Statistical tools are found useful in progressively increasing number of disciplines. In ancient times the statistics or the data regarding the human force and the wealth available in their land had been collected by the rulers. Nowadays the fundamental concepts of statistics are considered by many to be the essential part of their general knowledge. To quote an enthusiast:

'When the history of modern times is finally written, we shall read it as beginning with the age of steam and then progressing through the age of electricity to that of statistics.'

1. Origin and Growth

The origin of the word **'statistics'** has been traced to the Latin word **'status'**, the Italian word ***'statista'***, the French word **'statistique'** and the German word **'statistik'**. All these words mean political state. Cottfried Achenwall used this word first to say that statistics is a separate science. He called Statistics as "the political science of the several countries". In India, population statistics had been collected during the rule of Chandragupta Maurya. Todarmal had maintained land records during Akbar's rule. Statistical facts about the state administration in the country are found in Kautilya's "Arthashastra".

Statistics originated as statecraft and has grown markedly. It aids individuals as well as organizations. Governments and private enterprises alike increasingly use the statistical techniques. In science or humanity, agriculture or industry, the use of statistics is unavoidable. That is why it is said,

'Statistics without other sciences has no root and other sciences without statistics bear no fruit'.

2. Meaning

The word 'Statistics' is used in two different meanings. As a plural word it means data or numerical statements. As a singular word it means the science of Statistics and Statistical methods. The word 'Statistics' is also used currently as a singular to mean data. Hence, the meaning is to be understood from the context.

3. Definitions

Definitions in which the word 'statistics' means data or numerical statements are considered first.

By Statistics we mean aggregate 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

-Prof.Horace Secrist.

It is an exhaustive definition and gives the various characteristics of Statistical data.

Statistics is an aggregate of facts. It is not a single value. It is a set of values.

Statistics are affected to a marked extent by multiplicity of causes. There are many causes. Consequently, the resulting values differ. Consider a class of students. Their marks in a subject are not equal because of factors like grasp, understanding questions, method of answering, etc.

Statistics are numerically expressed. Qualities are not statistics. Proper quantities are statistics

Statistics are enumerated or estimated according to reasonable standards of accuracy. Statistics are either real values

5. Characteristics

1. Statistics is a Quantitative Science. It does not deal with qualities. It deals only with quantities like mean mark, the correlation coefficient between expenditure on advertisement and sales, etc. Even while determining the association between two attributes (qualities), those qualities are to be expressed only in numbers such as the number of literates, the number of persons employed, etc.

2. It never considers a single item. Only a set of items is considered and a single item is never considered. A physician considers the nature of blood of one patient but a statistician does not consider the mark of a student in one subject alone.

3. The values should be different. All the values in a set of items should not be one and the same. They should be different. Otherwise there is no use for any statistical measure. It is not necessary to calculate mean, standard deviation, etc. when all the values are equal.

4. Inductive logic is applied. Although in certain studies all the units are observed, most often sample surveys are conducted. A sample of units is observed and from the data so collected generalisation about the population is made. Total rainfall, agricultural production, etc. in India are estimated on the basis of suitable samples.

5. Statistical results are true on the average. The results in statistics are not as exact as in other sciences. For example, under specific conditions, the exact distance an object will fall in a given time can be estimated. But the estimated values in Statistics using regression or time series do not occur exactly. Some values are greater than the estimates and some others are less.

6. Statistics is liable to be misused. Statistics must be used by experts. Either due to their ignorance of the intricacies of the scientific concepts and techniques or deliberately people misuse Statistics. For example, a person of height 175 cms. who does not know swimming claims that he can cross safely a river with the mean depth of water being 150 cms. Let us see how he is wrong. The concept of mean can not be applied to the river water. Further, the mean values 175 and 150 cms. should not be compared for arriving at such a conclusion. At a particular spot, the depth of water can be more than 175 cms. or there can be a heavy undercurrent. As W.I.King says, *"Statistics are like clay of which one can make a god or devil as one pleases."*

6. Scope and Uses

The following are the importance of Statistics. The uses of Statistics are given in general first and in four specific disciplines later.

Statistics has pervaded almost all spheres of human activity. Statistical techniques such as sampling are applied by all people. Even a rustic examines a handful of rice before buying a sack. Everyone tastes one or two fruits before buying a bunch of grapes. Housewives examine only one grain of rice from a boiling pot. In examinations and interviews only a few questions are asked to each person. From these simple situations to the highest level of research and decision making, statistical tools are immensely useful. New drugs are tested statistically on guinea-pigs before prescribing for human beings. The role of Statistics and statistical data in planning and administration is known even to the common man. According to Tippett, "Statistics affects everybody and touches life at many points."

(i) Industry

Statistical methods and statistical data are very useful for an industry. They play the most complementary role.

Statistical methods help in the maintenance of records of inventory, purchase, production and marketing. They also help to do the difficult task of deciding when and where to purchase the raw materials, how to schedule the production, how to exploit the market conditions and how best to utilise the available men, machinery and capital. Each and every activity of an industry can be determined by using statistical methods. Industrialists need not any longer adopt trial and error methods. They can weigh the advantages and disadvantages of each course of action and choose the best among them.

Suppose an industrialist wants to select a suitable place for his new industry. He has to consider the available labour force, the distance and the means of transport of the raw

6. Scope and Uses

The following are the importance of Statistics. The uses of Statistics are given in general first and in four specific disciplines later.

Statistics has pervaded almost all spheres of human activity. Statistical techniques such as sampling are applied by all people. Even a rustic examines a handful of rice before buying a sack. Everyone tastes one or two fruits before buying a bunch of grapes. Housewives examine only one grain of rice from a boiling pot. In examinations and interviews only a few questions are asked to each person. From these simple situations to the highest level of research and decision making, statistical tools are immensely useful. New drugs are tested statistically on guinea-pigs before prescribing for human beings. The role of Statistics and statistical data in planning and administration is known even to the common man. According to Tippett, "Statistics affects everybody and touches life at many points."

(i) Industry

Statistical methods and statistical data are very useful for an industry. They play the most complementary role.

Statistical methods help in the maintenance of records of inventory, purchase, production and marketing. They also help to do the difficult task of deciding when and where to purchase the raw materials, how to schedule the production, how to exploit the market conditions and how best to utilise the available men, machinery and capital. Each and every activity of an industry can be determined by using statistical methods. Industrialists need not any longer adopt trial and error methods. They can weigh the advantages and disadvantages of each course of action and choose the best among them.

Suppose an industrialist wants to select a suitable place for his new industry. He has to consider the available labour force, the distance and the means of transport of the raw materials and the finished goods, the nature and the cost of energy, the wages and the taxation rates, the climatic, political and economic conditions, the scope for growth and other relevant factors. Statistical tools come handy to select the best location.

Once an industry is started , 'Statistical Quality Control' (S.Q.C. is the abbreviation) is useful for the production of quality goods at reduced inspection cost. It minimises wastage and rework. Shewharts control charts are drawn to find out whether manufacturing process is to be left as it is or whether any fault which is to be rectified has crept in. Acceptance sampling helps to estimate the quality of the manufactured products.

By using time series, regression and the like, the demand can be forecast. Further, market research reveals the likely changes in consumer preferences. Purchases, production and sales can be planned accordingly.

Inventory control helps to coordinate purchase of raw materials, production, stock of finished goods and sales. Less inventory may affect the production schedule sometimes. But more inventory increases the capital. More stock of finished goods is not desirable. But the manufacturer should be ready to face the situation arising out of the fluctuations in the market. Probability, decision theory and other related techniques are of great use in these circumstances.

(ii) Commerce and Business

Increasing size of the population and changing attitude of the people to spend more cause increase in the volume of business. Tastes and preferences of the consumers are changing. New fashions are introduced. Competition is growing. Cheaper substitutes are being invented. Goods are manufactured long before they are offered for sale. Manufacturers, marketing agents and consumers are strangers to each other. Business has turned more risky. Unplanned acts become pitfalls. Enterprising businessmen take calculated risks and reap the rewards of success.

The location and the size of a new business house can be determined statistically. Available opportunities for business, scope for improvement and prevailing nature of facilities help to choose a place and to decide a size.

Market survey gives the demand condition and also the likely changes. Buying and selling can be adjusted accordingly. Analysis of time series enables a businessman to forecast reliably. When the trend, cyclical fluctuation, etc. indicate larger business, the businessman should arrange to have necessary stock of goods, employ more personnel and have more godown facilities so that his profit might increase. When a decline in business is foreseen, he has to reduce his expenditure so that his profit may not be reduced heavily. For that purpose, he can reduce the stock of goods, send away temporary employees, save by surrendering the godwon, etc. Business Barometers, regression analysis and extrapolation also give good forecasts. A number of theories of business forecasting such as Action and Reaction theory and Sequence or time-lag theory have been developed.

Marketing methods can be examined and if advantageous, new strategies can be evolved using appropriate statistical techniques. The volumes of sales of various products in different regions are to be considered. Market surveys indicate the probable demand for the products, change in consumer tastes, whether a product is losing demand or whether new products are awaited in the market, etc. Suitable decisions can be taken after considering the response of the consumers and the marketing strategy of the competitors. Pricing a product, advertising, finding suitable marketing personnel and .answers for questions like whether to concentrate and improve the current market or to try for a new market, etc. are the aspects which are decided upon. Probability,estimation, tests of significance and decision theory are very useful tools for deciding upon the above factors.

