www.prepnext.com

Notes Prepared By:
RAKESH AGARWAL
M.Com, MBA, FIII
E-mail: messagerakesh@gmail.com
WhatsApp No: 8486118428

Business Statistics

Unit 1

Q: What do you understand by Statistics? What are the characteristics of Statistics? *(www.prepnext.com)*

Ans.:
The word 'statistics' comes from the Italian word 'statista' (meaning "Statesman") or the German word 'statistik' which means a Political State. It was first used by Professor Gottfried Achenwall in 1749 to refer to the subject-matter as a whole. The word 'statistics' appeared for the first time in the famous book, 'Elements of Universal Erudition' by Baron J.F. Von Bielefeld.

DEFINITIONS:

According to Yule and Kendall, "By Statistics we mean quantitative data affected to a marked extent by multiplicity of causes".

Prof. Horace Secrist defined statistics as: "By Statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other".

Characteristics of Statistics:

Numerical data must possess certain characteristics in order that they may be called statistics. These are as follows:

(i) **Statistics are Aggregates of Facts**: Single and isolated figures are not statistics for the simple reason that such figures are unrelated and cannot be compared. A single figure relating to production, sale, birth, employment, purchase, accident, etc., cannot be regarded statistics although aggregates of such figures would be statistics because of their comparability and relationship as parts of a common phenomenon.

(ii) **Statistics are Affected to a Marked Extent by Multiplicity of Causes**: Generally speaking, facts and figures are affected to a considerable extent by a number of forces operating together. For example, statistics of production of rice are affected by the rainfall, quality of soil, seeds and manure, method of cultivation, etc. The same is true of statistics of prices, imports, exports, sales, profits, etc.

(iii) **Statistics are Numerically Expressed**: All statistics are numerical statement of facts, i.e., expressed in numbers. Qualitative statements such as 'the population of India is rapidly increasing' do not constitute statistics. The reason is that such statements are vague and one cannot make out anything from them. On the other hand, the statement, "The population is India is expected to increase from 121 crore in 2011 to 137 crore in 2021" is a statistical statement.

(iv) **Statistics are Enumerated or Estimated According to Reasonable Standards of Accuracy**: Facts and figures about any phenomenon can be derived in two ways, viz., by actual counting and measurement or by estimate. Estimates cannot be as precise and accurate as actual counts or measurements. The degree of accuracy desired largely depends upon the nature and object of the enquiry. However, it is important that reasonable standards of accuracy should be attained, otherwise numbers may be altogether misleading.

(v) **Statistics are Collected in a Systematic Manner**: Before collecting statistics a suitable plan of data collection should be prepared and the work carried out in a systematic manner. Data collected in a haphazard manner would very likely lead to fallacious conclusions.

(vi) **Statistics are Collected for a Predetermined Purpose:** The purpose of collecting data must be decided in advance. The purpose should be specific and well-defined. A general statement of purpose is not enough.

(vii) **Statistics Should be placed in Relation to Each Other:** If numerical facts are to be called statistics, they should be comparable. Statistical data are often compared period-wise or region-wise. For instance, the population of India at a particular point of time may be compared with that of earlier years or with the population of other countries. Valid comparisons can be made only if the data are homogeneous, i.e., relate to the same phenomenon or subject.

In the absence of the above characteristics, numerical data cannot be called statistics and hence "all statistics are numerical statements of facts but all numerical statements of facts are not statistics".

Q: What are the functions of Statistics? *(www.prepnext.com)*

Ans.:

The following are the important functions of the science of Statistics:

- It presents facts in a definite form.
- It simplifies mass of figures
- It facilitates comparison
- It helps in formulating and testing hypothesis
- It helps in prediction
- It helps in the formulation of suitable policies

EXPLANATION:

Definiteness: One of the most important functions of statistics is to present general statements in a precise and definite form. Statements of facts conveyed in exact quantitative terms are always more convincing than vague utterances. Statistics presents facts in a precise and definite form and thus helps proper comprehension of what is stated.

Condensation: Statistics helps in condensing mass of data into a few significant figures. In away, statistical methods present a meaningful overall information from the mass of data.

Comparison: Unless figures are compared with the others of the same kind, they are often devoid of any meaning. Comparisons convey a definite meaning.

Formulating and Testing Hypothesis: Statistical methods are extremely helpful in formulating and testing hypothesis and to develop new theories. Hypothesis can be tested by appropriate statistical tools.

Prediction: Plans and policies of organisations are invariably formulated well in advance of the time of their implementation. Knowledge of future trends is very helpful in framing suitable policies and plans. Statistical methods provide helpful means of forecasting future events.

Formulation of Suitable Policies: Statistics provide the basic material for framing suitable policies. Whether it be peace or war or any problem relating to economic, social or political issue, in the absence of adequate, accurate and timely data it would be difficult, rather impossible, to frame suitable policies.

Q: What are the limitations of Statistics? *(www.prepnext.com)*

Ans.:
The following are the important limitations of the science of statistics:

Statistics does not Deal with Individual Measurements: Since statistics deals with aggregates of facts, the study of individual measurements lies outside the scope of statistics. Data are statistical when they relate to measurement of masses, not statistical when they relate to an individual item or event as a separate entity.

Statistics Deals only with Quantitative Characteristics: Statistics are numerical statements of facts. Such characteristics as cannot be expressed in numbers are incapable of statistical analysis.

Statistical Results are True only on an Average: The conclusions obtained statistically are not universally true – they are true only under certain conditions. This is because statistics as a science is less exact as compared to natural sciences.

Statistics is only one of the Methods of Studying a Problem: Statistical tools do not provide the best solution under all circumstances. Hence statistical conclusions should be supplemented by other evidences.

Statistics can be Misused: The greatest limitation of statistics is that it is liable to be misused. The misuse of statistics may arise because of several reasons. For example, if statistical conclusions are based on incomplete information, one may arrive at fallacious conclusions.

Also, it requires experience and skill to draw sensible conclusions from the data; otherwise, there is every likelihood of wrong interpretation. The very fact that it may lead to fallacious conclusions in the hands of inexperienced people limits the possibility of mass popularity of such a useful science. Statistics cannot be used to full advantage in the absence of proper understanding of the subject to which it is applied.

Q: Define Descriptive Statistics? *(www.prepnext.com)*

Ans.:
Descriptive statistics deals with the numerical facts or data. For example- Averages, Dispersion, time series.

Q: What do you mean by Inferential Statistics?
(www.prepnext.com)

Ans.:
Inferential Statistics covers those methods and techniques which are used to draw conclusion and inferences about the parameters of population on the basis of estimates derived from a sample. For example- the technique used are chi-square test, F- test, E- test, etc.

Q: What do you mean by Primary data and Secondary data? Differentiate between them. *(www.prepnext.com)*

Ans.:
There are two types of data- Primary data and Secondary data.

