequency of modal class
f_0 → Frequency of class prior to modal class
f_2 → Frequency of class next to modal class
i → modal class interval

Q- Calculate mode from the following.

Central size	10	20	30	40	50	60	70
Frequency	5	11	16	28	20	7	4

Soln

Range	Frequency
5-15	5
15-25	11
25-35	16
35-45	28
45-55	20
55-65	7
65-75	4

~~Mode=~~ modal class = 35-45

$$\text{mode} = 35 + \frac{28-16}{2 \times 28 - 16 - 20} \times 10$$

$$= 35 + \frac{12}{56-36} \times 10$$

$$= 35 + \frac{12}{20} \times 10 \Rightarrow 41 \quad \text{Ans}$$

Bi-modal distribution-

If the maximum frequency value and the second ~~maximum~~ maximum frequency value are significantly different then it is uni-modal distribution and it such case direct formula can be applied but think about the given distribution.

Central Size	10	20	30	40	50	60	70
Frequency	5	11	16	28	28	7	4

It is very clear that in this distribution, the highest and second highest value of frequency are same (or almost same) such distribution are known as bi-modal distribution.

> In such cases we use a tool known as analysis table to determine which is the modal class.

Q- Find out the mode of a following series:

Class	0-7	8-15	16-23	24-31	32-39	40-47	48-55
Frequency	1	2	10	4	10	9	2

Soln-

Class	Frequency	Grouping				
0 - 7.5	1	3				
7.5 - 15.5	2		12	13		
15.5 - 23.5	10	14			16	
23.5 - 31.5	4		14			24
31.5 - 39.5	10	19		23		
39.5 - 47.5	9				21	
47.5 - 55.5	2		11			

Analysis Table-

Class						
0-7.5						0
7.5-15.5						0
15.5-23.5					1	1
23.5-31.5		1	1		1	3
31.5-39.5	1	1	1	1	1	5
39.5-47.5	1		1	1		3
47.5-55.5				1		1

$\rightarrow$ modal class {31.5-39.5}

$$\text{Mode} = 31.5 + \frac{10-4}{2\times10-4-9} \times 8$$

$$= 31.5 + \frac{6}{20-13} \times 8$$

$$= 31.5 + \frac{6}{7} \times 8$$

$$= 31.5 + 6.8$$

$$= 38.2 \text{ Ans}$$

Q- Find the mode of the following distribution.

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	5	8	7	12	28	20	10	10

Sol^n.

Class	Fre.	Grouping				
0-10	5	13		20		
10-20	8		15		27	
20-30	7	19				47
30-40	12		40	60		
40-50	28	48			58	
50-60	20		30			40
60-70	10	20				
70-80	10					

Class						
0-10						0
10-20						0
20-30					1	1
30-40		1	1		1	3
40-50	1	1	1	1	1	5
50-60	1		1	1		3
60-70				1		1
70-80						0

∴ Modal class → 40-50

$$\text{Mode} = 40 + \frac{28-12}{2\times28-12-20} \times 10$$

$$= 40 + \frac{16}{56-32} \times 10$$

$$= 40 + \frac{\cancel{16}\ 4\ 2}{\cancel{24}} \times 10$$

$$\cancel{10}\ 20\ /3$$

$$= 40 + 20/3$$

$$= 40 + 6.67$$

$$= 46.67$$

Harmonic Mean -

• Individual ~~mean~~ values -

$$H.M = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}} = \frac{n}{\Sigma 1/x}$$

Q - Find harmonic mean of 5, 10, 15, 20, 25, 30

Soln -

$$H.M = \frac{6}{1/5 + 1/10 + 1/15 + 1/20 + 1/25 + 1/30}$$

$$= \frac{6}{\frac{60+30+20+15+12+10}{300}}$$

$$= \frac{6}{\frac{147}{300}}$$

$$= 12.24 \quad \text{Ans}$$

- For discrete & continuous value -

$$H.M = \frac{n}{f_1 \times \frac{1}{x_1} + f_2 \times \frac{1}{x_2} + f_3 \times \frac{1}{x_3} \cdots + f_n \times \frac{1}{x_n}} = \frac{n}{\Sigma f \, 1/x}$$

Q- Calculate the harmonic mean of the following data.