Statistical methods such as sampling are used in auditing. Price index numbers or price deflators are important for inflation accounting.

The importance of the study of the correlation between the expenditure on advertisement and the sales is understandable. Similarly, the correlations between recruitment test scores and actual performance of marketing personnel, capital and profit, rate of discount and volume of sales, rate of incentives to the merchants and the amount of sales of a product, etc. enable decision making.

(iii) Economics

Statistical techniques are very extensively used in Economics. The laws of Economics are not so exact as the laws of physical sciences. The nature of statistical methods is extremely suitable for examining the theoretical laws and empirical relations of Economics. Jevous felt even in 1871 that "The deductive science of economy must be verified and rendered useful by the purely inductive science of statistics." Statistics is so widely used in Economics to prompt Sir R.A. Fisher to complain as far back in 1926 of *"the painful misapprehension that Statistics is a branch of Economics."*

Statistical data and techniques are powerful aids in economic analysis. They are also useful in the calculation of national income, in the assessment of the gravity of poverty and in the evaluation of the magnitude of unemployment. They help economic planning and formulation of welfare schemes, Statistics of production over a period show the progress of a nation and also enable comparison between nations. Exchange statistics indicate the volume of transactions. Quantitative study of supply, demand and price gives clearly the condition of the market.

Statistical methods enrich the quantitative study of Economics. Economics is concerned with production, distribution, consumption, savings, investment, etc. Equitable distribution of national income and wealth is one of the set goals of socialist governments. Lorenz curve exhibits the disparity in their distributions. Consumption patterns of people depend upon their income, habits and customs. Consumption shows the way in which people of different strata spend their income. Savings augment investment and investment enhances production.

Index numbers are rightly called Economic Barometers. Index numbers of wholesale prices, index numbers of industrial production and the like indicate the nature of the economy and the direction in which it is moving.

Sampling techniques prove their use in Economics also. The theory of estimation enables the economists to estimate the unknown values of the population. Econometric models play an important role in forecasting. Many theories have been developed for forecasting. Economic Rhythm theory is one among them.

A new discipline called Econometrics has come into being. It is the result of the wide application of Statistics and Mathematics in Economics. Econometrics was first applied in the derivation of the demand function. It is difficult to analyse demand functions, cost functions, production functions and consumption functions. Statistical tools are of immense use to overcome this difficulty.

Statistical techniques remove the bottlenecks in economic thought and planning. They facilitate economic growth. For example, the study of family budgets in a town was the basis for Engel's law of consumption. Actual observations on the buyers in the market became the basis for the Revealed Preference Analysis of Prof.Samuelson.

The method of curve fitting by the principle of least squares and exponential smoothing are useful tools for making projections into the future. Economic planning comprises projection, laying standards, evaluating performance, etc. Statistical methods are indispensable in these spheres.

The importance of Statistics is clear from the following words of Dr.Bowley: "No student of Political Economy can pretend to know complete equipment unless he is master of the methods of statistics, knows its difficulties, can see where accurate figures are possible, can criticise the statistical evidence and has an almost instinctive perception of the reliance that he may place on the estimates given to him."

(iv) Management

In the very old days an entrepreneur might have been successful in personally managing all the activities of his business house. Many business houses catered to the needs of the local people. Their area of operation, quantum of production and sales, etc. were limited. Industrial revolution broke those barriers. The attitude and the opportunities of the business houses changed vastly. Recently World Trade Organisation (W.T.O.) and General Agreement on Trade and Tariff (G.A.T.T.) have paved the way for globalisation. Computers and other electronic devices have shrunk the world to become a small village. Size of the population, outlook of the people in earning and spending, availability of scarce raw materials and skilled labourers in limited localities, competitive spirit of the people to produce quality products at cheaper prices, ability to find substitutes, etc. have opened the flood gates to the young and enterprising managers.

Nowadays talented managers are in great demand to look after the various departments like purchase, production, marketing, finance and so on. Management has become a specilised job. Instead of the members of a family managing various departments of their own businesses, qualified managers are entrusted with those challenging tasks. This leads to the benefit of both-newly recruited managers and their employers.

Statistical data and statistical tools are indispensable. Lord Kelvin once remarked that when you can measure what you are speaking about and express it in numbers you know something about it but when you can not measure it, when you can not express it in numbers your knowledge is of meagre and unsatisfactory kind.

Managers are to be familiar with all the aspects of the business-starting from collection or compilation of relevant data till decision making and execution. It is not impossible for them in this computer age to deal with any amount of data, to classify them or to analyse them, to read the message they convey, to manipulate them according to the needs and the like. Statistical tools come handy. To quote Wallis and Roberts, 'Statistics may be regarded as a body of methods for making wise decisions in the face of uncertainty.'

Even statistical graphs and diagrams help in a small measure. The manager and all others in an organisation feel it a pleasure to know the past conditions, to have a glimpse of the future as it looks today and so on through charts. The role of diagrams in advertisements to catch both the sections of people - the careful and the carefree is not a small one.

A large number of business forecasting techniques are available. A few popular methods among them are

Sequence or Time-lag Theory
Action and Reaction Theory
Economic Rhythm Theory
Specific Historical Theory and
Cross - section Analysis

They tell what is in store. Hence, the managers can look ahead safely before they leap.

Similarly, the analysis of time series is useful not only to understand the past conditions but also to forecast the trend much reliably. Method of least squares and method of moving averages are to be appreciated for giving much reliable future values. Seasonal and Cyclical variations are also components of

The above details make us agree with Prof. Ya-Lin-Chou's opinion: 'Statistics is a method of decision making in the face of uncertainty on the basis of numerical data and calculated risks.'

7. Limitations

Limitations of Statistics are due to the characteristics of the science. Statistics is a very useful tool. But it can not be used for all purposes and in all situations as seen below.

1. Statistics does not deal with qualities. It is a quantitative science which does not deal with qualities directly. In Chemistry, the property (quality) of a gas is studied. In Statistics no quality is studied. But qualities in terms of numbers (frequencies) are considered. Number of males, number of persons cured from a disease and so forth may be considered when necessary.

2. Statistics does not consider a single item. A single item is not considered in Statistics. Only aggregate of items is considered. This is different from the situation where a doctor treats only one patient (at a time).

3. All the values should not be the same. The values in Statistics have to be different. When the amounts of sales in different periods are considered, they will not be equal. The daily productions in a factory will not be the same. But, in Physics or Chemistry laboratory, the readings are same as long as the conditions remain the same. In Statistics, the observations differ from one another.

4. Inductive logic is applied. Under induction, a sample is observed and generalisation for the whole population is made from the sample observations. Almost all statistical enquiries are of this type. Some statistical enquiries may involve population surveys. Even on such occasions deductive logic is not used.

5. Statistical results are not exact. The statistical results are not exact as in natural sciences. The volume of a gas under given pressure and temperature can be estimated accurately. But statistical forecasts using time series or regression do not coincide with the true values.

6. Statistics is one of the methods of studying a problem. Other methods may also be there for studying the problem. The demand for a product may be forecast by statistical techniques. Without applying those techniques, an ordinary businessman can forecast the demand. Because of his experience, he may be able to consider important factors and the cushion to be provided for market fluctuations.

7. Statistics can be misused. Misuse of Statistics has led to the following comments:

'An ounce of truth will produce tonnes of Statistics.'

'Statistics can prove anything'.

'There are three types of lies-lies, damn lies and Statistics-wicked in the order of their naming.'

Some people misuse Statistics deliberately with some ulterior motive. Some others misuse Statistics without properly understanding statistical concepts and techniques.

For example, consider an advertisement that a product is used in millions of families.

Figures like this are believable, convincing and psychologically more appealing. But, people misunderstand that another competing product is not used in more families, or those millions of families are regularly using the product or the product is good, etc.

Similarly, correlation is misunderstood to show the cause and effect relationship between the variables.

STATISTICAL SURVEY

Collection of data is the first stage of any statistical investigation. It is to be planned properly and executed carefully. This is a time when most people consider the entire globe as the area of their interest. The relevant data are enormous. Computers are there to process any amount of data. Computers do not malfunction generally. But data are to be relevant and free from mistakes.Carelessness at any stage including that of collection renders the data useless and the survey a waste. All the aspects of a survey, starting from planning and ending with the writing of the final report, are briefly considered under two broad heads, namely,

1. Planning a survey and 2. Executing a survey.

PLANNING

Various steps of planning are the following:

1. Purpose of the survey
2. Scope of the survey
3. Nature of information required
4. Units to be used
5. Sources of data
6. Techniques to be adopted
7. Choice of frame
8. Accuracy aimed
9. Other considerations.

The decision on every one of these aspects influences others. That is they cannot be thought about in isolation from each other. To start with, a tentative decision is made first on each aspect and then the plan is finalised as a whole.

1. Purpose of the survey

A statistical survey may be for a general purpose or a special purpose. The purpose of the survey should be very clear. Only on the basis of the purpose, the other aspects of planning are decided. Doubts such as whether some data are necessary or not, whether the coverage is to be in a particular method or not and others are bound to arise

... purpose or objective of the survey. The intended uses of the survey are the deciding factors on every aspect. Failure to set out the purpose clearly is bound to lead the survey to confusion. In short, a clear and detailed statement of the problem is essential to plan properly.

