

Please follow our audio (This Booklet comes with audio)

CHANDANSINGH CLASSES(9883306418/9477365464)

(Dare to become papaji of what you are going to learn)

Visit: www.chandansingh.co.cc

For notes on CPT, CS, IPCC, FINAL, B.Com

Statistics and Statistical Data

Term derived from

STATISTICS

Status (in Latin)

A political State

The term statistics has dual meaning.

Singular Sense

Plural Sense

**In singular sense/Specific View
statistic means technique or
method (e.g. Mean, SD etc.) to
predict some useful result from the
raw data or information**

In plural sense/General view statistics means raw data and information. I.e. quantitative information (record of employee in a factory, income of a nation etc)

DEFINITION: We may define Statistics as the science which deals with the collection, analysis and interpretation of numerical data.

Use of Statistics in Real World: - Suppose that you are the production manager of a car manufacturing co. and you have to determine the no of car's production this year. We may use the technique i.e. used by the TATA NANO. They collect a form with some advance deposit from their customer. This form gives them the idea of demand of NANO car in the market. Using this statistical data collection we can decide the no of car's production in the year.

We can use this technique to determine the average value of particular Stock (using AM) and can determine Risk of a share (using the measure of dispersion).

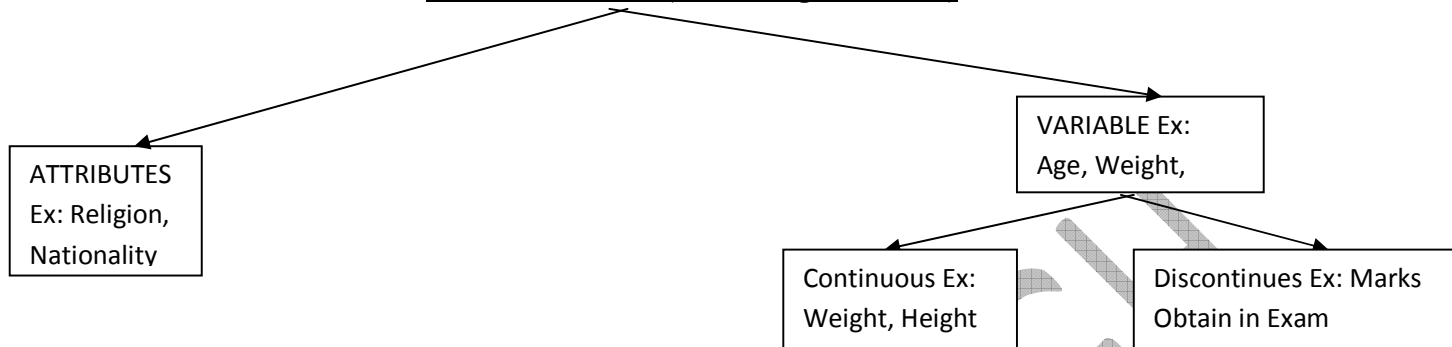
We can determine the relationship between two variables from the past data. Using this relation we can predict the one variable when other is known.

CHANDAN KUMAR SINGH. (9477365464/9883306418/9804943754).www.chandansingh.co.cc

Please follow our audio (This Booklet comes with audio)

Most important concern of statistics at present in our discussion is DATA. Generally data means information.

TYPE OF DATA (According to Nature)