1. **Primary data:-** It refers to the data which are originally collected for the first time by the investigator himself or his agent for a predetermined purpose or specific purpose.
2. **Secondary Data:-** The data which are collected from a published source is known as secondary data. Secondary data are compiled by someone other than the user. The sources from which secondary data are collected are known

as secondary sources. For example- the publication of data by CSO, NSS, RBI.

The differences between Primary data and Secondary data are:

PRIMARY DATA	SECONDARY DATA
1. Primary data are collection of original data for the first time.	**1.** Secondary data is basically compilation of existing data.
2. It is collected by the investigator or his agents.	**2.** It is compiled by persons other than who collected the primary data.
3. Its collection is relatively more costly.	**3.** Its collection is relatively less costly.
4. It is usually directly suitable to the purpose of enquiry.	**4.** It may or may not be directly suitable to the purpose of enquiry.
5. It may be used as it is for the purpose of enquiry.	**5.** It may require certain adjustments to be made to suit the purpose of enquiry.
6. There is possibility of personal prejudice in its collection.	**6.** There is no possibility of personal prejudice in its collection since such data are already collected.
7. For example- collection of population figure by a particular student in Tinsukia town.	**7.** The collection of population figure from the district office.

Q: Explain the different methods of collecting primary data.

(www.prepnext.com)

Ans.:

Primary Data *(qualitative or quantitative) collected by the researcher is unique to his research and, until he publishes, no one else has access to it.* The different methods of collecting primary data are:-

METHODS OF PRIMARY DATA COLLECTION:

1. **MAILED QUESTIONNAIRE METHOD:-** In this method, a questionnaire is sent (usually by post) to the respondents with a request to answer the questions and return the questionnaire. A questionnaire contains a number of questions in a definite order on a form or set of forms.

   MERITS:-

   a) It is economical in terms of time, money and labour.
   b) There is less scope for personal biasness.
   c) Covers a large number of people or organizations.
   d) Wide geographical coverage possible.

   LIMITATIONS:-

   a) It is not suitable where majority of respondents is illiterate.
   b) The information supplied may or may not be reliable.
   c) It involves uncertainty about the response from respondent.
   d) Response rate is low.
   e) Many reminders have to be given.
   f) There is a problem of incomplete questionnaire.
   g) There is a problem of designing questionnaire
   h) Respondents go through all the questions and then decide whether to complete it or not
   i) Responses are not spontaneous

2. **SCHEDULE SENT THROUGH ENUMERATOR:-** under this method the trained enumerators personally contact the informants, ask questions and record the replies to questions contained in the schedule.

   MERITS:-

   a) It can be used even where the informants are illiterates
   b) It can be used where the field of enquiry is wide.
   c) The chances of response are more
   d) The information received is more reliable and accurate.

LIMITATIONS:-

a) It is a costly method

b) It is time consuming

3. INTERVIEWS:

This method involves presentation of oral-verbal stimuli and reply in terms of oral-verbal response.

Major Types of Interview:

(i) Personal Interview

(ii) Telephone Interview

(iii) Focus Group Interviews

(i) Personal Interview: In this method, the interviewer asks questions in a face –to-face contact with the other person or persons. Personal Interview can be structured, semi-structured or unstructured.

Structured interviews *involve the use of predetermined questions and of highly standardized techniques of recording.*

In a ***semi-structures interview****, the interviewee is asked certain specific questions but gives the respondent scope to express himself at length.*

Unstructured or in-depth interview *do not follow a system of pre –determined questions. The interviewer has more flexibility. If needed he can ask supplementary questions or may omit certain questions if the situation so requires. Here the direction of the interview is decided by the respondent's initial responses.*

MERITS:-

a) It ensures high degree of accuracy

b) It helps to collect supplementary information which may be useful in the interpretation of the result.

c) It avoids the possibility of misinterpretation

d) It helps the investigator to adjust suitably depending upon the informant's reaction.

LIMITATIONS:-

a) It is a costly method where the number of persons to be interviewed is large and they are spread over a wide area.
b) It is a time consuming method.
c) There is a chance of biasness.

(ii) TELEPHONE INTERVIEW:

In this method of collecting information, the respondents are contacted over the telephone itself.

(iii) Focus Group Interviews:

The interviewer conducts the interview of a small group of respondents in a non-structured and natural manner. The main task of the interviewer is to confine the respondents to a discussion of specific issue of interest.

4. **Indirect oral Investigation:** under this method, investigator contact third parties who are suppose to possess the information on the problem under investigation. For example- this method is used to find out the cause of the fire, clues about the murders, etc. Under this method questions are put to different person and their answers are recorded. Enquiry commission usually adopts this method.

MERITS:-

a) It helps us to collect information about the persons who is reluctant to supply the required information.

LIMITATIONS:-

a) The information collected from some persons may not be reliable.

5. **Information through local correspondents:-** Under this method, the investigator appoints local agents or correspondents in different areas to collect and transmit information to the central office where data are processed and finally analysed. For Example- Newspaper agencies, TV channels adopt this method.

MERITS:-

a) It is economical.
b) Timely information is available.

LIMITATIONS:-

a) It may not ensure high degree of accuracy because of the personal prejudice and biasness of the correspondent.

6. **OBSERVATIONS:**

Under the observation method, the information is collected by the researcher's own direct observation without asking from the respondents. For example, the researcher may, instead of asking the brand of shoes used by the respondents, may himself look at the shoe.

There may be various forms of observations : Structured or unstructured, Disguised or undisguised, Natural or CONTRIVED, Personal or Mechanical, participant and non-participant.

*In **structured o**bservations, the researcher specifies in detail and in advance what will be observed and how the measurements will be recorded.*

*In **unstructured** observations, the researcher himself monitors all the relevant aspects of the phenomenon, without specifying in advance what will be observed and how the measurement will be recorded.*

*When the observer is observing in such a way that his presence is unknown to the people he is observing, such an observation is **disguised observation**.*

*In an **undisguised** observation, respondents are aware that they are being observed.*

*In a **natural** observation, behaviour of the respondents is observed as it takes place in the natural environment.*

*In a **contrived** observation, the behaviour of the respondents is observed in an artificial environment, according to a definite pre-arranged plan.*

*In a **personal** observation, a researcher himself observes and records actual behaviour as it takes place.*

*In a **mechanical** observation, mechanical devices such as closed circuit television, audio-video aids, are used to record what is being observed.*

*In a **participant observation**, the observer observes by making himself a member of the group he is observing in order to experience what the members of the group are experiencing.*

*In a **non-participant observation**, the observer observes, without making any effort to experience through participation what the members of the group are experiencing.*

Q: What are the sources of secondary data? *(www.prepnext.com)*

Ans.:

Secondary data means data that are already available. They have already been collected by someone else. Secondary data may be published data or unpublished data. *This is a quick method of data collection, but the information may not be suitable for the particular purpose.*

Usually **published data** are available in (a) various publications of the central, state and local governments; (b) various publications of foreign governments or of international bodies and their subsidiary organizations; (c) technical and trade journals; (d) books, magazines and newspapers; (e) reports and publications of various associations connected with business and industry, banks, stock exchanges, etc. (f) reports prepared by research scholars, universities, economists, etc. in different fields; (g) public records and statistics, historical documents, and other sources of published information.