2. Scope of the survey

Scope depends on the purpose and the availability of time and resources. Decisions on 'What is the geographical area to be covered ? From whom are the data to be collected ?', etc. are made at this stage.

3. Nature of information required

If the survey is about whole - sale price index number, whole - sale prices are needed. If the survey pertains to cost of living index number, retail prices need be known. Like every aspect of planning, the nature of information required depends on the proposed uses of the survey. The information may be required for a specific purpose such as the revision of salary of all personnel connected with textile industry in Coimbatore city. Or, it may be for a reference purpose. It may be stored and referred whenever necessary.

4. Units to be used

There are two kinds of units, viz, units of collection and units of analysis and interpretation. Rates, ratios, percentages and coefficients serve as units of analysis and interpretation. Those which help to count or measure the observations are the units of collection.

Units of collection are classified into (i) **simple units** and (ii) **composite units**. Simple units result from single conditions. Each condition with one or more restrictions causes a composite unit. For example, day, hour, rupee, kilometre, bale, house and ton are a few **simple units.** Of them, bale and house are **units of production.** The items are produced from natural resources for the use of human beings. Day, hour, kilometre and ton are **units of mensuration**. Rupee is a **pecuniary value unit.** Simple units on restrictions become **composite units**. For example, man - hour, labour - day and ton -kilometre are a few **composite units.** If a factory in which 200 workers are employed is under lock-out for 5 days, $5 \times 200 = 1000$ labour -days are lost. If another factory in which 50 workers are employed is under lock - out for 10 days, $10 \times 50 = 500$ labour-days are lost. On comparison, it is known that the loss in the first factory is twice that of the second. The composite unit enables persons a quicker comparison.

Before a survey, the suitable unit is to be decided. Production, for example, can be measured in rupees, tons, man - hours, etc. One of them is chosen and used throughout the survey. It may be arbitrary or conventional. Its desirable properties can be listed as follows:

(i) It should suit the purpose of the survey.

(ii) It should be simple to understand.

(iii) It should be clear cut and not vague.

(iv) It should be usable throughout the survey so that the results can be compared at different stages.

5. Sources of data*

There are two sources of data, viz., primary source and secondary source. The data which are collected by actual observation or measurement or count are primary data. Either the investigator individually or through his agents or employees collects the data. Secondary data, on the other hand, are compiled from the records of others. It is to be decided at this stage whether primary data or secondary data or both are to be used at each stage.

6. Technique to be adopted**

If the data are collected from every unit which belongs to a survey, the survey is called a population survey or a census survey. The data may be collected from a few selected units. The survey is then called a sample survey. The results of a sample survey are to be generalised for the population as a whole. A sample survey or a population survey is adopted on the basis of the nature, the scope, the cost, the time available and the accuracy aimed.

7. Choice of frame

A frame is a list of all the units of a survey. Each unit has its identification label. For example, students of a College have roll

*For details please refer to chapter 3 **For details please refer to chapter 4

numbers, houses in a municipal area have door numbers with names of the streets, etc. as their identification labels. The frame of every survey is not easily available as mentioned above. There is chance for the frames to be inaccurate, incomplete, inadequate, subject to duplication or out of date. The investigator has to scrutinize the frame and construct one, if necessary. Census of population, telephone directories, pay-rolls, previous such surveys, etc. may provide him the various details.

8. Accuracy aimed

The investigator decides about the degree of accuracy also. Absolute accuracy, that is, 100% accuracy is seldom attained due to the inherent nature of the surveys of this kind. There may be unintentional bias on the part of the investigator or enumerator or informant. The tools of units of measurements such as weighing machines may not be accurate. The degree of accuracy aimed at depends to a larger extent on the object of the survey. In weighing gold, even 1/10 th of a gram is important as it costs heavily. In the case of salt, even a few grams do not affect much. Investigator may wish to collect the data quickly rather than spend a lot of time and money for achieving a slightly higher degree of accuracy.

9. Other considerations

The investigator should consider whether the enquiry is (i) official or semi-official or non-official (ii) confidential (iii) regular or ad hoc iv) initial or repetitive and (v) direct or indirect.

An official enquiry is conducted by or on behalf of a Government. semi-official enquiry is conducted by a body which enjoys overnment patronage. A non-official enquiry is conducted by private gencies or individuals. Depending on the purpose of enquiry, legal nction can be obtained for collecting the data for an official enquiry. ajoling and coaxing may help collection of data for a semi- official quiry. Collecting the data for a non- official enquiry is quite difficult.

The findings of a confidential enquiry are kept secret. The results other enquiries are available to the public.

Enquires carried at regular intervals of time are known as regular enquiries. Census of India is a regular enquiry. Some of the enquiries are to be conducted as and when there is a necessity. They are called ad hoc enquires.

If an enquiry is conducted for the first time, it is an initial enquiry. If an enquiry is a continuation of previous enquiries, it is a repetitive enquiry. For an initial enquiry a plan of data collection is to be formulated. For a repetitive enquiry, a plan exists. It may or may not be necessary to modify it in the light of the past experience.

In a direct enquiry, there is possibility of measuring the characteristics directly. For example, weight and income of the respondents. In an indirect enquiry, there is no possibility of measuring the characteristics directly. For example, intelligence and efficiency of the respondents.

EXECUTION

The plan of any survey is to be followed by proper execution of the survey. The various phases of execution are as follows:

1. Setting up an administrative organisation.
2. Designing of forms.
3. Selecting, training and supervising the field investigators.
4. Controlling the accuracy of the field work.
5. Reducing non - response.
6. Presenting the information.
7. Analysing the information.
8. Preparing the reports.

1. Setting up an administrative organisation

Depending on the nature and the scope of the survey, the existing administrative organisation is to be utilised or a new one is to be set up. If the survey covers a large area, regional offices are to be set up. A central office is to be in charge of collecting all the information from the regional offices.

2. Designing of forms

Questionnaires or chedules or other forms necessary for collecting the information are to be carefully prepared.

3. Selecting, training and supervising the field investigators

Another important task is selecting the proper personnel for field work, imparting uniform training and supervising their field work closely. A well executed field work will make a survey a success and an ill executed one will mar it. If necessary, a preliminary test is to be conducted for the selection of proper personnel. Their pay and other facilities should be encouraging. One or more training courses are to be conducted. The works of the newly recruited personnel are to be carefully watched. For large scale surveys, a supervisor for every few investigators is to be appointed.

4. Controlling the accuracy of the field work

Accuracy is the most important aspect. Proper personnel who are given uniform training and who are supervised, collect accurate information. Close watch on the progress of the work helps to identify the problems and the necessary changes desired. Periodical and sample checks are also useful.

5. Reducing non-response

Whenever there is lack of response, steps are to be taken to collect the information from those units. The persons who are not available in their places or those who refuse to respond may possess certain peculiar characteristics. Those characteristics cannot be known from others.

Some method is to be found out to get the relevant information from them. It is necessary to avoid the loss of the representative character of the information collected.

6. Presenting information*

The collected details of information are to be presented in readily understandable forms. They are classified and presented in statistical tables, diagrams and graphs.

*For details please refer to chapter 5

7. Analysing the information **

The purpose of the survey is achieved at this stage. The collected data are carefully analysed for finding out the details. Rates, ratios, percentages, coefficients and statistical measures are the tools available for analysis and interpretation.

8. Preparing the reports

The final stage is the preparation of the reports of the survey. The reports show the purpose of the survey, the personnel involved, the time and the mode of collection of information, the accuracy, the nature, the coverage, the source of the information, etc. The findings are given. It is not an easy task. It is not just a typing work. On the basis of the knowledge of Statistics and others the findings are interpreted. Suggestions are spelt out.

EXERCISES

UNIT - II

Measures of Central tendency

Quantitative data in a mass exhibit certain general character, that show a tendency to concentrate at certain value tendency is called measure of central tendency (OR) Average.

1. Mean (AM)
2. Median (M)
3. Mode (Z)
4. Geometric Mean (GM)
5. Harmonic Mean (HM)

- Average
  - Mathematical Average
    - AM
    - GM
    - HM
    - Quadratic mean
  - Positional Average
    - median
    - mode
  - Commercial Average
    - moving Average
    - progressive Average
    - Composite Average.