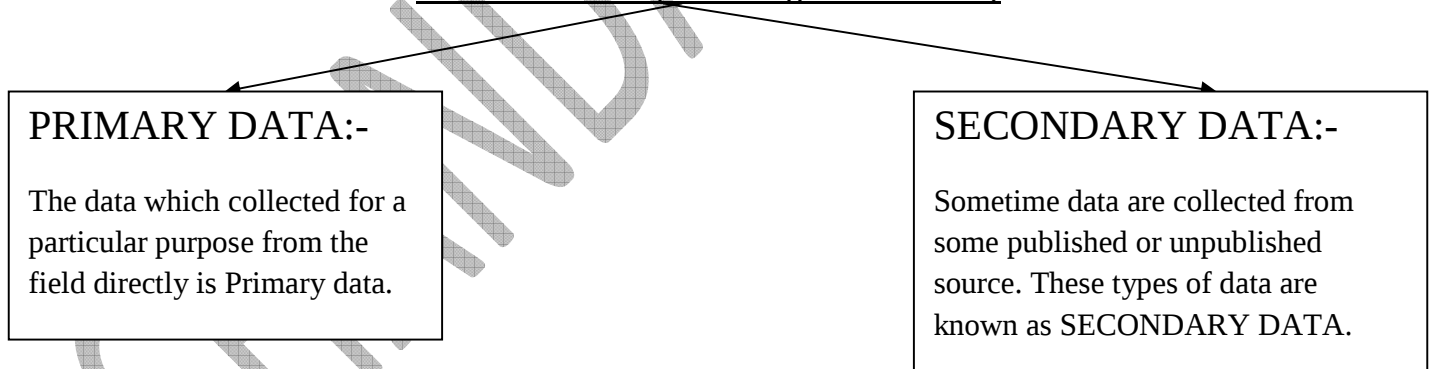


Attributes: - The data which can't take numerical value. Generally these data are represented by some words. For Example Religion of a person (generally answered as Hindu, Muslim, and Sikh etc.), Nationality of Person (Answered as Indian, American, etc.).

Continuous Variable: - The variable which can take any value. E.g. Weight of student (may take any say 45.5 kg., 49.8 kg. etc.).

Discontinuous Variable: - The variable which can't take all value. It only can takes predefined value. For example size of shoe.

TYPE OF DATA (According To Collection)



EX: 1. Suppose you have to collect a data about the age of the person who live in your locality.

You may collect it by yourself or you may contact your municipality for this. The first way is primary and second is secondary.

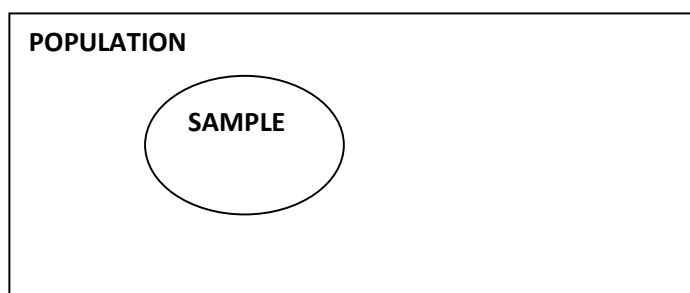
SAMPLE AND POPULATION

Please follow our audio (This Booklet comes with audio)

Suppose your mummy ask you to buy 50 Kg. of rice from the market. You go the shop and carry some small amount of rice in hand and check it out. If it is OK you ask shopkeeper to give that bag of the rice.

Here the small amount of rice in you hand is the **SAMPLE** and the bag of the rice is the **POPULATION**. And the technique that you applied to predict/estimate the quality of rice of full the full bag is **STATISTIC**.

We may Sample and Population using **VENN DIAGRAM**.



Classification and Presentation of DATA

Suppose you locality shopkeeper hire you to collect the price of different type of rice from the wholesale market. You collect them as follow.

Price of different type of RICE (Raw Data)

RATNA-35 per kg, MINIKET-40 per kg, BASKET- 37 per kg, BASMATI- 42 per kg, MANSURI-26 per kg, KALADAN -35 per kg, BASDAN- 28 per kg, KATRIN- 30 per kg.

This is the raw data that you have collected by yourself for a particular purpose. So this is the primary data.

Now you can present these data using three methods.

THE FIRST MEATHOD (TEXTUAL PRESENTATION):

Price of RATNA, MINIKET, BASKET, BASMATI, MANSURI, KALADAN, BASDAN and KATRIN are Rs.35, RS.40, Rs.37, Rs.42, Rs.26, Rs.35, Rs.28, and Rs.30 per Kg. respectively.

THE SECOND MEATHOD (TABULAR MEATHOD):