The sources of **unpublished data** are many: they may be found in diaries, letters, unpublished biographies and autobiographies and also may be available with scholars and research workers, trade associations, labour bureaus and other public/ private individuals and organizations.

Q: What do you mean by average? Mention the objectives of average.

OR

What do you mean by measure of central tendency? Mention its objectives. *(www.prepnext.com)*

Ans.:

An average is a single value which is used to represent all other values in the series. This single value describes the characteristic of the entire mass of unwieldy data. Such a value is called the central value or an 'average' or the expected value of the variable. Since the average lies somewhere in between the two extremes, that is the largest and the smallest items, it is also called the measure of central tendency.

The word average has been defined differently by various authors. Some important definitions are given below:

Clark – "Average is an attempt to find one single figure to describe whole of figures".

Ya-Lun-Chou –"An average is a typical value in the sense that it is sometimes employed to represent all the individual values in a series or of a variable."

Leabo – "The average is sometimes described as a number which is typical of the whole group".

Croxton & Cowden – "An average value is a single value within the range of the data that is used to represent all of the values in the series. Since an average is somewhere within the range of the data, it is also called a measure of central value".

Objectives of Average:

The main objectives of average are-

1. **To get single value that describes the characteristic of the entire group/ series**: Measures of central value, by condensing the mass of data in one single value, enable us to get a bird's eye view of the entire data. Thus one value can represent a large number of values. For example, when we obtain average income by dividing the total national income by total population we get one single value that represents the entire population.
2. **To facilitate comparison at a particular point of time or over a period of time:** Measures of central value, by reducing the mass of data to one single figure, enable comparison to be made. Comparison can be made either at a point of time or over a period of time. For example, we can compare the percentage results of the students of different colleges in a certain examination, say. B.Com for 2014.
3. **To facilitate statistical inference**- an average obtained from a sample is used in estimating the average of the population.
4. **To help the decision making process**.

Q: Mention the requisites of an ideal average.

OR

State the features/ properties of a good method of average.

(www.prepnext.com)

Ans.:

An ideal average should possess the following characteristics:-

(i) **Easy to Understand**: Since statistical methods are designed to simplify complexity, it is desirable that an average be such that can be readily understood.

(ii) **Simple to Calculate**: An average should be simple to compute so that it can be used widely. However, ease of computation should not be sought at the expense of accuracy.

(iii) **Based on all the Items** of the series : The average should depend upon each and every item of the series so that if any of the items is dropped the average itself is altered. For example, the arithmetic mean of 10, 20, 30, 40, 50, is 30. If we drop one item, say, 50, the arithmetic mean would be 25.

(iv) **Not be unduly Affected by Extreme Observations**: Although each and every item should influence the value of the average, none of the items should influence it unduly. If one or two very small or very large items unduly affect the average, i.e., either increase its value or reduce its value, the average cannot be really typical of the entire series.

(v) **Rigidly Defined**: An average should be properly defined so that it has one and only one interpretation. It should preferable be defined by algebraic formula so that if different people compute the average from the same figures they all get the same answer. The average should not depend upon the personal prejudice and bias of the investigator.

(vi) **Capable of Further Algebraic Treatment**: We should prefer to have an average that could be used for further statistical computations so that its utility is enhanced. For example, if we are given the data about the average income and number

of employees of two or more factories, we should be able to compute the combined average.

(vii) Sampling Stability : We should prefer to get a value which has 'sampling stability'. It means if one takes different samples from the same populations, the average of any sample should approximately turn out to be the same as those of other samples.

Q: What are the different types of averages?

(www.prepnext.com)

Ans.:

The following are the important types of averages:

(i) Arithmetic mean : (a) simple, and (b) weighted
(ii) Median
(iii) Mode
(iv) Geometric Mean
(v) Harmonic Mean

Besides these, there are less important averages like moving average, progressive average, etc. These averages have a very limited field of application.

Q: Define arithmetic mean. Mention its mathematical properties.

(www.prepnext.com)

Ans.:

Arithmetic mean is the most popular and widely used measure of representing the entire data by one value. Arithmetic mean is obtained by dividing the sum of the values of all items of a series by the number of items of that series. Normally, arithmetic mean is denoted by $\bar{x}$. It is popularly known as mean. Arithmetic mean may either be

(i) Simple arithmetic mean, or
(ii) Weighted arithmetic mean.

Mathematical Properties:-

1. The sum of the deviations of the items from the arithmetic mean (taking signs into account) is always zero, that is $\sum(x - \bar{x}) = 0$. *This is because the sum of the positive deviations from actual mean is equal to the sum of the negative deviations from it.*
2. For a given set of observations the sum of the squares of deviations is the minimum, when deviations are taken from the arithmetic mean.
3. If each of the values of a variable x is increased or decreased by some constant (c) the arithmetic mean also increases or decreases by c.
4. It is capable for further algebraic treatment.
5. Since $\bar{x} = \sum x / N$, therefore, $N\bar{x} = \sum x$

   In other words, if we replace each item in the series by the mean, then the sum of these substitutions will be equal to the sum of the individual items.
6. If we have the arithmetic mean and number of items of two or more than two related groups, we can compute combined average of these groups by applying the following formula:

$$\bar{x}_{12} = (N_1\bar{x}_1 + N_2\bar{x}_2)/ (N_1 + N_2)$$

$\bar{x}_{12}$ = combined mean of the two groups;

$\bar{x}_1$ = arithmetic mean of first group

$\bar{x}_2$ = arithmetic mean of second group;

N_1 = number of items in the first group

N_2 = number of items in the second group

Q: Mention the merits and demerits of arithmetic mean?

(www.prepnext.com)

Ans.:

Following are the **MERITS** of arithmetic mean-

1. Arithmetic mean is easy to understand
2. It is simple to calculate
3. It is based on all the items of the series
4. It is defined by a rigid mathematical formula with the result that everyone who computes the average gets the same answer.
5. It is capable of further algebraic treatment
6. It has sampling stability
7. It is relatively reliable in the sense that it does not vary too much when repeated samples are taken from one and the same population, at least not as much as some other kind of statistical descriptions.
8. The mean is typical in the sense that it is the centre of gravity, balancing the values on either side of it.
9. It is a calculated value, and not based on position in the series.

