STATISTICS

In our daily life, we have to collect facts which help us in answering most of the questions concerning the world in which we live. The facts we collect are often number facts such as the number of runs scored by Indian team against Pakistan.

The methods and techniques of collection, presentation, analyses and interpretations of numerical data in a logical and systematic manner so as to serve a purpose is known as '*statistics*'.

Statistics is a mathematical science pertaining to the collection, analysis, interpretation and presentation of data.

MEANING OF STATISTICS

Statistics is concerned with scientific method for collecting and presenting, organizing and summarizing and analyzing data as well as deriving valid conclusions and making reasonable decisions on the basis of this analysis.

ORIGIN AND GROWTH OF STATISTICS (HISTORY)

The word '*statistics*' and '*statistical*' are derived from the Latin word *status*, means political state.

- The German stastistik, first introduced by Gottfried ache wall (1749), originally designated the data analysis of state.

- It was used by the British mainly for administrative and governmental bodies.
- In particular census provides regular information about the population.
- Today however statistics had broadened Far beyond the service of a state or government ,it includes areas such as
- Business
- Natural and social sciences and
- Medicines
- Before 3000B.C. the Babylonians used small clay tablets' to record tabulations of agricultural yields and of commodities bartered or sold.
- The Egyptians analyzed the population and material wealth of their kingdoms.
- The Roman Empire was the first government to gather extensive data about population, area and wealth of territories that they controlled.

FUNDAMENTAL

CHARACTERSTICS OF STAISTICS

- They are related to each other and are comparable.
- They are aggregate of facts and not a single observation . Statistics do not take into account individual cases
- Statistics data are numerically expressed.
- Statistics are collection of data in a systematic manner.
- Statistics are collected for a predetermined purpose.
- Statistics deals with group and doesn't study individually.
- Statistics laws are not exact, they are true only on averages.

iCBSE

- The data collected by someone else, other than the investigator, are known as secondary data.
- The data obtained in the original form are called ungrouped data or raw data.
- An arrangement of raw numerical data in ascending or descending order of magnitude is called array.

SUB TOPICS

SOME RELATED DEFINATIONS

These fig. are in the ascending order

4,5,8,18,28,29,29,31,40,40,43,43,46,46,46,47,47,50,50,55,55,70,71,75, 75,80,90

Marks	**no. of students**	**Marks**	**No. of students**
4	1	46	3
5	1	47	2
8	1	50	2
18	2	55	3
28	1	70	1
29	2	71	1
31	1	75	2
40	2	80	1

The above data in the ascending order is called an '**arrayed**' and the way of arrangement is called an 'array'.

iCBSE.com

iCBSE

The way of arrangement of data in the table is known is known as '**frequency distribution**'.

Marks are called "**variates**" the no. of students who secured a particular no. of marks are frequency of variates is called "**frequency of the variate**".

The number of times a number has been repeated is called the "**frequency of the variate**".

CONTINOUS: Quantities which can take all numerical values within a certain interval .

DISCONTINOUS: Quantities or variable which can take only a finite set of values.

Each groups into which a raw data is condensed is called a "**class".** The size of class is known as the **"class interval".**

For ex. 10 is the class interval of class "**0-10**".

Each class is bounded by 2 fig. which are called the "**lower limits** "and "**20**"is the "**upper limit**".

The difference between the upper limit of the class and the lower limit of class is called as the "**class size**".

The value which lies midway between lower and upper limits of a class is known as its "**mid value or class mark**".

Class mark = **upper limit + lower limit**

The difference between the two extreme observations in an arranged data i.e. the difference between the maximum and minimum values of observations is known as the "**Range**".

Three measures of central tendency are

- **Mean**
- **Mode**
- **Median**

MEAN

Median of groped data :if $x_1, x_2, x_3, \ldots\ldots x_n$ are variables of a variable x , then the arithmetic mean or simply mean of these values is denoted by X and is defined as $\mathbf{X= \frac{x_1 + x_2 + x_3 + \ldots\ldots x_n}{n}}$

or $\mathbf{X=\sum_{i=1}^{n} x_i / n}$

ALGORITHM:

Step I= Prepare the frequency table in such a way that its first column consists of the values of the variate and the second column the ∑

Step II=multiply the frequency of each row with the corresponding values of variable to obtain third column containing fix;

Step III= Find the sum of all entries in column **III** to obtain $\mathbf{\sum f_i x_i}$.