Values	3	6	8	12	15	20	24
Frequency	2	3	4	1	2	3	2

Soln

$$H.M = \frac{17}{2 \times \frac{1}{3} + 3 \times \frac{1}{6} + 4 \times \frac{1}{8} + 1 \times \frac{1}{12} + 2 \times \frac{1}{15} + 3 \times \frac{1}{20} + 2 \times \frac{1}{24}}$$

$$= \frac{17}{2/3 + 1/2 + 1/2 + \frac{1}{12} + 2/15 + \frac{3}{20} + \frac{1}{12}}$$

$$= \frac{17}{2/3 + 1/2 + 1/2 + 1/12 + 2/15 + 3/20 + 1/12}$$

$$= \frac{17}{0.6 + 1 + 0.08 + 0.13 + 0.15 + 0.08}$$

$$= \frac{17}{2.04}$$

$$= 8.33 \quad \text{Ans}$$

Q. From the following data calculate the harmonic mean

C.I.	10-20	20-30	30-40	40-50	50-60
Fre.	5	8	7	6	3

Soln

C.I	Frequency	mid point (m)
10-20	5	15
20-30	8	25
30-40	7	35
40-50	6	45
50-60	3	55

Now, mode $= \dfrac{n}{\Sigma \frac{f}{m}}$

$$= \frac{29}{5\times\frac{1}{15} + 8\times\frac{1}{25} + 7\times\frac{1}{35} + 6\times\frac{1}{45} + 3\times\frac{1}{55}}$$

$$= \frac{29}{\frac{5\times1155 + 8\times693 + 7\times495 + 6\times385 + 3\times315}{17325}}$$

$$= \frac{29\times17325}{18039}$$

$$= \frac{502425}{18039}$$

$= 27.85$ Ans

$$\boxed{\text{Mode} = 3\,\text{median} - 2\,\text{Mean}}$$

Unit - 3

Measure of Dispersion

Measure of Dispersion - Whenever we talk about a frequency distribution, the central tendency value (mean, mode, median) can't represent the whole distribution alone.

Distribution 1 → 50, 50, 50

Dis. 2 → 30, 50, 70

Dis. 3 → 10, 50, 90

Here, we can clearly see all the 3 distribution have same mean i.e. 50 but by observation we can clearly see that all three distributions are different. It means we need something else along with central tendency value. to represent the distribution.

Therefore, we need another measure and it is called measure of dispersion.

→ measure of dispersion represents the distance of values of dispersion from the central tendency value.

There are two types of measures for dispersion -

- Absolute measure
- Relative Measure

Absolute measures -

Absolute measures are calculated in terms of given frequency distribution values. ~~where as~~

Relative measures-

Relative measures are calculated in terms of percentage or ratio.

{Ratio is always used for comparing two different values and when we have to talk about a single distribution its better to find out absolute measure.}

The first measure of dispersion is called as range and it is defined as the difference between the largest value and the smallest value.

$\boxed{Range = L - S}$ where, L → Largest and S → Smallest

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

& ratio (coefficient of range)

Q- Find Range, of 4, 14, 19, 25, 3, 9, 6, 4, 10

Sol^n^- Range = L − S = 25 − 3 = 22

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{25-3}{25+3} = \frac{22}{28} = \frac{11}{14}$$

Other measures of dispersion-

The next measure of dispersion is known as mean deviation.

To find out mean deviation of the distribution we follow the following step -

1) We find the mean of the given distribution.

2) We calculate the deviation of every item from the mean.

3) We find the mean of all the deviation. which is called mean deviation.

Ex- 30, 50, 70

① → mean 50

② → −20, 0, 20

③ → 0 ⇒ mean deviation of above series

while calculating the mean deviation whenever we calculate deviation its either positive or negative and hence have cancelling effect, to avoid this a new measure known as standard deviation was introduced. For calculating the standard deviation we follow the following steps–

1) We find the mean of the given distribution.
2) We calculate the deviation of every item from the mean
3) Square all the deviations
4) Find the mean of the squares of deviation.
5) Find the squareroot of the value obtained in step 4.