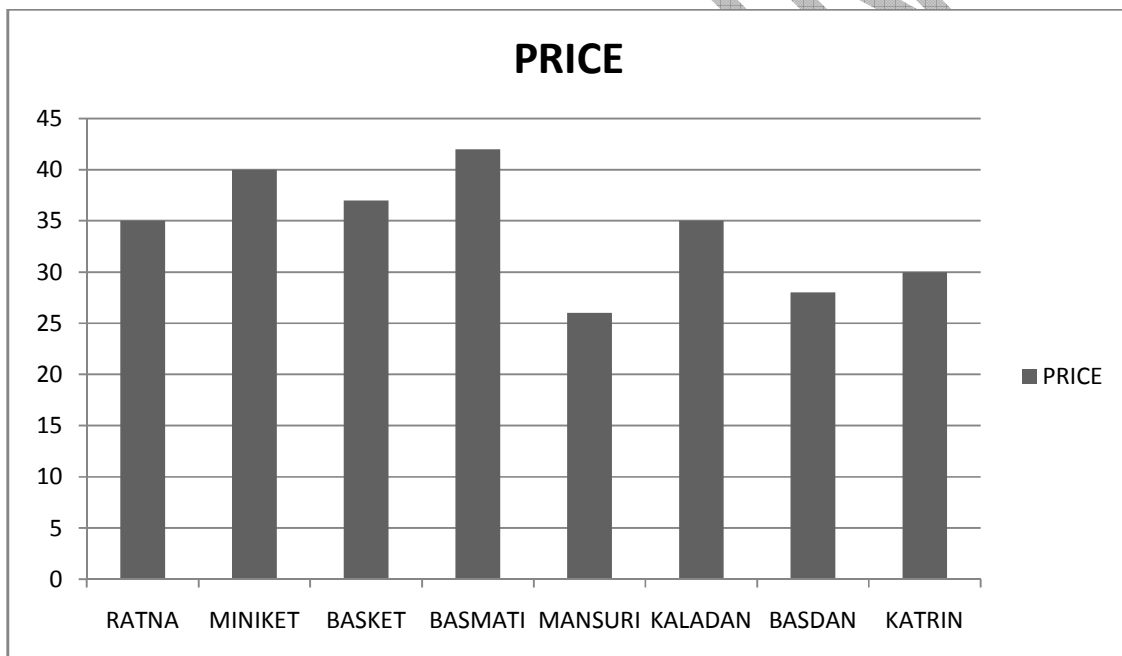
<u>TYPE</u>	<u>PRICE</u>
RATNA	35
MINIKET	40

Please follow our audio (This Booklet comes with audio)

BASKET	37
BASMATI	42
MANSURI	26
KALADAN	35
BASDAN	28
KATRIN	30

THE THIRD MEATHOD

(GRAPHICAL MEATHOD):



Here we present the Shopkeeper 'rice price in three ways. Out of them graphical approach is most proving. It is not possible to determine the exact price of rice at from the graphical approach directly. So generally this approach is used when comparison is to be done.



Please follow our audio (This Booklet comes with audio)



My friend please pick up the phone your primary school principle madam calling you for collecting a data of age of students in the school.

Suppose there are 50 students in school. You collect the following data.

5, 4, 7, 8, 3, 7, 5, 6, 7, 4, 6, 4, 8, 5, 4, 3,

6, 5, 4, 3, 7, 8, 6, 8, 5, 4, 7, 5, 6, 7, 8, 4, 7,

6, 4, 3, 6, 6, 4, 7, 4, 3, 6, 4, 3, 3, 6, 7, 5, 8

Now you present this one to your Madam Priya. Here it is seen that the ages 3, 4, 5, 6, 7, 8 repeated. So we will make a table with two columns. First column shows the age and second one shows the frequency (i.e. repetition of that age). This is generally known as frequency distribution.

FREQUENCY: - Repetition of particular observation in a given data. E.g. frequency of 5 is 6.

AGE	TALLEY MARK	Frequency
3		7
4		11
5		7
6		10
7		9
8		6

Here we draw the table from the age that we have collected already. First we read one age 5 and draw tally in front of 5. Similarly for other. After four Tally when we draw 5th one we just cut a line on 4 Tally.

Now I am very happy to inform you that you are an experienced statistician. Congratulation my friend, you are now selected to collect the age of the residents in a locality by KOLKATA MUNISPALITY. You collected data of 150 people are given below.

Age of the residents									
25	26	24	23	26	07	78	36	24	65
12	45	52	57	26	05	52	52	41	84
15	47	43	76	43	68	49	80	72	12
16	49	76	25	45	06	59	80	04	45
17	25	58	76	25	45	68	25	36	52

Please follow our audio (This Booklet comes with audio)

13	53	43	48	26	12	25	43	78	68
14	65	16	65	50	98	65	60	15	75
20	36	58	43	60	16	73	86	65	14
22	29	48	41	46	73	25	73	74	45
64	76	76	71	75	75	48	15	47	47
58	85	82	06	46	49	89	12	35	68
89	13	48	56	89	68	76	48	46	25
74	75	95	25	65	95	15	76	76	13
54	45	43	42	46	25	46	25	87	78
14	23	52	37	23	41	13	48	45	74

Here the observation may lie between large intervals. Let me define a term

RANGE: - This is the difference between maximum and minimum value of the observation.

$$\text{RANGE} = \text{MAX VALUE} - \text{MIN VALUE}$$

If the range is a big no. then we cannot present this data in a table as presented in last case. Here we make group data because number of different observation is many. So, we make different class for ages.

This is a Frequency distribution:-

Here, **Range** = $94 - 4 = 94$

So, here we divide the data into ten groups as shown below.