Arithmetic Mean

AM

AM is the total of the values of the items divided by their number.

denoted by $\bar{x}$

$$AM = \Sigma x / n$$

$$= \Sigma f x / n$$

$$= \Sigma f m / n$$

AM

Individual Observation	Discrete method	Class intervals
↓	↓	↓
$\Sigma x / n$	$\Sigma f x / n$	$\Sigma f m / n$

1) Calculate AM

$x = 30, 70, 10, 75, 500, 8, 42, 250, 40, 36$

$\Sigma x = 30 + 70 + 10 + 75 + 3500 + 8 + 42 + 250 + 40 + 36$

$\Sigma x = 1061$

$n = 10$

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

$= 1061/10$

$$\boxed{\bar{x} = 106.1}$$

2) Calculate Am

x -3, -2, -1, 0, 1, 0, 2, 3, 4, 4

$\Sigma \bar{x} = -3 - 2 - 1 + 0 + 1 + 0 + 2 + 3 + 4 + 4$

$\Sigma x = 8$

$n = 10$

$\bar{x} = 8/10$

$\boxed{\bar{x} = 0.8}$

3) Calculate mean

x	2	3	4	5	6
f	10	25	30	25	10

x	f	fx
2	10	20
3	25	75
4	30	120
5	25	125
6	10	60
	100	400

$\bar{x} = \Sigma fx/n$

$= 400/100$

$\boxed{\bar{x} = 4}$

4) Calculate mean

x	40	50	54	60	68	80
f	10	18	20	39	15	8

x	f	fx
40	10	400
50	18	900
54	20	1080
60	39	2340
68	15	1020
80	8	640
	110	6380

$\bar{x} = \Sigma fx / n$

$= 6380/110$

$\boxed{\bar{x} = 58}$

calculate mean

x	20-30	30-40	40-50	50-60	60-70	70-80
f	5	8	12	15	6	4

x	f	m	fm
20-30	5	25	125
30-40	8	35	280
40-50	12	45	540
50-60	15	55	825
60-70	6	65	390
70-80	4	75	300
	50		2460

$\bar{x} = \Sigma fm / n$

$= 2460/50$

$\boxed{\bar{x} = 49.2}$

calculate mean

x	0-10	10-20	20-30	30-40	40-50	50-60
f	5	10	25	30	20	10

x	f	m	fm
0-10	5	5	25
10-20	10	15	150
20-30	25	25	625
30-40	30	35	1050
40-50	20	45	900
50-60	10	55	550
	100		3300

$\bar{x} = \Sigma fm / n$

$= 3300/100$

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

Median

It is the value of the middle most item when all items are in the order of mangnitude. denoted by "M" (or) Me.

M

- Individual Observation → $\frac{N+1}{2}$
- Discrete method → $\frac{N+1}{2}$
- Class interval → $\frac{N+1}{2}$ (and) $L + \frac{i(N/2 - cf)}{f}$

1) calculate median: 6, 9, 21, 5, 7, -2, 0, 32, 9

-2, 0, 5, 6, $\boxed{7}$, 9, 9, 21, 32

$N = 9$

$M = 9+1/2 = 10/2$

$= 5$

$\boxed{M = 7}$

2) Calculate median. 57, 58, 61, 42, 38, 65, 72, 66

38, 42, 57, $\boxed{58, 61}$, 65, 66, 72

$M = \frac{N+1}{2}$

$= \frac{8+1}{2} = 9/2$

$= 4.5$

$= \frac{58+61}{2}$

$\boxed{M = 59.5}$

Calculate Median

x	3	4	5	6	7	8	9	10
f	1	5	6	7	10	15	10	5

x	f	cf
3	1	1
4	5	6
5	6	12
6	7	19
7	10	29
8	15	44
9	10	54
10	5	59
	59	

$M = \frac{N+1}{2}$

$= \frac{59+1}{2}$

$= 60/2$

$= 30$

$\boxed{M = 8}$

Calculate M

x	50	75	100	150	250
f	8	14	10	5	3

x	f	cf
50	8	8
75	14	22
100	10	32
150	5	37
250	3	40
	40	

$M = \frac{N+1}{2}$

$= \frac{40+1}{2}$

$= 41/2$

$= 20.5$

$\boxed{M = 75}$

Calculate Median

x	145-150	150-155	155-160	160-165	165-170	170-175
f	2	5	10	8	4	1

x	f	cf
145-150	2	2
150-155	5	7
155-160	10	17
160-165	8	25
165-170	4	29
170-175	1	30
	30	

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

$= 31/2$

$= 15.5$

$l = 155$; $i = 160 - 155 = 5$

$n/2 = 30/2 = 15$; $cf = 7$

$f = 10$

$$M = l + \frac{i\left(N/2 - cf\right)}{f}$$

$$= 155 + \frac{5(15-7)}{10}$$

$$= 155 + \frac{5(8)}{10}$$

$$= 155 + \frac{40}{10}$$

$$= 155 + 4$$

$$\boxed{M = 159}$$

Calculate Median

x	10-25	25-40	40-55	55-70	70-85	85-100
f	6	20	44	26	3	1

x	f	cf
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100
	$\underline{100}$	

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

$$= \frac{100+1}{2}$$

$$= 101/2$$

$$= 50.5$$

$L = 40$; $i = 15$; $f = 44$; $cf = 26$; $N/2 = 50$

$$M = L + \frac{i(N/2 - cf)}{f}$$

$$= 40 + \frac{15(50-26)}{44}$$

$$= 40 + \frac{15(24)}{44}$$

$$= 40 + 8.18$$

$$\boxed{M = 48.18 \simeq 48}$$

$$\boxed{M = 48}$$

Mode:

It is the value which has the greatest frequency density. Z (or) M_o denotes mode.

Mode

- Individual → No formula
- discrete → No formula
- class interval → $L + \left(\frac{i \cdot D_1}{D_1 + D_2}\right)$ (OR)

$$L + \left[\frac{i(f_1 - f_0)}{2f_1 - f_0 - f_2}\right] \text{ OR}$$

$$L + \left[\frac{i f_2}{f_1 + f_2}\right]$$

1) Determine mode

320, 395, 342, 444, 551, 395, 425, 417, 395, 401, 39

$\boxed{Z = 395}$

2) 3, 6, 7, 5, 8, 4, 9

$\boxed{Z = \text{No mode}}$

3) 25, 32, 24, 27, 32, 27, 25, 32, 24, 27, 25, 24

$\boxed{Z = 25, 32, 24, 27}$

· Determine the mode

1)

x	3	4	5	6	7	8	9
f	10	25	32	38	61	47	34

$$\boxed{Z = 7}$$

2)

x	18	20	22	24
f	55	120	108	45

$$\boxed{Z = 20}$$

3)

x	10	11	12	13	14	15	16	17	18
f	10	12	15	19	20	8	4	3	2

$$\boxed{Z = 13}$$

Determine the mode.

)

x	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
f	9	12	15	16	17	15	10	13

$L = 20$; $i = 25 - 20 = 5$; $D_1 = 17 - 6 = 1$; $D_2 = 17 - 15 = 2$

$$Z = L + \left[\frac{i D_1}{D_1 + D_2}\right]$$

$$= 20 + \frac{5(1)}{1+2}$$

$$= 20 + 5/3$$

$$= 20 + 1.67$$

$$\boxed{Z = 21.67}$$

· Determine the mode.

1)

x	3	4	5	6	7	8	9
f	10	25	32	38	61	47	34

$\boxed{Z = 7}$

2)

x	18	20	22	24
f	55	120	108	45

$\boxed{Z = 20}$

3)

x	10	11	12	13	14	15	16	17	18
f	10	12	15	19	20	8	4	3	2

$\boxed{Z = 13}$

Determine the mode.

1)

x	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
f	9	12	15	16	17	15	10	13

$L = 20$; $i = 25 - 20 = 5$; $D_1 = 17 - 6 = 1$; $D_2 = 17 - 15 = 2$

$$Z = L + \left[\frac{i D_1}{D_1 + D_2}\right]$$

$$= 20 + \frac{5(1)}{1+2}$$

$$= 20 + 5/3$$

$$= 20 + 1.67$$

$$\boxed{Z = 21.67}$$

x	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16
f	45	50	65	70	80	25	20	18

$L = 4$; $i = 2$; $D_1 = 65 - 50 = 15$; $D_2 = 70 - 65 = 5$

$$Z = L + \frac{i D_1}{D_1 + D_2}$$

$$= 4 + \frac{2(15)}{15+5}$$

$$= 4 + 30/20$$

$$= 4 + 1.5$$

$$\boxed{Z = 5.5}$$

Geometric Mean (GM)

GM of N values is the N^{th} root of the product of the N values.

GM

Individual → $\text{Antilog}\left(\frac{\Sigma \log x}{n}\right)$

Discrete → $\text{Antilog}\left(\frac{\Sigma f \log x}{n}\right)$

Class interval → $\text{Antilog}\left(\frac{\Sigma f \log m}{n}\right)$.

find Gm 3, 6, 24, 48

x	$\log x$
3	0.4771
6	0.7782
24	1.3802
48	1.6812
	4.3167

$$Gm = \text{Antilog}\left(\frac{\Sigma \log x}{n}\right)$$

$$= \text{Antilog}\left(\frac{4.3167}{4}\right)$$

$$= \text{Antilog}(1.0792)$$

$$\boxed{Gm = 12}$$

---

find Gm 120, 130, 145

x	$\log x$
120	2.0792
130	2.1139
145	2.1614
	6.3545

$$Gm = \text{Antilog}\left(\frac{\Sigma \log x}{n}\right)$$

$$= \text{Antilog}\left(\frac{6.3545}{3}\right)$$

$$= \text{Antilog}(2.1182)$$

$$\boxed{Gm = 131.28}$$

Calculate Gm

x	10	15	25	40	50
f	4	6	10	7	3

x	f	$\log x$	$f \log x$
10	4	1	4
15	6	1.1761	7.0566
25	10	1.3979	13.9790
40	7	1.6021	11.2147
50	3	1.6990	5.0970
	30		41.3473

$$Gm = \text{Antilog}\left(\frac{\Sigma f \log x}{n}\right)$$

$$= \text{Antilog}\left(\frac{41.3473}{30}\right)$$

$$= \text{Antilog}(1.3782)$$

$$\boxed{Gm = 23.89}$$

find GM

x	20	21	22	23	24	25
f	4	2	7	1	3	1

x	f	$\log x$	$f \log x$
20	4	1.3010	5.2040
21	2	1.3222	2.6444
22	7	1.3424	9.3968
23	1	1.3617	1.3617
24	3	1.3802	4.1406
25	1	1.3979	1.3979
	18		24.1454

$GM = \text{Antilog}\left(\frac{\Sigma f \log x}{n}\right)$

$= \text{Antilog}\left(\frac{24.1454}{18}\right)$

$= \text{Antilog}(1.3414)$

$\boxed{GM = 21.95}$

compute GM

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

x	f	m	$\log m$	$f \log m$
0-10	5	5	0.6990	3.4950
10-20	7	15	1.1761	8.2327
20-30	15	25	1.3979	20.9685
30-40	25	35	1.5441	38.6025
40-50	8	45	1.6532	13.2256
	60			84.5243

$GM = \text{Antilog}\left(\frac{\Sigma f \log m}{n}\right)$

$= \text{Antilog}(84.5243/60)$

$= \text{Antilog}(1.4087)$

$\boxed{GM = 25.63}$

Harmonic Mean

HM is the reciprocal of the mean of the reciprocal of the values.