Step IV= Find the sum of all the frequencies in column **II** to obtain $\mathbf{\sum f_i = N}$

Step V= Use the formula: $X = \Sigma f_i x_i / N$

iCBSE

For Ex= Find the missing frequencies in the following frequency distribution if it is known that the mean of the distribution is 1.46 No. of accidents (x): 0 1 2 3 4 5 Total

Frequency (f): 46 ? ? 25 10 5 200

COMPUTATION OF ARITHMATIC MEAN

X_i	f_i	f_ix_i
0	46	0
1	f_1	f_1
2	f_2	$2f_2$
3	25	75
4	10	40
5	5	25

$\sum f = N = 86+f_1+f_2$ $\sum f_1x_1=140+f_1+2f_2$

N= 200

- $200= 86 +f_1+f_2$
- $f_1+f_2=114+f_2+f_2$
- $f_1+f_2=114$

Also, Mean=1.46 ----------1

- $1.46= \sum f_ix_{i/N}$
- $1.46=^{140+f1+2f2}/_{200}$
- $292=140+f_1+2f_2$
- $F_1+2f_2=150$ ----------2

Solving 1,2

iCBSE.com

iCBSE

- F_1=76 and f_2=38

STEP DEVIATION METHOD

Step I- Obtain the frequency distribution and prepare the frequency table in such a way that its first column consists of the values of the variable and the second column corresponding frequencies.

Step II- Choose a number 'A'(generally known as the assumed mean) and take deviations $d_i=x_i-A$ about A. Write these deviations against the corresponding di's in the **IV** column.

Step IV- Multiply the frequencies in second column with the corresponding u_i's in **IV** column to prepare **V** column of f_iu_i.

Step V- find the sum of all entries in **V** column to obtain $(\sum^{n}_{i=1} f_ix_i)$ and the sum of all frequencies in column to obtain $N=(\sum^{n}_{i=1}f_i)$. Use formula:$X=A+h\{{}^{1}/_{N}\sum_{i=0}^{n} f_ix_i\}$

MEDIAN

The median is the middle value of a distribution is the value of the variable which divides it into two equal parts.

Step I-Arrange the observation $x_1,x_2,\ldots\ldots\ldots x_n$ in ascending or descending order of magnitude.

Step II- Determine the total no. of observation, say, n.

Step III- If n is odd , then median is the value of $\left(\mathbf{n+\frac{1}{2}}\right)$ observation.

iCBSE

For ex.-Calculate the median from the following distribution:

Class:	5-10	10-15	15-20	20-25	25-30
Frequency:	5	6	15	10	5

30-35	35-40	40-45
4	2	2

SOLUTION: First cumulative table to complete median.

CLASS	FREQUENCY	CUMULATIVE FREQUENCY
5-10	5	5
10-15	6	11
15-20	15	26
20-25	10	36
25-30	5	41
30-35	4	45
35-40	2	47
40-45	2	49

SO,

N=49 & $^{N}/_{2}$=24.5→The cumulative frequency just greater than $N/_2$ is 26 and corresponding class is 15-20

(Median class) L=15, f=15, F=11, h=5

:. Median = $\left(\frac{^{l+N}}{_2}-F\right)/_f = {}^{15+24.5-11}/_{15}*5 = 19.5$

MODE

iCBSE

The mode or modal value of a distribution is that value of the variable for which the frequency is maximum. In order to compute the mode of a series of individual observations. We first convert it into a discrete series frequency distribution by preparing a frequency table. From the frequency table, we identify the value having maximum frequency. The value of variable to obtain is the mode or modal value.

FOR EX.

Obtain the value of the following:

l → lower limit

h → width

f_x → frequency

f_1 → frequency of the class preceding

f_2 → frequency of the class following.

MODE= $l + {}^{(f-f1)}/_{(2f-f1f2)} * h$

Relationship among mean, median and mode

MODE= 3median – 2mean

OR MEDIAN = mode+⅔ (mean – mode)

OR MEAN = mode + $^3/_2$ (median – mode)