CLASS	TALLY MARK	Frequency
0-10		
10-20		
20-30		
30-40		
40-50		
50-60		
60-70		
70-80		
80-90		

Please follow our audio (This Booklet comes with audio)

90-100		
--------	--	--

CLASS:-

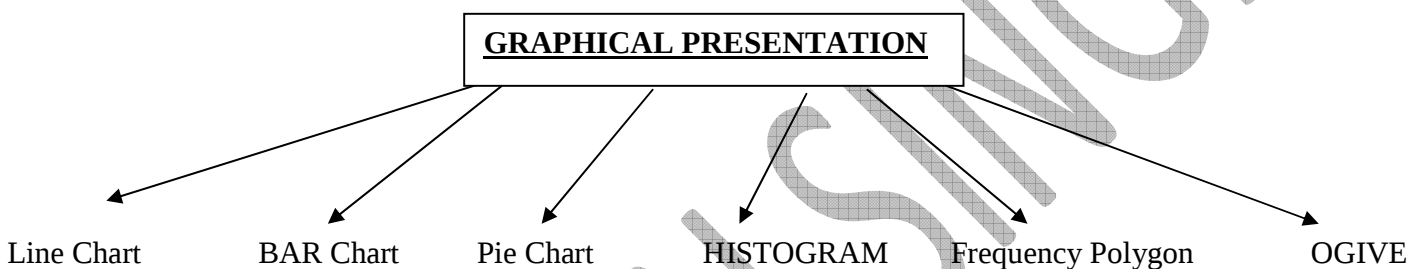
CLASS LIMIT/INTERVAL:- Separate one class in a grouped frequency distribution from another. The limits could actually appear in the data and have gaps between the upper limit of one class and the lower limit of the next. Table show the sale of packet HALDI (for different Quantity)

WEIGHT(IN GRAM)	SALES
50-500	150
550-1000	200
1050-1500	250
1500-2000	150

CLASS BOUNDRY:- Separate one class in a grouped frequency distribution from another. The boundaries have one more decimal place than the raw data and therefore do not appear in the data. There is no gap between the upper boundary of one class and the lower boundary of the next class. The lower class boundary is found by subtracting 0.5 units from the lower class limit and the upper class boundary is found by adding 0.5 units to the upper class limit.

CLASS INTERVAL	CLASS BOUNDRY	CLASS BOUNDRY		FREQU ENCY	MIDVAL UE(X)	Width of the class	Frequ ency densit y	Relativ eFrequ ency
		LOWER	UPPER					
1-10	0.5-10.5	0.5	10.5	5	5.5	10	0.5	0.11
11-20	10.5-20.5	10.5	20.5	7	15.5	10	0.7	0.15
21-30	20.5-30.5	20.5	30.5	8	25.5	10	0.8	0.18
31-40	30.5-40.5	30.5	40.5	10	35.5	10	1	0.22
41-50	40.5-50.5	40.5	50.5	9	45.5	10	0.9	0.2
51-60	50.5-60.5	50.5	60.5	6	55.5	10	0.6	0.13

- Width of the class measure using Class Boundary not using class interval
- CLASS WIDTH= UPPER BOUNDARY – LOWER BOUNDARY
- FREQUENCY DENSITY= CLASS FREQUENCY $\div$ CLASS WIDTH
- RELATIVE FREQUENCY = CLASS FREQUENCY $\div$ TOTAL FREQUENCY
- LOWER CLASS LIMIT= LOWER BOUNDARY- $\frac{1}{2}$ (DIFFERENCE)
- UPPER CLASS LIMIT= UPPER BOUNDARY – $\frac{1}{2}$ (DIFFERENCE)
- DIFFERENCE= Lower Boundary of nth class – Upper of (n-1)th class



LINE CHART: - This is a two dimensional presentation of data. In this chart we measure one on the X axis and other on the Y axis. Let us explain this with the help of an example.