HM

Individual	Discrete	Class interval
$\frac{N}{\Sigma(1/x)}$	$\frac{N}{\Sigma(f/x)}$	$\frac{N}{\Sigma(f/m)}$

1) calculate HM. 6, 15, 35, 40, 900, 520, 300, 400, 1800, 2000.

x	$1/x$
6	0.1667
15	0.0667
35	0.0286
40	0.0250
900	0.0011
520	0.0019
300	0.0033
400	0.0025
1800	0.0006
2000	0.0005
	0.2969

$$HM = \frac{N}{\Sigma(1/x)}$$

$$= \frac{10}{0.2969}$$

$$\boxed{HM = 33.68}$$

calculate HM

x	10	12	14	16	18	20
f	5	18	20	10	6	1

x	f	f/x
10	5	0.5
12	18	1.5
14	20	1.4286
16	10	0.6250
18	6	0.3333
20	1	0.0500
	60	4.4369

$$HM = \frac{N}{\Sigma(f/x)}$$

$$= 60/4.4369$$

$$\boxed{HM = 13.52}$$

calculate HM

x	0-10	10-20	20-30	30-40	40-50
f	8	12	20	6	4

x	f	m	f/m
0-10	8	5	1.6
10-20	12	15	0.8
20-30	20	25	0.8
30-40	6	35	0.1714
40-50	4	45	0.0889
	50		3.4603

$$HM = \frac{N}{\Sigma[f/m)}$$

$$= \frac{50}{3.4603}$$

$$\boxed{HM = 14.45}$$

calculate Am, Gm and Hm 120, 130, 145

x	$\log x$	$1/x$
120	2.0792	0.0083
130	2.1139	0.0077
145	2.1614	0.0069
395	6.3545	0.0229

$Am = \Sigma x / n$

$= 395/3$

$\boxed{Am = 131.67}$

$Gm = \text{Antilog}\left(\frac{\Sigma \log x}{n}\right)$

$= \text{Antilog}\left(\frac{6.3545}{3}\right)$

$= \text{Antilog}(2.1182)$

$\boxed{Gm = 131.28}$

$Hm = \frac{N}{\Sigma(1/x)}$

$= 3/0.0229$

$\boxed{HM = 131}$

Show that $Am > Gm > HM$

x	20	21	22	23	24	25
f	4	2	7	1	3	1

x	f	$\log x$	$f \log x$	f/x	fx
20	4	1.3010	5.2040	0.2	80
21	2	1.3222	2.6444	0.0952	42
22	7	1.3424	9.3968	0.3182	154
23	1	1.3617	1.3617	0.0435	23
24	3	1.3802	4.1406	0.1250	72
25	1	1.3979	1.3979	0.04	25
	18		24.1454	0.8219	396

$$Am = \Sigma fx / n$$

$$= 396 / 18$$

$$\boxed{Am = 22}$$

$$Gm = \text{Antilog}\left(\frac{\Sigma f \log x}{n}\right)$$

$$= \text{Antilog}\left(\frac{24.1454}{18}\right)$$

$$= \text{Antilog}\ (1.3414)$$

$$\boxed{Gm = 21.95}$$

$$Hm = \frac{N}{\Sigma (f/x)}$$

$$= \frac{18}{0.8219}$$

$$\boxed{HM = 21.90}$$

$$Am > Gm > Hm$$

$$22 > 21.95 > 21.90$$

Calculate Am, Gm and Hm

x	0-19	20-39	40-59	60-79	80-99
f	5	15	35	15	10

x	f	m	fm	log m	f log m	f/m
0-19	5	9.5	47.5	0.9777	4.8885	0.5263
20-39	15	29.5	442.5	1.4698	22.0470	0.5085
40-59	35	49.5	1732.5	1.6946	59.3110	0.7071
60-79	15	69.5	1042.5	1.8420	27.6300	0.2158
80-99	10	89.5	895	1.9518	19.5180	0.1117
	80		4160	-	133.3945	2.0694

AM

$$\bar{x} = \Sigma fm/n$$

$$= 4160/80$$

$$\boxed{\bar{x} = 52}$$

GM

$$GM = \text{Antilog}\left(\frac{\Sigma f \log m}{n}\right)$$

$$= \text{Antilog}\left(\frac{133.3945}{80}\right)$$

$$= \text{Antilog}\,(1.6674)$$

$$\boxed{GM = 46.50}$$

HM:

$$= \frac{N}{\Sigma(f/m)}$$

$$= 80/2.0694$$

$$\boxed{HM = 38.66}$$

Formulas

	Individual	Discrete	Class interval
Mean (x)	$\Sigma x / n$	$\Sigma f x / n$	$\Sigma f m / n$
Median	$\frac{N+1}{2}$	$\frac{N+1}{2}$	$\frac{N+1}{2}$ and $L + \frac{i(N/2 - cf)}{f}$
Mode	–	–	$L + \left(\frac{i D_1}{D_1 + D_2}\right)$
Harmonic mean	$\frac{N}{\Sigma(1/x)}$	$\frac{N}{\Sigma(f/x)}$	$\frac{N}{\Sigma(f/m)}$
Geometric mean	$\text{Antilog}\left(\frac{\Sigma \log x}{n}\right)$	$\text{Antilog}\left(\frac{\Sigma f \log x}{n}\right)$	$\text{Antilog}\left(\frac{\Sigma f \log m}{n}\right)$.

UNIT - 3

Measurement of dispersion

Absolute measure
- Range
- QD
- MD
- SD
- Variance

dispersion measure
- Co-eff of range
- Co-eff of QD
- Co-eff of MD
- Co-eff of variance.

Range

Difference between the greatest and lowest values.

$$\text{Range} = L - S$$

$$\left.\begin{array}{r}\text{Co-eff of}\\ \text{Range}\end{array}\right\} = \frac{L-S}{L+S}$$

Range

Individual
$$R = L - S$$
$$= \frac{L-S}{L+S}$$

Class interval
$$L - S$$
$$\frac{L-S}{L+S}.$$

find the value of range 8, 10, 5, 9, 12, 11

$L = 12$

$S = 5$

$\text{Rang} = L - S$

$= 12 - 5$

$\boxed{\text{Range} = 7}$

$$\left.\begin{matrix}\text{Co-eff of} \\ \text{range}\end{matrix}\right\} = \frac{L - S}{L + S}$$

$$= \frac{12 - 5}{12 + 5}$$

$$= 7/17$$

$\boxed{\text{Co-eff} = 0.4118}$

calculate range

x	60-62	63-65	66-68	67-71	72-74
f	5	18	42	27	8

$L = 74.5$; $S = 59.5$

$\text{Range} = L - S$

$= 74.5 - 59.5$

$\boxed{\text{Range} = 15}$

$$\left.\begin{matrix}\text{Co-eff of} \\ \text{range}\end{matrix}\right\} = \frac{L - S}{L + S}$$

$$= \frac{74.5 - 59.5}{74.5 + 59.5} = 15/134$$

$\boxed{\text{Co-eff of range} = 0.1119}$

Quartile deviation (QD)

It is the half of the diffrence between the first and the third quartiles. Hence it is called semi inter Quartile range.

$$QD = \frac{Q_3 - Q_1}{2}$$

$$\text{co-eff} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$Q_1 = \frac{N+1}{4}$$

$$Q_3 = 3\left(\frac{N+1}{4}\right).$$

1) find QD 391, 384, 591, 407, 672, 522, 777, 733, 1490, 2488.