PIE-CHART

iCBSE.com

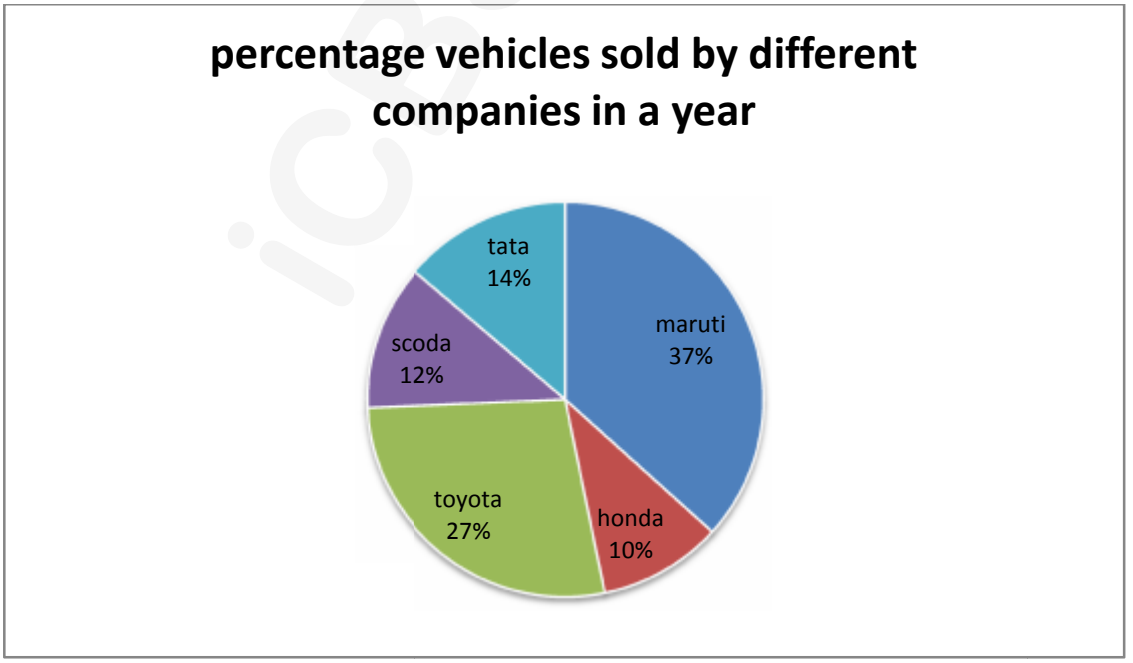
A pie – chart displays data as a percentage as a percentage of the whole Each pie has a label and percentage . A total data no. is included . These have a circle divided into parts or sectors of different sizes to show different amounts of data.

ADVANTAGES

- Visually appealing
- Shows % of total for each category

DISADVANTAGE

- No exact numerical data.
- Hard to compare two data sets.
- "Other" category can be a problem.
- Total unknown unless specified.
- Best for 3 to 7 only.



iCBSE

BAR - GRAPH

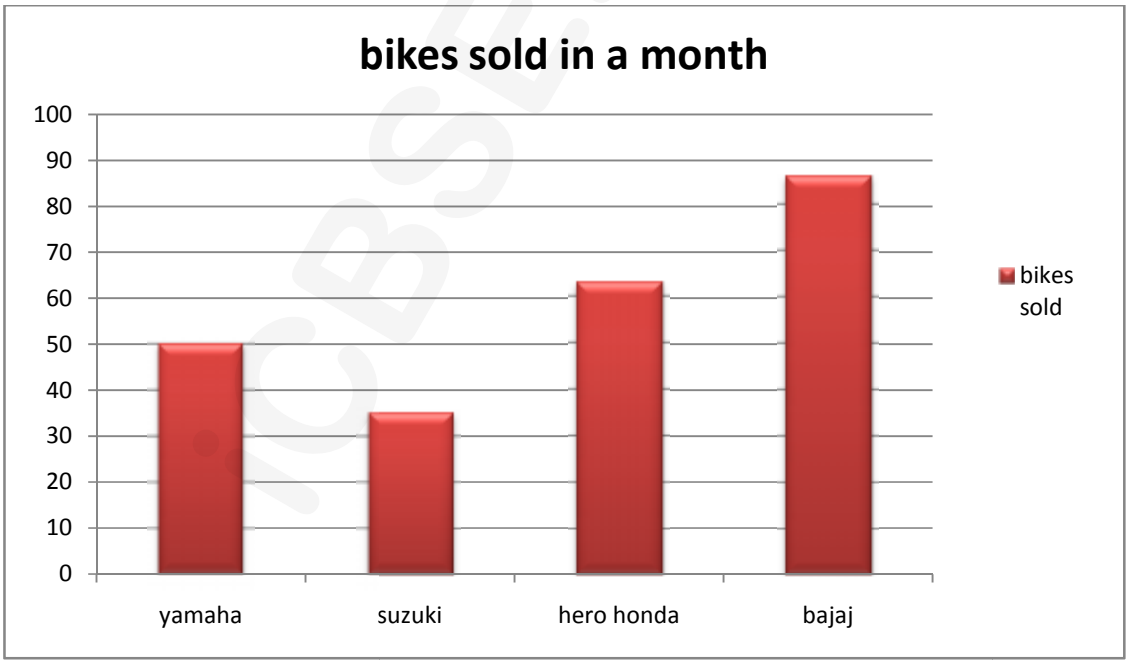
A bar graph display data in separate columns. It data is on a continous scale, such as height, the bars touch each other. The bars can be vertical or horizontal.