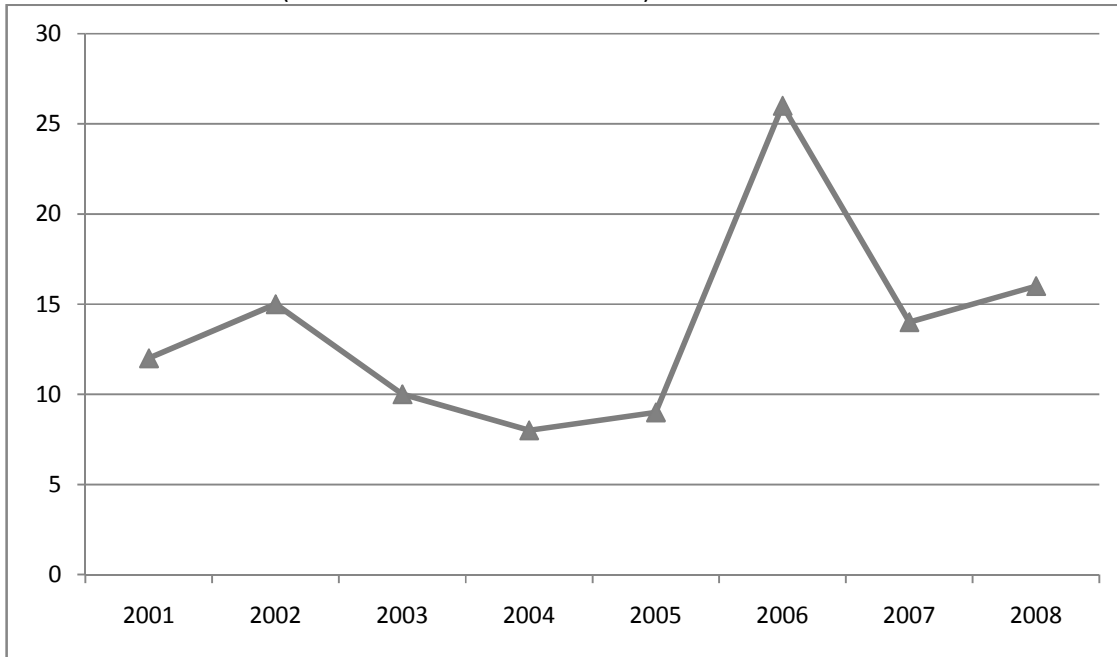
The table shows the chronological trend of production of Rice.

Year	2001	2002	2003	2004	2005	2006	2007	2008
Production	12	15	10	8	9	26	14	16

Table: X

SOLUTION:- following is line graph of the given chronological trend.

Please follow our audio (This Booklet comes with audio)



If there are more than one variable then we draw another variable line on the same graph.

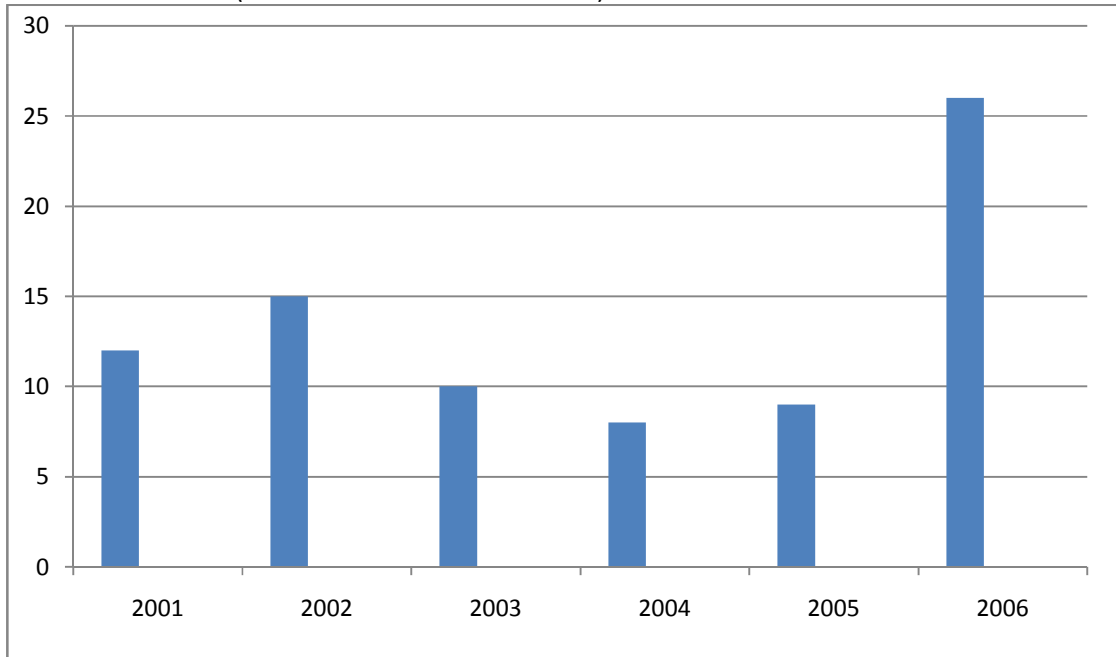
BAR CHART: -

1. The width of the each bar is same.
2. We may use vertical or horizontal bar chart.
3. If there is any negative figure then it drawn opposite side.

This is explaining in the example.