384, 391, 407, 522, 591, 672, 733, 777, 1490, 2488

$Q_1 = N+1/4 \quad = \frac{10+1}{4}$

$= 11/4 = 2.75 \simeq 3$

$\boxed{Q_1 = 407}$

$Q_3 = 3\left(\frac{N+1}{4}\right)$

$= 3(2.75)$

$= 8.25 \simeq 8$

$\boxed{Q_3 = 777}$

$$QD = \frac{Q_3 - Q_1}{2}$$

$$= \frac{777 - 407}{2}$$

$$= \frac{370}{2}$$

$$\boxed{QD = 185}$$

$$\text{Co-eff} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{777 - 407}{777 + 407}$$

$$= \frac{370}{1184}$$

$$\boxed{\text{Co-eff } QD = 0.3125}$$

---

find QD

x	100	200	400	500	600
f	5	8	21	12	6

x	f	cf	
100	5	5	
200	8	13	
400	21	34	Q_1
500	12	46	Q_3
600	6	52	
	52		

$Q_1 = \frac{N+1}{4} = \frac{52+1}{4} = 53/4$

$= 13.25$

$\boxed{Q_1 = 400}$

$Q_3 = 3\left(\frac{N+1}{4}\right) = 3(13.25)$

$= 39.75$

$\boxed{Q_3 = 500}$

$QD = \frac{Q_3 - Q_1}{2}$

$= \frac{500 - 400}{2}$

$= 100/2$

$\boxed{QD = 50}$

$\text{Co-eff} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{500 - 400}{500 + 400}$

$= 100/900$

$\boxed{= 0.1111}$

find the value of QD

x	351-500	501-650	651-800	801-950	951-1100
f	48	189	88	47	28

x	f	cf	
351-500	48	48	
501-650	189	237	Q_1
651-800	88	325	Q_3
801-950	47	372	
951-1100	28	400	
	400		

Q_1

$= \frac{N+1}{4}$

$= 401/4 = 100.25$

$L = 501$; $i = 650 - 501 = 149$; $N/4 = 100$

$cf = 48$; $f = 189$

$= L + \left(i \frac{(N/4 - cf)}{f}\right)$

$= 501 + \frac{149(100-48)}{189}$

$= 501 + \frac{149(52)}{189}$

$= 501 + 40.9947$

$\boxed{Q_1 = 541.9947}$

$\underline{Q_3}$

$= 3\left(\frac{N+1}{4}\right)$

$= 3(100.25)$

$= 300.75$

$L = 651$; $i = 800 - 651 = 149$; $cf = 237$

$f = 88$; $3(N/4) = 3(100) = 300$

$$= L + \frac{i\,(3n/4 - cf)}{f}$$

$$= 651 + \frac{149\,(300 - 237)}{88}$$

$$= 651 + \frac{149\,(63)}{88}$$

$$= 651 + \frac{9387}{88}$$

$= 651 + 106.6704$

$\boxed{Q_3 = 757.6704}$

$$QD = \frac{Q_3 - Q_1}{2}$$

$$= \frac{757.6704 - 541.9947}{2}$$

$= 215.6757/2$

$\boxed{QD = 107.8379}$

$$\text{Co-eff} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{757.6704 - 541.9947}{757.6704 + 541.9947}$$

$$= \frac{107.8379}{1299.6651}$$

$$\boxed{\text{Co-eff} = 0.0829}$$

Mean deviation (OR) Average deviation.

It is the arithmetic mean of the absolute deviations of the values about their arithmetic mean (or) median (or) mode.

MD →

about mean

$$MD = \frac{\Sigma |x - \bar{x}|}{N}$$

$$\text{Co-eff} = \frac{MD}{\text{mean}}$$

about median

$$MD = \frac{\Sigma |x - M|}{N}$$

$$\text{Co-eff} = \frac{MD}{\text{median}}$$

about mode

$$MD = \frac{\Sigma |X - Z|}{N}$$

$$\text{Co-eff} = \frac{MD}{\text{Mode}}$$

1) Calculate MD by mean 1, 2, 3, 4, 5

$\Sigma x = 1+2+3+4+5$

$\Sigma x = 15$

$\bar{x} = \Sigma x / n = 15/5$

$\boxed{\bar{x} = 3}$

x	$\lvert x - \bar{x} \rvert$
1	2
2	1
3	0
4	1
5	2
	6

$MD = \frac{\Sigma \lvert x - \bar{x} \rvert}{n}$

$= 6/5$

$\boxed{MD = 1.2}$

$\text{Co-eff} = \frac{MD}{\bar{x}}$

$= 1.2/3$

$\boxed{\text{Co-eff} = 0.4}$

---

2) Calculate MD by median 7, 4, 10, 9, 15, 12, 7, 9, 7.

4, 7, 7, 7, 9, 9, 10, 12, 15

$m = \frac{N+1}{2} = 9+1/2 = 10/2 = 5$

$M = 9$

x	$\lvert x - M \rvert$
4	5
7	2
7	2
7	2
9	0
9	0
10	1
12	3
15	6
	21

$MD = \frac{\Sigma \lvert x - M \rvert}{n}$

$= 21/9$

$\boxed{MD = 2.33}$

$\text{Co-eff} = \frac{MD}{M}$

$= \frac{2.33}{9}$

$\boxed{\text{Co-eff} = 0.2589}$

3) Calculate MD by mode

2, 4, 4, 6, 3, 1, 7, 9, 5

$Z = 4$

x	$\lvert x - Z \rvert$
2	2
4	0
4	0
6	2
3	1
1	3
7	3
9	5
5	1
	17

$$MD = \frac{\Sigma \lvert x - Z \rvert}{n}$$

$$= 17/9$$

$$\boxed{MD = 1.8889}$$

$$\text{Co-eff} = \frac{MD}{Z}$$

$$= \frac{1.8889}{4}$$

$$\boxed{\text{Co-eff} = 0.4722}$$

Calculate MD by mean, median and mode.

32, 51, 23, 46, 20, 78, 57, 56, 57, 30

x	$\lvert x - \bar{x} \rvert$	$\lvert x - M \rvert$	$\lvert x - Z \rvert$
32	13	16.5	25
51	6	2.5	6
23	22	25.5	34
46	·1	2.5	11
20	25	28.5	37
78	33	29.5	21
57	12	8.5	0
56	11	7.5	1
57	12	8.5	0
30	15	18.5	27
450	150	148	162

$\bar{x} = \frac{\Sigma x}{n} = 450/10$

$\boxed{\bar{x} = 45}$

$m = \frac{N+1}{2}$

$= 11/2 = 5.5$

$= \frac{46+51}{2} = 48.5$

$\boxed{m = 48.5}$

$\boxed{Z = 57}$

MD by $\bar{x}$:

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

$= 150/10$

$\boxed{MD = 15}$

$\text{Co-eff} = \frac{MD}{\bar{x}}$

$= \frac{15}{45}$

$\boxed{\text{Co-eff} = 0.3333}$

MD by m

$= \frac{\Sigma|x-m|}{n}$

$= 148/10$

$\boxed{MD = 14.8}$

$\text{Co-eff} = \frac{MD}{m}$

$= \frac{14.8}{48.5}$

$\boxed{\text{Co-eff} = 0.3052}$

MD by Z

$= \frac{\Sigma|x-Z|}{n}$

$= 162/10$

$\boxed{MD = 16.2}$

$\text{Co-eff} = \frac{MD}{Z}$

$= \frac{16.20}{57}$

$\boxed{\text{Co-eff} = 0.2842}$

Calculate MD by mean

x	2	4	6	8	10
f	1	4	6	4	1

x	f	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
2	1	2	4	4
4	4	16	2	8
6	6	36	0	0
8	4	32	2	8
10	1	10	4	4
	16	96		24

$\bar{x} = \Sigma fx/n = 96/16$

$\boxed{\bar{x} = 6}$

$$MD = \frac{\Sigma f\lvert x-\bar{x}\rvert}{n}$$

$= 24/16$

$\boxed{MD = 1.5}$

$$\left.\begin{matrix}\text{Co-eff of}\\ \text{MD}\end{matrix}\right\} = \frac{MD}{\bar{x}}$$

$$= \frac{1.5}{6}$$

$\boxed{\text{Co-eff of MD} = 0.25}$

x	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	6	5	8	15	7	6	3

x	f	m	fm	$\lvert m-\bar{x}\rvert$	$f\lvert m-\bar{x}\rvert$
0-10	6	5	30	28.4	170.4
10-20	5	15	75	18.4	92
20-30	8	25	200	8.4	67.2
30-40	15	35	525	1.6	24
40-50	7	45	315	11.6	81.2
50-60	6	55	330	21.6	129.6
60-70	3	65	195	31.6	94.8
	50		1670		659.2

$\bar{x} = \Sigma fm / n$

$= 1670/50$

$\boxed{\bar{x} = 33.4}$

$MD = \dfrac{\Sigma f\lvert m-\bar{x}\rvert}{n}$

$= \dfrac{659.2}{50}$

$\boxed{MD = 13.18}$

x	21	25	27	32	41	46	50	55
f	2	3	10	20	15	10	8	2

x	f	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$	cf	$\lvert x-M\rvert$	$f\lvert x-M\rvert$	$\lvert x-Z\rvert$	$f\lvert x-Z\rvert$
21	2	42	16.31	32.62	2	15.5	31	11	22
25	3	75	12.31	36.93	5	11.5	34.5	7	21
27	10	270	10.31	103.1	15	9.5	95	5	50
32	20	640	5.31	106.2	35	4.5	90	0	0
41	15	615	3.69	55.35	50	4.5	67.5	9	135
46	10	460	8.69	86.9	60	9.5	95	14	140
50	8	400	12.69	101.52	68	13.5	108	18	144
55	2	110	17.69	35.38	76	18.5	37	23	46
	70	2612		558			558		558

$\bar{x} = \Sigma fx/n$

$= 2612/70$

$\boxed{\bar{x} = 37.31}$

$M = N+1/2$

$= 71/2$

$= 35.5$

$M = 36.5$

$Z = 32$

mean deviation by mean

$MD = \frac{\Sigma f|x - \bar{x}|}{n}$ Co-eff $= \frac{MD}{\bar{x}}$

$= \frac{558}{70}$ $= \frac{7.97}{37.31}$

$= 0.2136$.