Formula for ~~step de~~ standard deviation –

$$\boxed{\sigma = \sqrt{\frac{1}{N}\cdot\Sigma(x-\bar{x})^2} = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}}$$

shortcut method

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

Q- Find the standard deviation of the following data
50, 55, 56, 32, 48, 44, 36, 46

Soln –

x	x^2	$d = x-A$	d^2
50	2500	2	4
55	3025	7	49
56	3136	8	64
32	1024	−16	256
48	2304	0	0
44	1936	−4	16
36	1296	−12	144
46	2116	−2	4

$$\sigma = \sqrt{\frac{17337}{8} - \frac{134689}{64}}$$

$$\sigma = \sqrt{\frac{138696 - 134689}{64}}$$

$$\sigma = \sqrt{\frac{4007}{64}} = \sqrt{62.6093}$$

$$\boxed{\sigma = 7.9126}$$

$\Sigma x = 367$, $N = 8$

$(\Sigma x)^2 = 134689$, $\Sigma x^2 = 17337$

$N^2 = 64$

$\Sigma d = -17$, $\Sigma d^2 = 537$

~~$(\Sigma d)^2 =$~~

$$\sigma = \sqrt{\frac{537}{8} - \left(-\frac{17}{8}\right)^2}$$

$$\sigma = \sqrt{67.13 - (-2.13)^2}$$

$$= \sqrt{67.13 - 4.53}$$

$$= \sqrt{62.60}$$

$$\boxed{\sigma = 7.91}$$

Standard deviation of discrete and continuous series–

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

Q– Calculate the standard deviation of the following

Marks	1	2	3	4	5
No. of std.	5	10	15	12	10

Soln

x	f	x^2	fx^2	fx	$(fx)^2$
1	5	1	5	5	25
2	10	4	40	20	400
3	15	9	135	45	2024
4	12	16	192	48	2304
5	10	25	250	50	2500
	52		622	168	7253

$$\cancel{\sigma = \sqrt{\frac{622}{52} - \frac{7253}{2704}} = \sqrt{11.96 - 2.68} = \sqrt{9.28} = 3.04}$$

$$\sigma = \sqrt{\frac{622}{52} - \left(\frac{168}{52}\right)^2} = \sqrt{11.96 - (3.23)^2} = \sqrt{11.96 - 10.43} = \sqrt{1.53}$$

$$= 1.23$$

Q- Find out the standard deviation of the following data which was oven study of 50 people.

Age under (in years)	10	20	30	40	50
No. of person dying	10	23	37	45	50

Sol^n ~~Age nu~~

Class	f	m	m^2	fm^2	fm
0-10	10	5	25	250	50
10-20	13	15	225	2925	195
20-30	14	25	625	8750	350
30-40	8	35	1225	9800	280
40-50	5	45	2025	10125	225
				31850	1100

$$\sigma = \sqrt{\frac{31850}{50} - \left(\frac{1100}{50}\right)^2}$$

$$= \sqrt{637 - 484} = \sqrt{153} = 12.369$$

Q- The standard deviation calculated from the set of 32 observations is 5. If the sum of the observation is 80. Find the sum of squares of these observation.

Sol^n Given,

$N = 32, \quad \sigma = 5, \quad \Sigma x = 80, \quad \Sigma x^2 = ?$

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

$$5 = \sqrt{\frac{\Sigma x^2}{32} - \left(\frac{80}{32}\right)^2}$$

$$5^2 = \frac{\Sigma x^2}{32} - 6.25$$

$$25 = \frac{\Sigma x^2}{32} - 6.25$$

$$\therefore 31.25 = \frac{\Sigma x^2}{32}$$

$$[\Sigma x^2 = 1000]$$

Q- Given that, $N=100$, $\bar{x}=2$, $\sigma=4$, find Σx and Σx^2

Sol^n.
$$\sigma = \sqrt{\frac{1}{N}\cdot\Sigma(x-\bar{x})^2} = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}$$