DEMERITS of Arithmetic mean-

1. It is affected by extreme values. *Since the value of mean depends upon each and every item of the series, extreme items, i.e., very small and very large items, unduly affect the value of the average. For example, if in a tutorial group, there are 4 students and their marks in a test are 60, 70, 10 and 80 the average marks would be (60 + 70 + 10 + 80)/ 4 = 55. One single value, i.e., 10, has reduced the average marks considerably*.
2. It is not useful for studying the qualitative phenomenon
3. For the open-ended distribution mean can not be calculated accurately. : *In a distribution with open-end classes the value of mean cannot be computed without making assumptions regarding the size of the class interval of the open-end classes. If such classes contain a large proportion of the values, then mean may be subject to substantial error.*

4. The arithmetic mean is not always a good measure of central tendency. The mean provides a "characteristic" value, in the sense of indicating where most of the values lie, only when the distribution of the variable is reasonably normal (bell-shaped). In case of a U-shaped distribution the mean is not likely to serve a useful purpose.

Q: What do you mean by Median? Why median is known as positional average? *(www.prepnext.com)*

Ans.:

Median is the central value of the variable that divides the series into two equal parts in such a way that half of the items lies above this value and the remaining half lies below this value. In other words, Median refers to the middle value in a distribution. The median is just the 50^{th} percentile value below which 50 per cent of the values in the sample fall, it splits the observation into two halves.

Median is called a positional average because it is based on the position (place) of a given observation in a series arranged in ascending or descending order and the position of the median is such that an equal number of items lie on either side of it. For example, if the income of five employees is Rs. 5000, 5400, 5650, 5700 and 5900 the median would be the middlemost value - 5650.

Q: Mention mathematical properties of Median:

(www.prepnext.com)

Ans.:

The sum of the deviations of the items from median, ignoring signs, is the least. This means that this total is smaller than the one obtained if deviations are taken from any other value. For example, the median of 4, 6, 8, 10, 12 is 8. The deviations from 8 ignoring signs are 4, 2, 0, 2, 4 and the total is 12. Thus if deviations are taken from 7, values ignoring signs would be 3, 1, 1, 3, 5, and the total is 13.

Q: What are the Merits and Limitations of Median?

(www.prepnext.com)

Ans.:

MERITS:

1. It is especially useful in case of open-end classes since only the position and not the values of items must be known. The median is also recommended if the distribution has unequal classes, since it is easier to compute than the mean.
2. Extreme values do not affect the median as strongly as they do the mean. For example, the median of 10, 20, 30, 40 and 150 would be 30 whereas the mean 50. Hence very often when extreme values are present in a set of observations, the median is a more satisfactory measure of the central tendency than the mean.
3. In markedly skewed distributions such as income distributions or price distributions where the arithmetic mean would be distorted by extreme values, the median is especially useful. Consequently, the median income for some purposes be regarded as a more representative figure, for half the income earners must be receiving at least the median income. One can say as many receive the median income and as many do not.
4. It is the most appropriate average in dealing with qualitative data, i.e., where ranks are given or there are other types of items that are not counted or measured but are scored.
5. The value of median can be determined graphically whereas the value of mean cannot be graphically ascertained.
6. Perhaps the greatest advantage of median is, however, the fact that the median actually does indicate what many people incorrectly believe the arithmetic mean indicates. The median indicates the value of the middle item in the distribution. This is a clear-cut meaning and makes the median a measure that can be easily explained.

LIMITATIONS:

1. For calculating median it is necessary to arrange the data; other averages do not need any arrangement.
2. Since it is a positional average, its value is not determined by each and every observation.
3. It is not capable of algebraic treatment. For example, median cannot be used for determining the combined median of two or more group as is possible in case of mean.
4. The value of median is affected more by sampling fluctuations than the value of the arithmetic mean.
5. The median, in some cases, cannot be computed exactly as the mean. When the number of items included in a series of data is even, the median is determined approximately as the mid-point of the two middle terms.
6. It is erratic if the number of items is small.

RELATED POSITIONAL MEASURES: *Besides median, there are other measures which divide a series into equal parts. Important amongst these are quartiles, deciles and percentiles. Quartiles are those values of the variate which divide the total frequency into four equal parts, deciles divide the total frequency into 10 equal parts and the percentiles divide the total frequency into 100 equal parts.*

Computation of Quartiles, Percentiles, etc.

Q_1 = Size of $(N + 1)/ 4^{th}$ item (individual observations and discrete series).

Q_1 = Size of $N/4^{th}$ item (in continuous series).

Q_3 = Size of $3(N + 1)/ 4^{th}$ item (in individual and discrete series).

Q_3 = Size of $3N/4^{th}$ item (in continuous series).

D_4 = Size of $4(N + 1)/ 10^{th}$ item (in individual and discrete series).

D_4 = Size of $4N/10^{th}$ item (in continuous series).

P_{60} = Size of $60(N + 1)/ 100^{th}$ item (in individual and discrete series).

P_{60} = Size of $60N/100^{th}$ item (in continuous series).

Q: Define Mode? *(www.prepnext.com)*

Ans.: The mode or the modal value is that value in a series of observation which occurs with greatest frequency. For example- the mode of the series 3, 5, 8, 5, 4, 5, 9, 3, is 5, since this value occurs more frequently than any of the others.
Croxton and Cowden, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values."

Mode is the value around which the items tend to be most heavily concentrated. A set of data may have a single mode in which case it is said to be unimodal, it may have two modes which makes it bimodal or it may have several modes and be called multimodal.

Q: What are the Merits and Limitations of Mode?
(www.prepnext.com)

Ans.:
MERIT:

1. By definition mode is the most typical or representative value of a distribution. If the modal wage in a factory is Rs. 916 then more workers receive Rs. 916 than any other wage.
2. Like median, the mode is not unduly affected by extreme values. Even if the high values are very high and the low values are very low we choose the most frequent value of the data to the modal value.
3. Its value can be determined in open-end distributions without ascertaining the class limits.
4. The value of mode can also be determined graphically whereas the value of mean cannot be graphically ascertained.

LIMITATIONS:

1. The value of mode cannot always be determined. In some case we may have a bimodal series.
2. It is not capable of algebraic manipulations. For example, from the modes of two sets of data we cannot calculate the overall mode of the combined data. The value of mode is not based on each and every item of the series.
3. It is not a rigidly defined measure. There are several formulae for calculating the mode, all of which usually give somewhat different answers. In fact, mode is the most unstable average and its value is difficult to determine.
4. While dealing with quantitative data, the disadvantages of the mode outweigh its good features and hence it is seldom used.

Q: Mention the relationship between mean, median and mode?

(www.prepnext.com)

Ans.:

A distribution in which the value of mean, median and mode coincide (i.e., mean = median = mode) is known as symmetrical distribution. In moderately skewed or asymmetrical distributions a very important relationship exists among mean, median and mode. Karl Pearson has expressed this relationship as follows:

Mode= 3 median – 2 mean

OR

Mode = Mean – 3 (Mean – Median)

And, Median = Mode + 2/3 (Mean – Mode)

Q: Define Harmonic Mean? *(www.prepnext.com)*

Ans.: The Harmonic mean for *'n'* observations x_1, x_2, ….. x_n is the total number divided by the sum of the reciprocals of the numbers. In other words, harmonic mean of a series is the reciprocal of the arithmetic mean of the reciprocal of the individual observations.