$MD = 7.97$

mean deviation by median

$MD = \frac{\Sigma f|x - m|}{n}$ Co-eff $= \frac{MD}{m}$

$= \frac{558}{70}$ $= \frac{7.97}{36.5}$

$= 0.2184$

$MD = 7.97$

mean deviation by mode

$MD = \frac{\Sigma f|x - z|}{n}$ Co-eff $= \frac{MD}{z}$

$= \frac{558}{70}$ $= \frac{7.97}{32}$

$MD = 7.97$ $= 0.2491$.

---

x	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60
f	8	15	13	20	11	7	3	2	1

Calculate MD by median.

x	f	cf	m	|m - m|	f |m - m|
16 - 20	8	8	18	13.5	108
21 - 25	15	23	23	8.5	127.5
26 - 30	13	36	28	3.5	45.5
31 - 35	20	56	33	1.5	30
36 - 40	11	67	38	6.5	71.5
41 - 45	7	74	43	11.5	80.5
46 - 50	3	77	48	16.5	49.5
51 - 55	2	79	53	21.5	43
56 - 60	1	80	58	26.5	26.5
	80				582

median :

$= \frac{N+1}{2} = 81/2$

$= 40.5$

$L = 31$; $f = 20$; $cf = 36$; $N/2 = 80 = i = 4$

$= 31 + \frac{4(80-36)}{20}$

$= 31 + \frac{4(44)}{20}$

$= 31 + 176/20$

$= 31 + 8.8$

$\boxed{m = 39.8}$

$MD = \frac{\Sigma f|m-m|}{n}$

$= 582/80$

$\boxed{MD = 7.2750}$

$\text{Co-eff} = \frac{MD}{m}$

$= \frac{7.2750}{39.8}$

$\boxed{\text{Co-eff} = 0.1828}$

calculate MD by mode.

x	0-5	5-10	10-15	15-20	20-25	25-30
f	19	28	50	22	10	7

x	f	m	$\lvert m-z \rvert$	$f\lvert m-z \rvert$
0-5	19	2.5	9.7	184.3
5-10	28	7.5	4.7	131.6
10-15	50	12.5	0.3	15
15-20	22	17.5	5.3	116.6
20-25	10	22.5	10.3	103
25-30	7	27.5	15.3	107.1
	136			657.6

$$Z = L + \left(\frac{iD_1}{D_1+D_2}\right)$$

$L = 10$, $D_1 = 50-28 = 22$; $D_2 = 50-22 = 28$; $i = 5$

$$= 10 + \frac{5(22)}{22+28}$$

$$= 10 + \left(\frac{110}{50}\right)$$

$$= 10 + 2.2$$

$$\boxed{Z = 12.2}$$

$$MD = \frac{\Sigma f \lvert m-z \rvert}{n}$$

$$= \frac{657.6}{136}$$

$$\boxed{MD = 4.84}$$

$$\text{Co-eff} = \frac{MD}{Z} = \frac{4.84}{12.2}$$

$$\boxed{\text{Co-eff} = 0.3967}$$

Standard deviation (SD)

It is the root mean square deviation of the values from their arithmetic mean. Also called root mean square deviation.

$$SD \begin{cases} \text{individual} \rightarrow \sqrt{\dfrac{\Sigma x^2}{n} - \left(\dfrac{\Sigma x}{n}\right)^2} \\ \text{discrete} \rightarrow \sqrt{\dfrac{\Sigma f x^2}{n} - \left(\dfrac{\Sigma f x}{n}\right)^2} \\ \text{class interval} \rightarrow \sqrt{\dfrac{\Sigma f m^2}{n} - \left(\dfrac{\Sigma f m}{n}\right)^2} \end{cases}$$

1) 10 students of B.Com class of a college have obtained the following marks in business maths out of 100. Calculate standard deviation. 5, 10, 20, 25. 40, 42, 45, 48, 70. 80

x	x^2
5	25
10	100
20	400
25	625
40	1600
42	1764
45	2025
48	2304
70	4900
80	6400
385	20143

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

$$= \sqrt{\frac{20143}{10} - \left(\frac{385}{10}\right)^2}$$

$$= \sqrt{2014.3 - (38.5)^2}$$

$$= \sqrt{2014.3 - 1482.25}$$

$$= \sqrt{532.05}$$

$$\boxed{\sigma = 23.07}$$

calculate SD for following series.

x	6	9	12	15	18
f	7	12	13	10	8

x	f	fx	x^2	fx^2
6	7	42	36	252
9	12	108	81	972
12	13	156	144	1872
15	10	150	225	2250
18	8	144	324	2592
	50	600		7938

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

$$= \sqrt{\frac{7938}{50} - \left(\frac{600}{50}\right)^2}$$

$$= \sqrt{158.76 - (12)^2}$$

$$= \sqrt{158.76 - 144}$$

$$= \sqrt{14.76}$$

$$\boxed{\sigma = 3.84}$$

find SD

x	0	1	2	3	4	5
f	1	2	4	3	0	2

x	f	x^2	fx	fx^2
0	1	0	0	0
1	2	1	2	2
2	4	4	8	16
3	3	9	9	27
4	0	16	0	0
5	2	25	10	50
	12		29	95

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

$$= \sqrt{\frac{95}{12} - \left(\frac{29}{12}\right)^2}$$

$$= \sqrt{7.9167 - (2.4167)^2}$$

$$= \sqrt{7.9167 - 5.8404}$$

$$= \sqrt{2.0763}$$

$$\boxed{\sigma = 1.44}$$

The following data were obtained while observing the life span of a ~~few~~ few neon lights of a Company calculate SD.

life span (yrs)	4-6	6-8	8-10	10-12	12-14
No. of lights	10	17	32	21	20

x	f	m	m^2	fm	fm^2
4-6	10	5	25	50	250
6-8	17	7	49	119	833
8-10	32	9	81	288	2592
10-12	21	11	121	231	2541
12-14	20	13	169	260	3380
	100			948	9596

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

$$= \sqrt{\frac{9596}{100} - \left(\frac{948}{100}\right)^2}$$

$$= \sqrt{95.96 - (9.48)^2}$$

$$= \sqrt{95.96 - 89.8704}$$

$$= \sqrt{6.0896}$$

$$\boxed{\sigma = 2.47}$$

Combined Standard deviation

When two (or) more groups merge, the mean and standard deviation of the combined group are calculated as follows

Group	Size	mean	SD
1	N_1	$\bar{x}_1$	σ_1
2	N_2	$\bar{x}_2$	σ_2

The mean of the combined group

$$\bar{x}_{12} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N_1 + N_2}$$

The standard deviation of the combined group

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

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

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

The mean and standard deviation of 63 children on an average test are respectively 27.6 and 7.1. To them are added a new group of 26 who have had less training and whose mean is 19.2 and standard deviation is 6.2. How will the value of combined group differ from those of the original 63 children as to (i) mean and (ii) SD?

Sol:

$N_1 = 63$; $\bar{x}_1 = 27.6$; $\sigma_1 = 7.1$

$N_2 = 26$; $\bar{x}_2 = 19.2$; $\sigma_2 = 6.2$

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

$$= \frac{(63 \times 27.6) + (26 \times 19.2)}{63 + 26}$$

$$= \frac{1738.8 + 499.2}{89}$$

$$= \frac{2238}{89}$$

$$\boxed{\bar{x}_{12} = 25.15}$$

combined SD:

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$$d_1 = 27.6 - 25.15 = 2.45$$

$$d_2 = 19.2 - 25.15 = -5.95$$

$$= \sqrt{\frac{63 \times (7.1)^2 + 26 \times (6.2)^2 + 63 \times (2.45)^2 + 26(-5.95)^2}{63 + 26}}$$

$$= \sqrt{\frac{3175.83 + 999.44 + 378.1575 + 920.4650}{89}}$$

$$= \sqrt{\frac{5473.8925}{89}}$$

$$= \sqrt{61.5044}$$

$$\boxed{\sigma_{12} = 7.84}$$

Merge of three groups

There are 20, 30 and 50 employees to the three branches of a concern. Their mean salaries are ₹. 15, 12 and 18 thousands. SD for their salaries are ₹. 3, 5, and 6 thousands respectively. find the mean salary and the SD of salaries of employees of the concern as a whole.