Bar chart of Table: X

Please follow our audio (This Booklet comes with audio)



PIE CHART: - Pie **charts** are useful to compare different parts of a whole amount. They are often used to present financial information. E.g. A Company's expenditure can be shown to be the sum of its parts including different expense categories such as salaries, borrowing interest, taxation and general running costs (i.e. rent, electricity, heating etc).

Example: - The following are the expenses occurred by the company in a particular year. Prepare a pie chart to present it.

Expenses	Electricity	Rent	Salary	AMC	Office Expenses
Amount(\$)	\$500	\$700	\$300	\$250	\$750

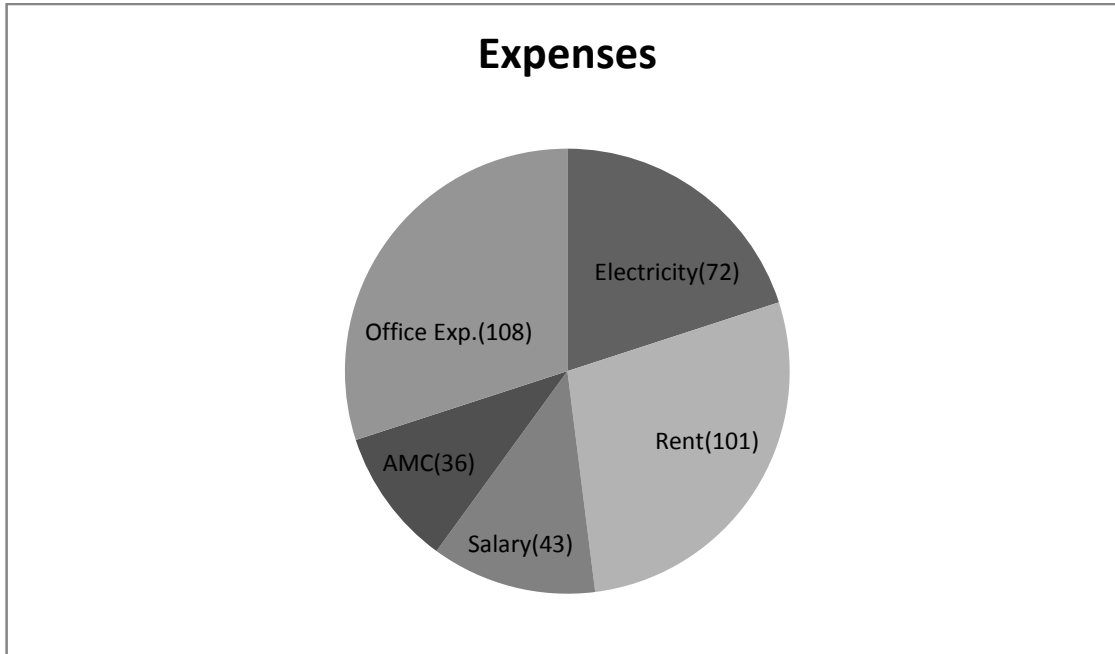
Solution: - Following is the working table for pie chart.

Expenses	Amount(\$)	Calculation for degree	Expenses In degree
Electricity	\$500	_____	72
Rent	\$700	_____	101(real is 100.8)
Salary	\$300	_____	43(real is 43.2)

Please follow our audio (This Booklet comes with audio)