ADVANTAGE

- Visually strong.
- Can compare 2 or 3 data sets.

DISADVANTAGE

- Graph categories can be recorded to emphasis certain effects.



iCBSE.com

iCBSE

LINE GRAPH

A line graph plots continuous data as points and then joins them with a line. Multiple data sets can be grouped together, but a key must be used.

ADVANTAGES

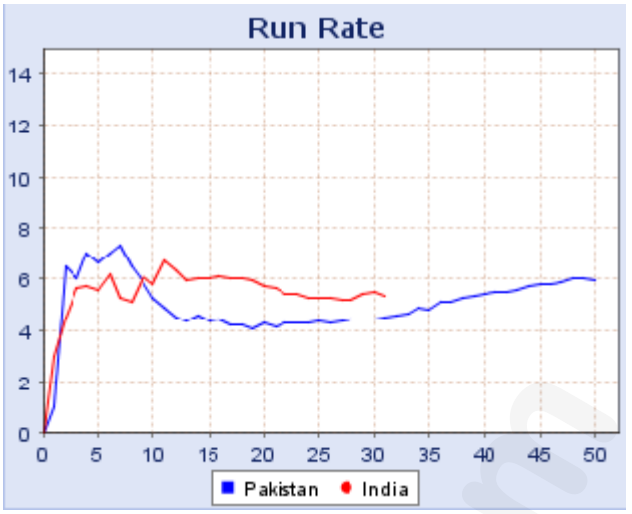
- Can compare multiple continuous data sets easily.
- Interim data can be inferred from graph line

DISADVANTAGE

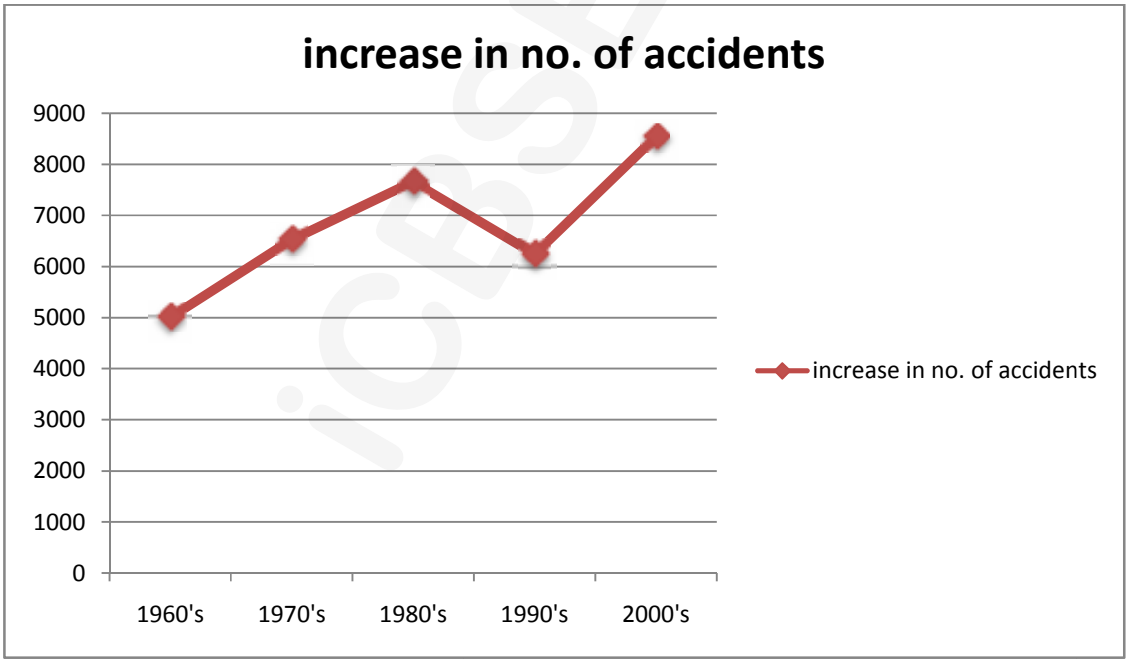
- Use only with continuous data.