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

$$2 = \frac{\Sigma x}{100}$$

$$[\Sigma x = 200]$$

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

$$4 = \sqrt{\frac{\Sigma x^2}{100} - \left(\frac{200}{100}\right)^2}$$

$$4 = \sqrt{\frac{\Sigma x^2}{100} - 2^2}$$

$$16 = \frac{\Sigma x^2}{100} - 4$$

$$20 = \frac{\Sigma x^2}{100}$$

$$[\Sigma x^2 = 2000]$$

Q- The mean and standard deviation of a sample of 100 observations were calculated as 40 and 5.1 respectively. Later it was found that one value originaly 40 was taken 50 by mistake find the correct mean and standard deviation.

Sol^n- Given, old $\bar{x} = 40$, old $\sigma = 5.1$

$N = 100$

$$\therefore \text{old } \bar{x} = \frac{\Sigma x}{N}$$

$$40 = \frac{\Sigma x}{100} \quad \therefore \quad \text{old } \Sigma x = 4000$$

New, $\Sigma x = 4000 + 40 - 50$

$= 4000 - 10$

$= 3990$

New mean $= \frac{\Sigma x}{N} = \frac{3990}{100} = 39.9$

Now,

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

$5.1 = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{4000}{100}\right)^2}$

$5.1 = \sqrt{\frac{\Sigma x^2}{100} - \cancel{0.16}(40)^2}$

$26.01 = \frac{\Sigma x^2}{100} - \cancel{0.16}\,1600$

$26.17 = \frac{\Sigma x^2}{100}$

$2617 = \Sigma x^2$

$\therefore$ old $\Sigma x^2 = 2617$

New $\Sigma x^2 = 2617 + (40)^2 - (50)^2$

$= 2617 + 1600 - 2500$

$= 1717$

Now,

New $\sigma = \sqrt{\frac{3990}{100} - \left(\frac{1717}{100}\right)^2}$

$\sigma = \sqrt{39.9 - (17.17)^2}$

New $\sigma = \sqrt{\frac{1717}{100} -}$

$26.01 + 1600 = \frac{\Sigma x^2}{100}$

$1626.01 = \Sigma x^2$

$162601 = \Sigma x^2$

New $\Sigma x^2 = 162601 + 1600 - 2500$

$= 161701$

$\therefore$ New $\sigma = \sqrt{\frac{161701}{100} - (39.9)^2}$

$= \sqrt{1617.01 - 1592.01}$

$= \sqrt{25}$

$= 5$

UNIT- 4

Probability

Before studying the probability we must define some terms related to probability -

- **Experiment** - Experiment is the process which does not have a fixed outcome. It has several possible outcomes.

- **Event** - Every outcome of an experiment is called an event

- **Sample space** - It is a set containing all the possible outcomes of an experiment.

  There are two types of sample space -

  - **Finite** : If the no. of element in sample space is finite then it is called finite sample space.

  - **Infinite** : If the no. of element in sample space is infinite then it is called infinite sample space.

- **Exhaustive Events** - Collection of those event is called exhaustive event which constitute the sample event.

- **Mutually exclusive event** - Those events are known as mutually exclusive if on occurence of one others don't occur.

- **Independent event** - Two events are called independent if the occurence of one does not affect the occurence of other.

- **Dependent event** - Two events are called dependent event if occurence of one affects the occurence of other.

$$\boxed{P(A) = \frac{\text{No. of favourable cases}}{\text{No. of total cases}}}$$

Demorgan's law -

- $(A \cup B)' = A' \cap B'$
- $(A \cap B)' = A' \cup B'$

Q- Find the sample space of the following experiments-

(a) a coin is tossed thrice

(b) there are 5 students A, B, C, D, E select a team of any two

(c) You are sowing 6 seeds find the number of plants.

(d) A coin is tossed twice, if in second toss head comes a dice is tossed. otherwise coin is tossed.