AMC	\$250	_____	36
Office Exp.	\$750	_____	108

Following is the graph



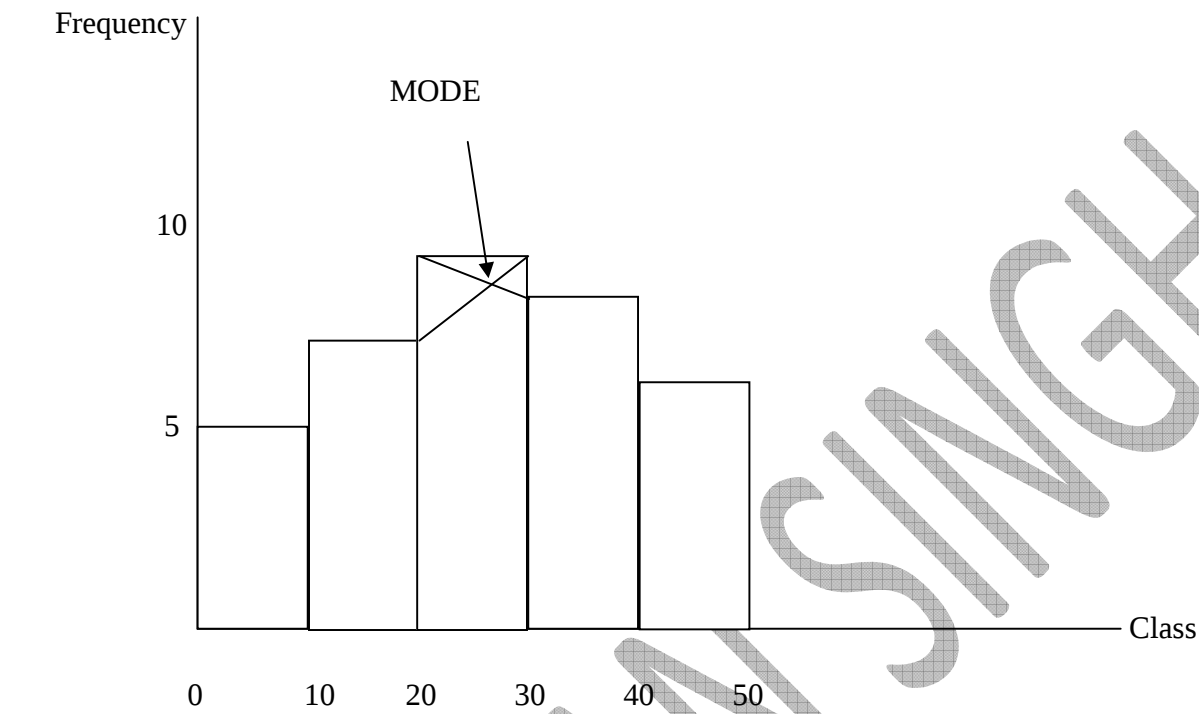
Up to now we discuss about the graphical presentation of Ungroup data without frequency now see some technique how to present frequency distribution.

HISTOGRAM: - This is a Graphical presentation to present the data which have frequency. The histogram is the summary graph showing a count of the data points falling in various ranges.

CLASS	FREQUENCY
0-10	5
10-20	7
20-30	9
30-40	8
40-50	6

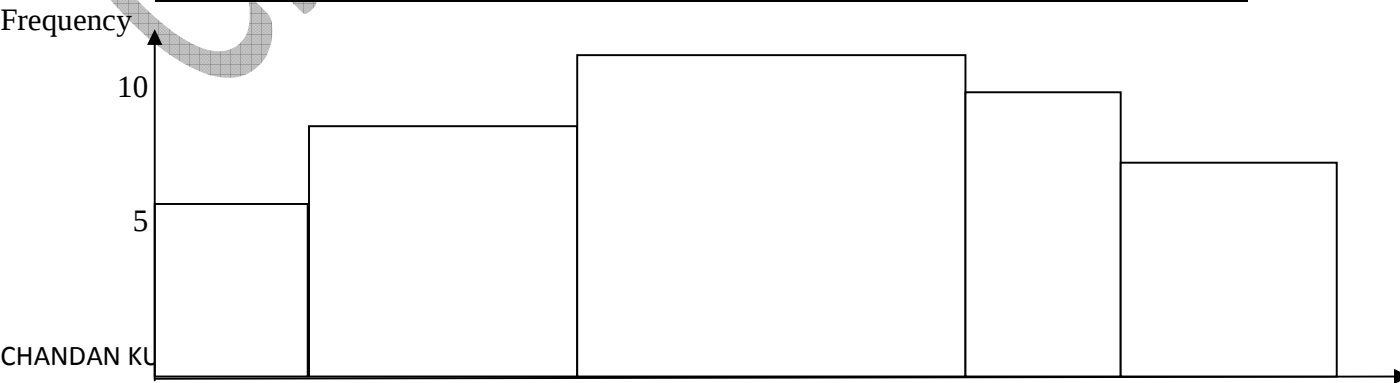
Please follow our audio (This Booklet comes with audio)

The HISTOGRAM of the above graph is shown below.



Here Width of the BAR depends on the class width. This is shown in the below example.

CLASS	FREQUENCY	Class Width	Frequency Density	Relative Frequency
0-20	5	20	$4/20=4$	$5/35=0.14$
20-50	7	30	$7/30=0.23$	$7/35=0.2$
50-100	9	50	$9/50=0.18$	$9/35=0.26$
100-120	8	20	$8/20=0.4$	$8/35=0.23$
120-150	6	30	$6/30=0.2$	$6/35=0.17$



Please follow our audio (This Booklet comes with audio)

0 20 50 100 120 150 class

CUMULATIVE FREQUENCY DISTRIBUTION and OGIVE:-

There are two type of cumulative distribution one is “LESS THAN” and another is “MORE THAN”. Less than frequency shows the number of observation less than a particular specified value. More than frequency shows the number of observation more than a particular specified value.