Harmonic Mean = N/ $(1/x_1 + 1/x_2 + 1/x_3 + \ldots\ldots 1/x_n)$

The harmonic mean is a measure of central tendency for data expressed as rates such as kms. per hour, kms. per litre, etc.

Q: Mention the merits and demerits of harmonic mean?

(www.prepnext.com)

Ans.:

Merits:-

1. Its value is based on every item of the series.
2. It is capable of further algebraic treatment.
3. It is rigidly defined.
4. In problems relating to time and rates it gives better results than other averages.

Demerits:-

1. It is difficult to understand.
2. It is difficult to calculate.
3. It cannot be calculated when one or more items are zero.
4. It gives largest weight to smallest items.
5. Its value cannot be computed when there are both positive and negative items in a series.

Q: What do you mean by geometric mean? Mention its properties?

(www.prepnext.com)

Ans.:

Geometric mean is defined as the N^{th} root of the product of 'N' items or values. Symbolically, geometric mean =

$$\sqrt[n]{(x_1 \times x_2 \times x_3 \times \ldots\ldots x_n)}$$

Where, x_1, x_2, x_3, etc., refer to the various items of the series.

Thus the geometric mean of 3 values 2, 3, 4 would be:

G.M. = $\sqrt[3]{(2 \times 3 \times 4)} = \sqrt[3]{24} = 2.885$

Mathematical Properties:-

1. The product of the values of series will remain unchanged when the value of the geometric mean is substituted for each individual value. For example, the geometric mean for series 2, 4, 8 is 4; therefore, we have
   $2 \times 4 \times 8 = 64 = 4 \times 4 \times 4$
2. The sum of the deviations of the logarithms of the original observations above or below the logarithms of the geometric mean is equal.

Q: Why geometric mean is considered as the best average? (Uses of geometric mean) *(www.prepnext.com)*

Ans.: Uses of Geometric Mean:

1. The geometric mean is used to find the average per cent increase in sales, production, population or other economic or business series. *For example, from 2012 to 2014 prices increased by 5, 10 and 18 per cent respectively. The average annual increase is not 11 per cent [(5 + 10 + 18)/ 3 = 11] as given by the arithmetic average but 10.9 per cent as obtained by the geometric mean.*
2. Geometric mean is theoretically considered to be the best average in the construction of index numbers. It satisfies the time reversal test and gives equal weight to equal ratio of change.
3. This average is most suitable when large weights have to be given to small items and small weights to large items, situations which we usually come across in social and economic fields.

Q: What are the Merits and Limitations of Geometric Mean? *(www.prepnext.com)*

Ans.:

MERITS:-

1. It is based on each and every item of the series.
2. It is rigidly defined
3. It is capable for further algebraic treatment. For example, if the geometric average of two or more series and their number of items are known, a combined G.M. can be easily calculated.

LIMITATIONS:-

1. It is difficult to understand.
2. It is difficult to compute and to interpret and so has restricted application.
3. It cannot be computed when there are both negative and positive values in a series
4. It cannot be computed when one or more of the values are zero.

Q: Mention the relationship among A.M., G.M. and H.M.

(www.prepnext.com)

Ans.:

$$AM \geq GM \geq HM$$

In any distribution when the original items differ in size the value of A.M., G.M. and H.M. would also differ and will be in the following order:

$$AM \geq GM \geq HM$$

i.e. arithmetic mean is greater than geometric mean and geometric mean is greater than harmonic mean. *The equality signs hold only if all the numbers X_1, X_2 ……. X_n are identical.*

Q: What are the general limitations of 'Average'?

(www.prepnext.com)

Ans.:

1. *Since an average is a single value representing a group of values it must be properly interpreted; otherwise there is every possibility of jumping to wrong conclusions.*

2. *An average may give us a value that does not exist in the data. For example, the arithmetic mean of 100, 300, 250, 50, 100 is = 160 a value that does not exist in the data.*
3. *At times the average may give a very absurd result. For example, if we are calculating size of a family we may get a value 4.8. But this is impossible as persons cannot be in fractions. However we should remember that it is an average value representing the entire group.*
4. *Measures of central value fail to give an idea about the formation of the series. Two or more series may have the same central value but may differ widely in composition.*
5. *An average is a measure of central tendency. Hence unless the data show a clear single concentration of observations an average may not be meaningful at all. This evidently precludes the use of any average to typify a bimodal, a U-shaped or a J-shaped distribution.*

WHICH AVERAGE TO USE:

No one average can be regarded as best for all circumstances. The following considerations influence the selection of an appropriate average:

1. *The purpose which the average is designed to serve.*
2. *Would the average be used for further computations?*
3. *The type of data available. Are they badly skewed (avoid the mean), gappy around the middle (avoid the median) or unequal in class interval (avoid the Mode)?*
4. *The typical value required in the particular problem:- The main requirement is to know what each average means and then select one that fulfils the purpose in hand. Is a composite average of all absolute or relative values needed (arithmetic mean or geometric mean) or is middle value wanted (median) or the most common value (mode)?*

On occasions it may even be advisable to work out more than one average and present them.

Median: *The median is generally the best average in open-end grouped distributions.*

Mode: *The mode can be used in problems involving the expression of preferences where quantitative measurements are not possible. If we want to compare consumer preferences for different kinds of products or different kinds of advertising we can compare the modal preferences expressed by different groups of people but we cannot calculate the median or mean. The mode is a particularly useful average (IS BEST SUITED) where there is an outstandingly large frequency.*

Geometric Mean: *Geometric mean is useful in averaging ratio and percentages, and in computing average rates of increase or decrease. It is particularly important in Economics and Business Statistics in index number construction.*

Harmonic Mean: *Harmonic mean is useful in problems in which values of a variable are compared with a constant quantity of another variable e.g., distance covered within certain time.*

Leaving aside the above specific cases where either median, mode, geometric mean or harmonic mean is more appropriate in other cases we should apply as a rule of thumb the arithmetic mean – the most popular and widely used average in practice.

In the following cases arithmetic mean should not be used:

(i) In highly skewed distributions.

(ii) In distributions with open-end intervals.

(iii) When the distribution is unevenly spread concentration being smaller or large at irregular points.

(iv) The arithmetic mean should not be used to average ratios and rates of change. In such cases the geometric mean is more suitable.

(v) When there are very large and very small items arithmetic mean would be seriously misleading on account of undue influence from extreme items.

Q: What do you mean by dispersion? Mention the significance of it?
(www.prepnext.com)

Ans.:
Dispersion measures the extent to which the items vary from some central value. It is also known as variation.

According to A.L. Bowley, "Dispersion is the measure of the variation of the items."

According to Spiegel, "The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data."

According to Brooks & Dick, "Dispersion or spread is the degree of the scatter or variation of the variable about a central value."

It is clear from above that dispersion (also known as scatter, spread or variation) measures the extent to which the items vary from some central value. The measures of dispersion are also called **averages of the second order** because these measures give an average of the differences of various items from an average.