Solⁿ- (a) HHH, HHT, HTH, HTT, TTT, TTH, THT, THH

(b) AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

(c) 0, 1, 2, 3, 4, 5, 6

(d) HTH, HTT, TTT, TTH, HH1, HH2, HH3, HH4, HH5, HH6, TH1, TH2, TH3, TH4, TH5, TH6

Q- A dice is thrown find the probability of a no. divisible by 3.

Solⁿ- Sample space = 1, 2, 3, 4, 5, 6

$$P(A) = \frac{\text{No. of favourable cases}}{\text{No. of total cases}} = \frac{2}{6} = \frac{1}{3}$$

Addition Rule -

Let there are two events A and B happening then the probability of happening of ~~either~~ anyone event is defined as

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If the both events are mutually exclusive then,

$$P(A \cup B) = P(A) + P(B)$$

Q- If two dices are thrown together, find the probability of getting ~~the probability of~~ a sum of 8 or 12.

Soln- Let A is the event of getting a sum of 8 = {(2,6), (3,5), (4,4), (5,3), (6,2)}

And B is the event of getting a sum of 12 = {(6,6)}

$P(A) = 5/36$ and $P(B) = 1/36$

Since these two events are mutually exclusive here sum rule will be applied,

$P(A \cup B) = P(A) + P(B) = 5/36 + 1/36 = \frac{6}{36} = 1/6$

Q. A bag contains 25 balls bearing the no. 1 to 25, find the probability that the number on a ball drawn is either divisible by 11 or by 3.

Soln- Let A be the event of getting the no. on ball divisible by 11 = {11, 22}

And B be the event of getting the no. on ball divisible by 3 = {3, 6, 9, 12, 15, 18, 21, 24}

$P(A) = 2/25$ and $P(B) = \frac{8}{25}$

$\therefore P(A \cup B) = P(A) + P(B) = \frac{2}{25} + \frac{8}{25} = \frac{10}{25} = \frac{2}{5}$

Q- A card is drawn from a pack of normal 52 cards deck. What is the probability that the drawn card is either a diamond or a queen.

Soln- $P(A) = 13/52$ {P(getting diamond)}

$P(B) = 4/52$ {P(getting queen)}

$P(A \cap B) = 1/52$

$P(A \cup B) = \frac{13}{52} + 4/52 - 1/52$

$= \frac{16}{52} = \frac{4}{13}$

Q- If $P(A) = 0.25$ and $P(B) = 0.15$ and $P(A \cup B) = 0.30$

Find

ⓐ $P(A \cap B)$

ⓑ $P(A \cap \bar{B})$

ⓒ $P(\bar{A} \cap B)$

ⓓ $P(\bar{A} \cap \bar{B})$

ⓔ $P(\bar{A} \cup \bar{B})$

Sol^n Given, $P(A) = 0.25$, $P(B) = 0.15$, $P(A \cup B) = 0.30$

ⓐ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

$$= 0.25 + 0.15 - 0.30$$

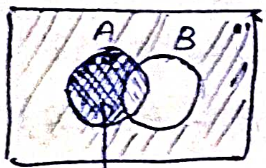
$$= 0.40 - 0.30$$

$$[P(A \cap B) = 0.10]$$

ⓑ $\cancel{P(A \cap \bar{B}) = P(A) \cup P(\bar{B})}$, $P(\bar{A}) = 1 - 0.25 = 0.75$

$\cancel{= 0.25}$ $P(\bar{B}) = 1 - 0.15 = 0.85$

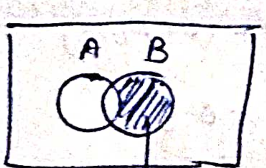
ⓑ



$$P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.25 - 0.10$$

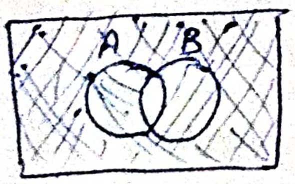
$$= 0.15$$

ⓒ



$$P(\bar{A} \cap B) = P(B) - P(A \cap B) = 0.15 - 0.10 = 0.05$$

ⓓ



$\Rightarrow P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B) = 1 - 0.3 = 0.7$