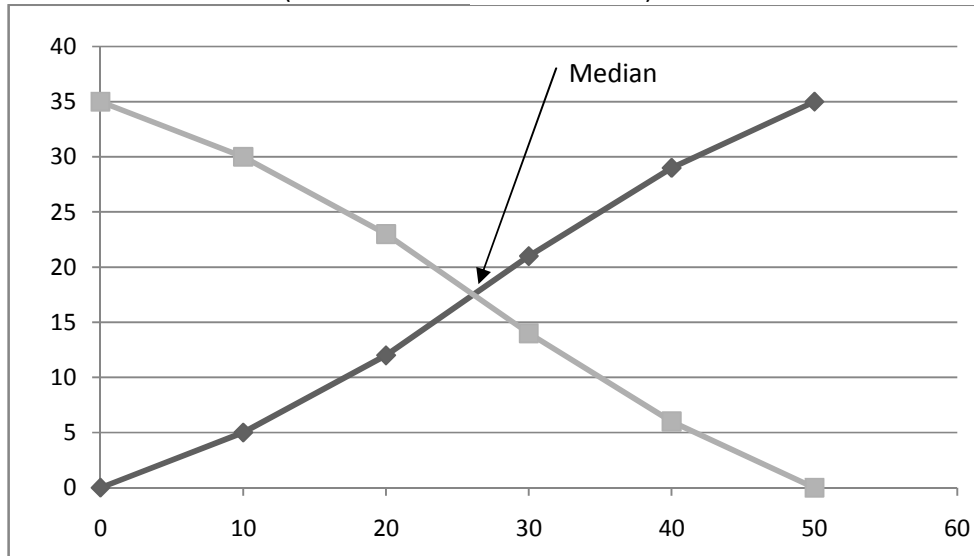
Illustration:

CLASS	FREQUENCY	Cumulative Frequency	
		Less Than	More Than
0-10	5	5	$30+5=35$
10-20	7	$5+7=12$	$23+7=30$
20-30	9	$12+9=21$	$14+9=23$
30-40	8	$21+8=29$	$6+8=14$
40-50	6	$29+6=35$	6

In the column “LESS THAN” the first row is 5 it means number of observation less than 20 is 5. Similarly in the 2nd row 12 shows the number of observation less than 50 is 12. And so on.

In the column “MORE THAN” the first row is 35 it means number of observation more than 0 is 35. Similarly in the 2nd row 30 shows the number of observation more than 20 is 30. And so on Now let us explain this with the help of Graph, i.e. OGIVE. The graphical representation of cumulative frequency diagram is known as OGIVE.

Please follow our audio (This Booklet comes with audio)



Graph of Ogive.

PRACTICE TIME:-

- Go to your locality and collect the following information from Local shopkeepers
Shop's Name, Number of product, number of employee, address, Phone number. Present the raw data and then prepares a tabular presentation of the raw data.
- Last time when I go for the ZOO school collect the age of 50 Elephant. This is as follows

4	5	3	5	6	7	5	6	5	4
5	3	4	6	4	7	4	5	3	5
3	4	5	3	5	4	6	4	4	5
7	4	6	4	5	7	5	6	4	3
4	5	3	4	6	7	4	6	4	5

Prepare a frequency table for the above data.

- Out of 1000 persons 25% were industrial worker and rest were agriculture worker. 300 person enjoyed world cup matches on TV. 30% of people who had not watch world cup matches were industrial workers. What is the number of agricultural worker who had enjoyed world cup matches on TV?
- Prepare an OGIVE from following information.

CLASS	0-10	10-20	20-30	30-40	40-50	50-60	60-70
FREQUENCY	8	10	12	15	13	9	8

- Draw a Pie Chart from following Information.

EXPORT	Wheat	Rice	Sugar	Mustard	Maize	Coconut
Quantity(TON)	200	250	350	450	150	400

- Draw the Histogram from following data.

CLASS	0-10	10-30	30-70	70-90	90-110
FREQUENCY	8	12	15	13	10

From the histogram prepare the frequency polygon.

- Draw a bar chart for the following data.

EXPORT	Wheat	Rice	Sugar	Mustard	Maize	Coconut
--------	-------	------	-------	---------	-------	---------

Please follow our audio (This Booklet comes with audio)

Quantity(TON)	200	250	350	450	150	400
---------------	-----	-----	-----	-----	-----	-----

8. Explain the difference between bar chart and Histogram.
9. Explain the concept of attributes, continuous variable, and discontinuous variable with example.

CHANDAN SINGH