Significance of measures of dispersion:-

(Purposes of Dispersion)

(Why do you need measures of dispersion?)

Measures of variations are needed for the following purposes:-

1. **To determine the reliability of an average**: It points out as to how far an average is representative of the entire data.
2. To serve as a basis for the control of the variability**:** Measures of dispersion determine the nature and cause of variation in order to control the variation itself.
3. To compare two or more series with regard to their variability**:** It helps to compare two or more distributions with regard to their variability. A high degree of variation would mean little uniformity or consistency whereas a low degree of variation would mean great uniformity or consistency.

4. **To facilitate the use of other statistical measures:** Many powerful analytical tools in statistics such as correlation analysis, the testing of hypothesis, analysis of variance, the statistical quality control, regression analysis are based on measures of variation of one kind or another.

Q: Mention properties of a good measure of dispersion or variation?

(www.prepnext.com)

Ans.:

A good measure of dispersion should possess as far as possible the following properties:-

- It should be simple to understand
- It should be easy to compute
- It should be rigidly defined
- It should be based on each and every item of the distribution
- It should be capable of further algebraic treatment
- It should have sampling stability
- It should not be unduly affected by extreme items

Q: Mention the methods of studying variation. (Dispersion)

(www.prepnext.com)

Ans.:

The following are the important methods of studying variation:

- The range
- The interquartile range or the quartile deviation
- The mean deviation or average deviation
- The standard deviation
- The Lorenz curve

Of these the first two, namely, the range and quartile deviations, are positional measures because they depend on the values at a particular position in the distribution. The other two, the average deviation and the standard deviation, are called calculation measures of deviation because all of the values are employed in their calculation and the last one is a graphic method.

Absolute and Relative Measures of Variation:

Measures of dispersion may be either absolute or relative. Absolute measures of dispersion are expressed in the same statistical unit in which the original data are given such as rupees, kilograms, tonnes, etc. These values may be used to compare the variations in two distributions provided the variables are expressed in the same units and of the same average size. In case the two sets of data are expressed in different units, however, such as quintals of sugar versus tonnes of sugarcane, the absolute measures of dispersion are not comparable. In such cases measures of relative dispersion should be used. A measure of relative dispersion is the ratio of a measure of absolute dispersion to an appropriate average.

Q: What do you mean by Range? *(www.prepnext.com)*

Ans.:

Range is defined as the difference between the largest value and the smallest value in the distribution.

Range= L-S

Where, L= largest item

S=smallest item

Q: What do you mean by co-efficient of range?

(www.prepnext.com)

Ans.: The relative measure corresponding to range is known as co-efficient of range.

Co-efficient of range= (L-S)/ (L+S)

Q: What are the Merits and Limitations of Range?

(www.prepnext.com)

Ans.:

Merits:

1. Amongst all the methods of studying dispersion range is the simplest to understand and the easiest to compute.
2. It takes minimum time to calculate the value of range. Hence, if one is interested in getting a quick rather than a very accurate picture of variability one may compute range.

Limitations:

1. Range is not based on each and every item of the distribution.
2. It is subject to fluctuations of considerable magnitude from sample to sample.
3. Range cannot tell us anything about the character of the distribution within the two extreme observations. The range takes no account of the form of the distribution within the range. Range is, therefore, unreliable.
4. Range cannot be computed in case of open-end distributions.

Q: What are the uses/ utility of range? *(www.prepnext.com)*

Ans.:

USES:

Despite serious limitations range is useful in the following cases:

(i) Quality Control: *The object of quality control in a manufacturing unit is to keep a check on the quality of the product without 100% inspection. In this range plays a very important role. If the range – the difference between the largest and smallest mass produced items – increases beyond a certain point, the production machinery should be examined to find out why the items produced have not followed their usual more consistent pattern.*

(ii) ***Fluctuations in the Share Prices:*** *Range is useful in studying the variations in the prices of stocks and shares and other commodities that are sensitive to price changes from one period to another. For example, by computing range we can get an idea about the range of variation of, say, gold prices. If the minimum price for 10 gm. gold in the year 2013-14 was Rs. 27,500 and the maximum price Rs. 32,000 this at once tells us about the range of variation, i.e., (32,000 – 27,500) is Rs. 4,500.*

(iii) ***Weather Forecasts.*** *The meteorological department does make use of the range in determining, say, the difference between the minimum temperature and the maximum temperature. This information is of great concern to the general public because they know as to within what limits the temperature is likely to vary on a particular day.*

(iv) ***Everyday Life:*** *The range is a most commonly used measure of dispersion in everyday life. Questions of the form "what is the minimum and maximum temperature on a particular day"? "What is the difference between the wages earned by workers of a particular factory"? - are all usually answered in the form of range. Answers to questions such as these are usually given in the form of "Between such and such". The answer is usually in a range.*

Q: Define inter quartile range. *(www.prepnext.com)*

Ans.:

Interquartile range includes the middle 50 per cent of the distribution. That is, one quarter of the observations at the lower end, another quarter of the observations at the upper end of the distribution are excluded in computing the interquartile range. In other words, interquartile range represents the difference between the third quartile and the first quartile.

Interquartile range = $Q_3 - Q_1$

Q: What is Quartile Deviation or Semi- interquartile range?

(www.prepnext.com)

Ans.:

Quartile deviation is ½ the difference between upper quartile (Q_3) and lower quartile (Q_1). Quartile deviation indicates the average amount by which the two quartiles differ from the median. The interquartile range is reduced to the form of the Semi- interquartile range or quartile deviation by dividing it by 2.

Symbolically,

$$\text{Quartile Deviation or Q.D.} = (Q_3 - Q_1)/ 2$$

Q: What is Coefficient of quartile deviation? *(www.prepnext.com)*

Ans.:

Quartile deviation is an absolute measure of dispersion. The relative measure corresponding to this measure, called the coefficient of quartile deviation, is calculated as follows.

$$\text{Coefficient of Q.D.} = [(Q_3 - Q_1)/2] \;/\; [(Q_3 + Q_1) / 2]$$
$$= (Q_3 - Q_1)/ (Q_3 + Q_1)$$

Q: What are the Merits and Limitations of quartile deviation?

(www.prepnext.com)

Ans.:

<u>Merits</u>:

In certain respects it is superior to range as a measure of dispersion.

1. It has a special utility in measuring variation in case of open end distributions.
2. It is also useful in erratic or badly skewed distributions, where the other measures of dispersion would be warped by extreme values. The quartile deviation is not affected by the presence of extreme values.

Limitations:

1. Quartile deviation ignores 50% items, i.e., the first 25% and the last 25%. As the value of quartile deviation does not depend upon every item of the series, it cannot be regarded as a good method of measuring dispersion.
2. It is not capable of mathematical manipulation.
3. Its value is very much affected by sampling fluctuations.
4. It is in fact not a measure of dispersion as it really does not show the scatter around an average but rather a distance on a scale, i.e., quartile deviation is not itself measured from an average, but it is a positional average.