Run Rate Graph

iCBSE.com



LINE GRAPH



iCBSE.com

iCBSE

HISTOGRAM

A histograph displays continuous data in order column. Categories are of continuous measures such as time, inches, temperature, etc.

ADVANTAGES

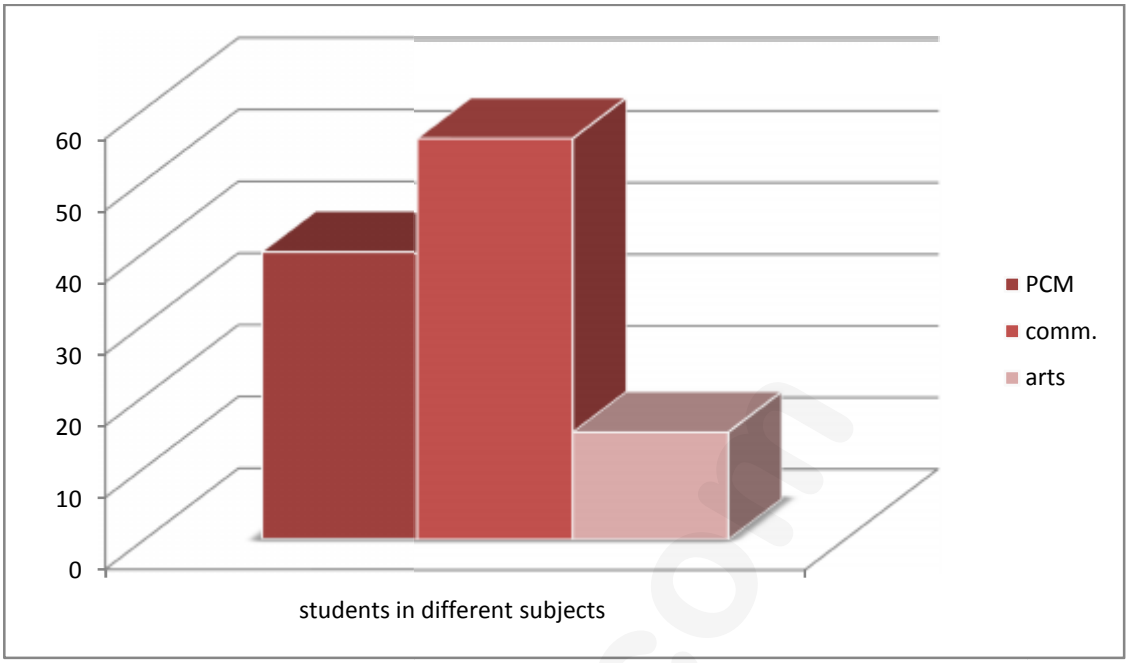
- Visually strong
- Can compare to normal curve.
- Usually vertical axis is a frequency count of items falling into each category.

DISADVANTAGES

- Cannot read exact values because data is grouped into categories.
- More difficult to compare two data sets
- Use only with continuous data.

HISTOGRAM

iCBSE.com



USES AND APPLICATIONS OF STATISTICS

Statistics and its studies have been used to answer questions such as:-

INDUSTRIES AND BUSINESS

- Report of early sales & comparison others.
- It shows where the factory or its sales lack and where they are good

iCBSE

AGRICULTURE

- What amount of crops are grown this year in comparison to previous year or in comparison to required amount of crop for the country
- Quality and size of grains grown due to use of different fertilizer.

FORESTERY

- How much growth has been occurred in area under forest or how much forest has been depleted in last 5 years?
- How much different species of flora and fauna have increased or decreased in last 5 years?

EDUCATION

- Money spend on girls education in comparison to boys education?
- Increase in no. of girl students who seated in who seated for different exams?
- Comparison for result for last 10 years.

ECOLOGICAL STUDIES

- Comparison of increasing impact of pollution on global warming?
- Increasing effect of nuclear reactors on environment?

MEDICAL STUDIES

- No. of new diseases grown in last 10 year.
- Increase in no. of patients for a particular disease.

iCBSE.com

iCBSE

SPORTS

- Used to compare run rates of to different teams.
- Used to compare to different players.

CONCLUSION

Statistics has been a great learning experience and is a very interesting experience and an important topic that is very-very helpful for people of all ages and for teacher to clear their concept and increase their intelligence level.

iCBSE.com