ⓔ $P(\bar{A} \cup \bar{B}) = 1 - P(A \cap B) = 1 - 0.1 = 0.9$

Q. If probability of getting A contract is 0.25, getting B contract is 0.15 and the getting probability of at least one contract is 0.30. Find

(a) P(getting both contracts) → $P(A \cap B)$

(b) P(getting contract A only) → $P(A \cap B')$

(c) P(getting contract B only) → $P(A' \cap B)$

(d) P(not getting both the contract) → $P(\bar{A} \cap \bar{B})$

Soln Given, $P(A) = 0.25$, $P(B) = 0.15$, $P(A \cup B) = 0.30$

(a) P(getting both contracts) $= P(A) + P(B) - P(A \cup B)$

$= 0.25 + 0.15 - 0.30$

$= 0.40 - 0.30$

$= 0.10$

(b) P(getting contract A only) $= P(A) - P(A \cap B)$

$= 0.25 - 0.10$

$= 0.15$

(c) P(getting contract B only) $= P(B) - P(A \cap B)$

$= 0.15 - 0.10$

$= 0.05$

(d) P(not getting both the contract) $= 1 - P(A \cup B)$

$= 1 - 0.30$

$= 0.7$

Multiplication rule-

when two events A and B occurs and both are independent event then the probability of occurrence of both the events is defined as-

$$\boxed{P(A \cap B) = P(A) \times P(B)}$$

Q- There are two speakers S1 and S2, S1 speaks truth 60% times, where as S2 speaks truth 70% of time. What is the probability that they both contradict each other.

Soln- Let A be the event that S1 speaks truth and B be the event that S2 speaks thruth.

$P(A) = 60\% = \frac{60}{100} = 0.6$

$P(\bar{A}) = 1 - 0.6 = 0.4$

$P(B) = 70\% = \frac{70}{100} = 0.7$

$P(\bar{B}) = 1 - 0.7 = 0.3$

For contradictory statement there are two cases -

Case 1:- S1 - True, S2 - False

$P(A \cap \bar{B}) = P(A) \times P(\bar{B}) = 0.6 \times 0.3 = 0.18$

Case 2:- S1 - False, S2 - True

$P(\bar{A} \cap B) = P(\bar{A}) \times P(B) = 0.4 \times 0.7 = 0.28$

Total probability $= 0.18 + 0.28 = 0.46$

Q- There are three shooters A, B and C, A shoots three out of 6 targets. B shoots 4 out of 6. And C shoots 3 out of 5. find the probability at at least two of them shoots the target.

Soln Let A be the event, A shoots,
B be the event, B shots
C be the event C shoots

$P(A) = \frac{3}{6} = \frac{1}{2} = 0.5$

$P(\bar{A}) = 1 - 0.5 = 0.5$

$P(B) = \frac{24}{63}\!\!\!\!\!\!\!\quad = 0.6$

$P(\bar{B}) = 1 - 0.6 = 0.4$

$P(C) = \frac{3}{5} = 0.6$

$P(\bar{C}) = 1 - 0.6 = 0.4$

Case 1: $P(A \cap B \cap \bar{C}) = P(A) \times P(B) \times P(\bar{C}) = 0.5 \times 0.6 \times 0.4 = 0.12$

case 2: $P(A \cap \bar{B} \cap C) = P(A) \times P(\bar{B}) \times P(C) = 0.5 \times 0.4 \times 0.6 = 0.12$

case 3: $P(\bar{A} \cap B \cap C) = P(\bar{A}) \times P(B) \times P(C) = 0.5 \times 0.6 \times 0.6 = 0.18$

Case 4: $P(A \cap B \cap C) = P(A) \times P(B) \times P(C) = 0.5 \times 0.6 \times 0.6$

$= 0.18$

Total probability $= 0.12 + 0.12 + 0.18 + 0.18$

$= 0.12 + 0.12 + 0.18 + 0.18$

$= 0.6$

Conditional Probability –

Normally the probabilities are calculated in absolute term which means conditions does not have any rule to play like toss of coin.

But in some cases the probability of occurence of one event is dependent on the occurence of other event this type of probability is known as conditional probability.