Because of the above limitations quartile deviation is not often useful for statistical inference.

PERCENTILE RANGE*: Like semi- interquartile range, the percentile range is also used as a measure of dispersion. Percentile range of a set of data is defined as:*

Percentile Range = $P_{90} - P_{10}$

Where P_{90} and P_{10} are the 90^{th} and 10^{th} percentiles respectively. The semi-percentile range, i.e., $[(p_{90} - p_{10})/ 2]$ can also be used, but is not commonly employed.

Q: What is mean deviation or average deviation? What are its merits and limitations? *(www.prepnext.com)*

Ans.:

Mean deviation is the average difference between the items in a distribution and the median or mean of that series. It is also known as average deviation.

In practice arithmetic mean is used for mean deviation and median is used rarely.

If $X_1, X_2, X_3, \ldots\ldots X_n$ are N given observations then the deviation about an average **A** is given by

$$\text{Mean Deviation (M.D.)} = \sum|D|/\ N$$

Where $|D| = |X - A|$; $|X - A|$ is the modulus value or absolute value of the deviation ignoring plus and minus signs.

(**N.B.** When deviations are taken from median, the sum of deviations of items from median is minimum **when signs are ignored**)

Merits:

1. Average deviation is simple to understand and easy to compute.
2. It is based on each and every item of the data. Consequently change in the value of any item would change the value of mean deviation.
3. Mean deviation is less affected by the value of extreme items than the standard deviation.
4. Since deviations are taken from a central value, comparison about formation of different distributions can easily be made.

Limitations:

1. The greatest drawback of this method is that algebraic signs are ignored while taking the deviations of the items. *For example, if from twenty, fifty is deducted we write 30 and not – 30.* This is mathematically wrong and makes the method **non-algebraic.**
2. This method may not give us very accurate results. The reason is that mean deviation gives us best results when deviations are taken from median. But median is not a satisfactory measure when the degree of variability in series is very high.
3. It is not capable of further algebraic treatment.
4. It is rarely used in sociological studies.

Because of these limitations its use is limited and it is overshadowed as a measure of variation by the superior standard deviation.

Q: Explain usefulness of average deviation. *(www.prepnext.com)*

Ans.:
Average deviation is simple to understand and compute. It is especially effective in reports presented to the general public or to groups not familiar with statistical methods. This measure is useful for small samples with no elaborate analysis required. *Incidentally, it may be mentioned that the National Bureau of Economic Research has found, in its work on forecasting business cycle, that the average deviation is the most practical measure of dispersion to use for this purpose*

Q: What is coefficient of mean deviation? *(www.prepnext.com)*

Ans.:
The relative measure corresponding to the mean deviation, called the coefficient of mean deviation, is obtained by dividing mean deviation by the particular average used in computing mean deviation. Thus, if mean deviation has been computed from median, the coefficient of mean deviation shall be obtained by dividing mean deviation by median.

Coefficient of M.D. = M.D./ Median.

If mean has been used while calculating the value of mean deviation, in such as case coefficient of mean deviation shall be obtained by dividing mean deviation by the mean.

Q: What do you mean by standard deviation? Mention one of its properties. *(www.prepnext.com)*

Ans.: Standard deviation is the square root of the arithmetic mean of the squares of deviations of all items of the distribution from arithmetic mean. The standard deviation was introduced by Karl Pearson in 1823. It is the most important and widely used measure of studying dispersion.

*(Standard Deviation is also known as **root mean square deviation** for the reason that it is the square root of the mean of the squared deviation from the arithmetic mean. Standard deviation is denoted by the small Greek letter ∂ (read as sigma).*

Q: Differentiate between mean deviation and standard deviation. *(www.prepnext.com)*

Ans.:

Mean Deviation	Standard Deviation
Actual ± signs are ignored and all deviations are taken as positive.	Actual ± signs are not ignored
Mean deviation can be calculated from arithmetic mean, median or mode.	It is calculated from arithmetic mean

---

Q: What are the Mathematical Properties of Standard Deviation? What are its merits and limitations? *(www.prepnext.com)*

Ans.:

1. **Combines Standard Deviation:** Just as it is possible to compute combined mean of two or more than two groups, similarly we can also compute combined standard deviation of two or more groups.
2. The sum of the squares of the deviations of items in the series from their arithmetic mean is minimum. This is the reason why standard deviation is always computed from the arithmetic mean.
3. **Standard Deviation of first n Natural Numbers**: The standard deviation of the first n natural numbers can be obtained by the following formula:

$$\partial = \sqrt{[1/12\ (N^2 - 1]}$$

Thus the standard deviation of natural numbers 1 to 10 will be

$$\partial = \sqrt{[1/12\ (10^2 - 1]} = \sqrt{[1/12 \times 99]} = \sqrt{8.25} = 2.87$$

Merits:

1. The standard deviation is the best measure of variation because of its mathematical characteristics. It is based on every item of the distribution. Also it is amenable to algebraic treatment and is less affected by fluctuations of sampling than most other measures of dispersion.
2. It is possible to calculate the combined standard deviation of two or more groups. This is not possible with any other measure.
3. For comparing the variability of two or more distributions coefficient of variation is considered to be most appropriate and this is based on mean and standard deviation.
4. Standard deviation is most prominently used in further statistical work. For example, in computing skewness, correlation, etc., use is made of standard deviation. It is keynote in sampling and provides a unit of measurement for the normal distribution.

Limitations:

1. As compared to other measures it is difficult to compute.
2. It gives more weight to extreme values and less to those which are near the mean. It is because of the fact that the squares of the deviations which are big in size would be proportionately greater than the squares of those deviations which are comparatively small. The deviations 2 and 8 are in the ratio of 1:4 but their squares, i.e., 4 and 64, would be in the ratio of 1 : 16.

Q: What is Coefficient of Variation? *(www.prepnext.com)*

Ans.:

The standard deviation is an absolute measure of dispersion. The corresponding relative measure is known as the coefficient of variation. This measure developed by Karl Pearson is the most commonly used measure of relative variation. It is used in such

problems where we want to compare the variability of two or more than two series. That series (or group) for which the coefficient of variation is greater is said to be more variable or conversely less consistent, less uniform, less stable or less homogeneous. On the other hand, the series for which coefficient of variation is less is said to be less variable or more consistent, more uniform, more stable or more homogeneous. Coefficient of variation is denoted by C.V. and is obtained as follows:

$$\text{Coefficient of Variation or C.V.} = (\partial / \bar{x}) * 100$$

Q: What is Variance? *(www.prepnext.com)*

Ans.:
The term variance was used to describe the square of the standard deviation by R.A. Fisher in 1913. Variance is defined as follows:

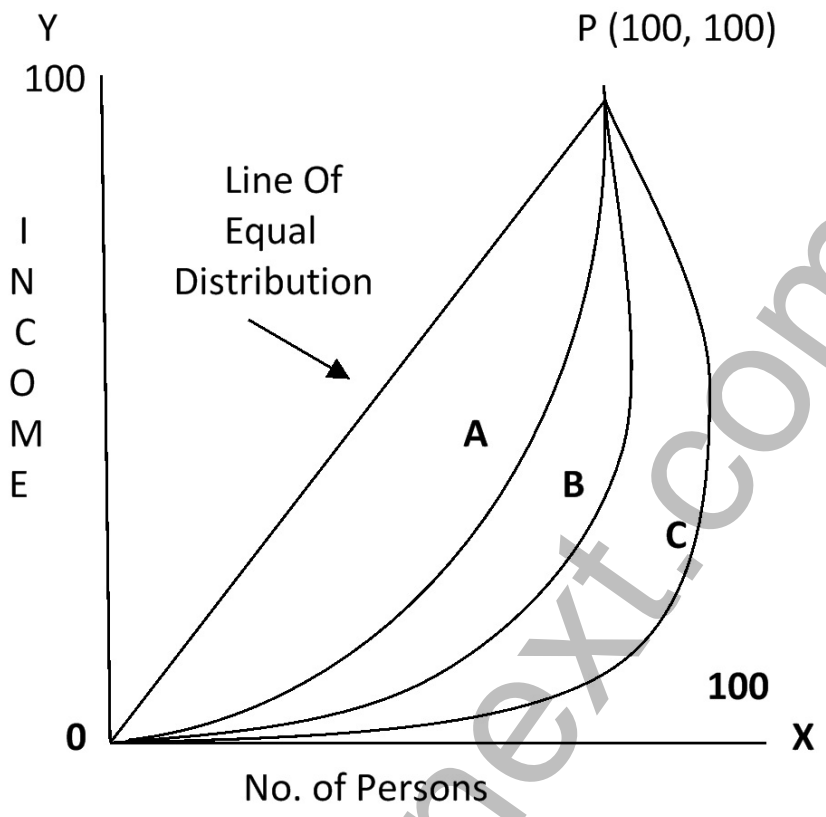
$$\text{Variance} = \sum[(x - \bar{x})^2 / N]$$

Thus, variance is nothing but the square of the standard deviation.

$$\text{i.e., Variance} = \partial^2$$
$$\text{or, } \partial = \sqrt{\text{Variance}}$$

Q: What do you mean by Lorenz Curve? *(www.prepnext.com)*

Ans.:
The Lorenz curve, devised by Max O. Lorenz, is a graphical method of studying dispersion. Basically, Lorenz curve is a cumulative percentage curve in which the percentage of item is combined with the percentage of other things as wealth, income, sales, profit etc. The most common use of this curve is in the study of the degree of inequality in the distribution of income and wealth between countries or between different periods of time. The curve is also used to study the distribution of profits, wages, turnover, etc.



While drawing the Lorenz curve the following procedure is adopted:

(i) The size of items (variable values) and frequencies are both cumulated. Taking grand total for each as 100, percentages are obtained for these various cumulative values.

(ii) On the X-axis start from 0 to 100 and take the per cent of cumulative frequencies.

(iii) On the Y-axis start from 0 to 100 and take the per cent of the cumulated values of the variable.

(iv) Draw a diagonal line joining O (0,0) with the point P (100, 100) as shown in the diagram above. The line OP will make an angle of 45° with the Y-axis and is called the line of equal distribution. Any point on this diagonal shows that same per cent on X as on Y.

(v) Plot the percentages of the cumulated values of the variable (Y) against the percentages of the corresponding cumulated frequencies (X) for the given distribution and join these points with a smooth freehand curve. For any given distribution this

will never cross the line of equal distribution OP. It will always lie below OP unless the distribution is uniform in which case it will coincide with OP. The greater the variability, the greater is the distance of the curve from OP.

In the above diagram OP is the line of equal distribution. The points lying on the curve OAP indicate a less degree of variability as compared to the points lying on the curve OBP. When the points lie on the curve OCP, variability is still greater. Thus a measure of variability of the distribution is provided by the distance of the curve of the cumulated percentages of the given distribution from the line of equal distribution.

WHICH MEASURE OF DISPERSION TO USE:

The choice of a suitable measure depends on the following two factors:

(i) **The type of data available.** If they are few in number, or contain extreme values, avoid the standard deviation. If they are generally skewed, avoid the mean deviation as well. If they have gaps around the quartiles, the quartile deviation should be avoided. If there are open-end classes, the quartile measure of dispersion should be preferred.

(ii) The purpose of investigation. In an elementary treatment of statistical series in which a measure of variability is desired **only for itself** any of three measures, namely, range, quartile deviation and average deviation, would be acceptable. Probably the average deviation would be better. However, in usual practice, the measure of variability is employed in **further statistical analysis**. For such a purpose the standard deviation, by far, is the most popularly used. Practically all advanced statistical methods deal with variability and centre around the standard deviation. Hence, unless the circumstances warrant the use of any other measure, we should make use of standard deviation for measuring variability.

KEY TERMS:

Qualitative Data: Qualitative Data reflect non-numeric features or qualities of experimental units.

Quantitative Data: Data that possess numerical properties are known as quantitative data.

Variable: A variable is a characteristic that may take on different values at different times, places or situations.

Investigator : Investigator (or enumerator) is a person who collects the information.

Questionnaire: A list of questions properly selected and arranged pertaining to the investigation.

Respondent: A person who fills the questionnaire or supplies the required information.

Classification: Classification is the grouping of related facts into classes.

Class Limits: The class limits are the lowest and the highest values that can be included in the class. For example, take the class 30-50. The lowest value of the class is 30 and the highest 50. The two boundaries of class are known as the lower limit and the upper limit of the class. The **Lower Limit** of a class is the value below which there can be no item in the class. The **Upper Limit** of a class is the value above which no item can belong to that class.

Class Intervals: The difference between the Upper and Lower Limit of a class is known as class interval of that class. For example, in the class 100 – 200, the class interval is 100 (i.e., 200 minus 100).

Class Frequency: The number of observations corresponding to a particular class is known as the frequency of that class or the class frequency.

Class Mid-Point or Class Mark: It is the value lying half-way between the lower and upper class limits of a class-interval. Mid-point of a class is ascertained as follows:

Mid-point of a class =
(Upper limit of the class + Lower limit of the class)/ 2

There are two methods of classifying the data according to class-intervals, namely (i) 'exclusive' method, and (ii) 'inclusive' method.

(i) **'Exclusive' Method:** When the class intervals are so fixed that the upper limit of one class is the lower limit of the next class it is known as the exclusive method of classification.

(ii) **'Inclusive' Method:** Under the 'inclusive' method of classification, the upper limit of one class is included in that class itself.

Open-end Classes: Open-end classes are those in which lower limit of the first class and the upper limit of the last class are not known.

Spurious or nonsense correlation: Correlation observed between variables that cannot conceivably be causally related is called spurious or nonsense correlation.