Sol:

$N_1 = 20 \; ; \; \bar{x}_1 = 15 \; ; \; \sigma_1 = 3$

$N_2 = 30 \; ; \; \bar{x}_2 = 12 \; ; \; \sigma_2 = 5$

$N_3 = 50 \; ; \; \bar{x}_3 = 18 \; ; \; \sigma_3 = 6$

$$\bar{x}_{123} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{x}_3}{N_1 + N_2 + N_3}$$

$$= \frac{20\times15 + 30\times12 + 50\times18}{20+30+50}$$

$$= \frac{300 + 360 + 900}{100}$$

$$= 1560/100$$

$$\boxed{\bar{x}_{12} = 15.6}$$

$$\sigma_{123} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

$d_1 = 15 - 15.6 = -0.6$

$d_2 = 12 - 15.6 = -3.6$

$d_3 = 18 - 15.6 = 2.4$

$$= \sqrt{\frac{20\times3^2 + 30\times5^2 + 50\times6^2 + 20\times(-0.6)^2 + 30\times(-3.6)^2 + 50\times(2.4)^2}{20+30+50}}$$

$$= \sqrt{\frac{180 + 750 + 1800 + 7.2 + 388.8 + 288}{100}}$$

$= \sqrt{\frac{3414}{100}}$

$= \sqrt{34.14}$

$\boxed{\sigma_{1,3} = 5.84}$

Co-efficient of variance

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

1) find CV 2, 4, 8, 8, 6, 10

x	x^2
2	4
4	16
8	64
8	64
6	36
10	100
38	284

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

$= 38/6$

$\boxed{\bar{x} = 6.3333}$

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

$= \sqrt{\frac{284}{6} - \left(\frac{38}{6}\right)^2}$

$= \sqrt{47.3333 - (6.3333)^2}$

$= \sqrt{47.3333 - 40.1107}$

$= \sqrt{7.2226}$

$\boxed{\sigma = 2.6875}$

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

$= \frac{2.6875}{6.3333} \times 100$

$\boxed{CV = 42.4344}$

The mean and standard deviation values for the number of runs of 2 players A and B are 55, 65 and 4.2, 7.8 respectively who is more consistent player.

Player A.

$\bar{x} = 55$

$\sigma = 4.2$

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

$= \frac{4.2}{55} \times 100$

$\boxed{CV = 7.64}$

Player B

$\bar{x} = 65$

$\sigma = 7.8$

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

$= \frac{7.8}{65} \times 100$

$\boxed{CV = 12}$

CV of player A is less so player A is more consistent.

From the followings find which firm have greater variability in individual wage.

	firm 1	firm 2
No. of worker	100	200
mean	7	8
SD of wage	2	2.5

Firm 1

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

$= \frac{2}{7} \times 100$

$\boxed{CV = 2.57}$

Firm 2

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

$= \frac{2.5}{8} \times 100$

$\boxed{CV = 31.25}$

CV of Firm 2 is more, Hence there is greater variability in individual wage in firm 2

Calculate CV

40, 41, 45, 49, 50, 51, 55, 59, 60, 60

x	x^2
40	1600
41	1681
45	2025
49	2401
50	2500
51	2601
55	3025
59	3481
60	3600
60	3600
510	26,514

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

$= 510/10$

$\boxed{\bar{x} = 51}$

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

$= \sqrt{\frac{26514}{10} - \left(\frac{510}{10}\right)^2}$

$= \sqrt{2651.4 - (51)^2}$

$= \sqrt{2651.4 - 2601}$

$= \sqrt{50.4}$

$\boxed{\sigma = 7.0993}$

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

$= \frac{7.0993}{51} \times 100$

$\boxed{CV = 13.92}$

Calculate CV

x	50	40	30	20	10
f	80	65	46	25	12

x	x^2	f	fx	fx^2
50	2500	80	4000	2,00,000
40	1600	65	2600	1,04,000
30	900	46	1380	41,400
20	400	25	500	10,000
10	100	12	120	1200
		228	8600	356600

$\bar{x} = \Sigma fx / n$

$\boxed{\bar{x} = 37.7193}$

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

$$= \sqrt{\frac{356600}{228} - \left(\frac{8600}{228}\right)^2}$$

$$= \sqrt{1564.0351 - (37.7193)^2}$$

$$= \sqrt{1564.0351 - 1422.7456}$$

$$= \sqrt{141.2895}$$

$\boxed{\sigma = 11.8865}$

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

$= \frac{11.8865}{37.7193} \times 100$

$\boxed{CV = 31.51}$

Calculate CV.

x	0-10	10-20	20-30	30-40	40-50
f	12	13	21	19	15

x	m	m^2	f	fm	fm^2
0-10	5	25	12	60	300
10-20	15	225	13	195	2925
20-30	25	625	21	525	13125
30-40	35	1225	19	665	23275
40-50	45	2025	15	675	30375
			80	2120	70,000

$\bar{x} = \Sigma fm / n$

$= \cancel{70,000/80}\ 2120/80$

$\boxed{\bar{x} = 26.5}$

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

$$= \sqrt{\frac{70,000}{80} - \left(\frac{2120}{80}\right)^2}$$

$$= \sqrt{875 - (26.5)^2}$$

$$= \sqrt{875 - 702.25}$$

$$= \sqrt{172.75}$$

$$\boxed{\sigma = 13.1434}$$

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

$$= \frac{13.1434}{26.5} \times 100$$

$$\boxed{CV = 49.5977}$$

From the following price of gold in a week, find the city in which the price was more stable.

Day	Mon	Tues	Wed	Thu	Fri	Sat
City A	498	500	505	504	502	509
City B	500	505	502	498	496	505

x_1	x_1^2	x_2	x_2^2
498	248004	500	250000
500	250000	505	255025
505	255025	502	252004
504	254016	498	248004
502	252004	496	246016
509	259081	505	255025
3018	15,18,130	3006	15,06,074

City A

$\Sigma x_1 = 3018$

$\Sigma x_1^2 = 15,18,130$

$n = 6$

$\bar{x}_1 = \Sigma x_1 / n$

City B

$\Sigma x_2 = 3006$

$\Sigma x_2^2 = 1506074$

$n = 6.$

$\bar{x}_2 = \Sigma x_2 / n$

$= 3018/6$

$$\boxed{\bar{x}_1 = 503}$$

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

$$= \sqrt{\frac{1518130}{6} - \left(\frac{3018}{6}\right)^2}$$

$$= \sqrt{253021.6667 - (503)^2}$$

$$= \sqrt{253021.6667 - 253009}$$

$$= \sqrt{12.6667}$$

$$\boxed{\sigma = 3.5590}$$

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

$$= \frac{3.5590}{503} \times 100$$

$$\boxed{CV = 0.7076}$$

$= 3006/6$

$$\boxed{\bar{x}_2 = 501}$$

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

$$= \sqrt{\frac{1506074}{6} - \left(\frac{3006}{6}\right)^2}$$

$$= \sqrt{251012.3333 - (501)^2}$$

$$= \sqrt{251012.3333 - 251001}$$

$$= \sqrt{11.3333}$$

$$\boxed{\sigma = 3.3665}$$

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

$$= \frac{3.3665}{501} \times 100$$

$$\boxed{CV = 0.6719}$$

CV of Price of city B is less. Hence the price was more stable.

Goals scored by 2 teams A and B
football are observed as follows

No. of match	Score: Team A	Score: Team B
0	5	4
1	7	5
2	5	5
3	3	4
4	2	3
5	3	3

x	x^2	f_1	f_1x	f_1x^2	f_2	f_2x	f_2x^2
0	0	5	0	0	4	0	0
1	1	7	7	7	5	5	5
2	4	5	10	20	5	10	20
3	9	3	9	27	4	12	36
4	16	2	8	32	3	12	48
5	25	3	15	75	3	15	75
		25	49	161	24	54	184.

Team A.

$\bar{x} = \Sigma fx / n$

$= 49/25$

$\boxed{\bar{x} = 1.96}$

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

$= \sqrt{\frac{161}{25} - \left(\frac{49}{25}\right)^2}$

$= \sqrt{6.44 - (1.96)^2}$

$= \sqrt{6.44 - 3.8416}$

$= \sqrt{2.5984}$

$\boxed{\sigma = 1.61}$

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

$= \frac{1.61}{1.96} \times 100$

$\boxed{CV = 82.14}$

Team B

$\bar{x}_2 = \Sigma fx / n$

$= 54/24$

$\boxed{\bar{x}_2 = 2.25}$

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

$= \sqrt{\frac{184}{24} - \left(\frac{54}{24}\right)^2}$

$= \sqrt{7.6667 - (2.25)^2}$

$= \sqrt{7.6667 - 5.0625}$

$= \sqrt{2.6042}$

$\boxed{\sigma = 1.61}$

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

$= \frac{1.61}{2.25} \times 100$

$\boxed{CV = 71.56}$

The marks of B. maths in 2 Classes are given below. find variable.

x	20-30	30-40	40-50	50-60	60-70
Sec A	5	13	24	5	3
Sec B	7	14	25	12	2.

x	m	m^2	f_1	f_1m	f_1m^2	f_2	f_2m	f_2m^2
20 - 30	25	625	5	125	3125	7	175	4375
30 - 40	35	1225	13	455	15925	14	455	17150
40 - 50	45	2025	24	1080	48600	25	1080	50625
50 - 60	55	3025	5	275	15125	12	275	36300
60 - 70	65	4225	3	195	12675	2	195	8450
			50	2130	95450	60	2180	116900

Section A

$\bar{x} = \Sigma fm/n$

$= 2130/50$

$\boxed{\bar{x} = 42.6}$

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

$= \sqrt{\frac{95450}{50} - \left(\frac{2130}{50}\right)^2}$

$= \sqrt{1909 - 42.6}$

$= \sqrt{1909 - 1814.76}$

$= \sqrt{94.24}$

$\boxed{\sigma = 9.7077}$

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

$= \frac{9.7077}{42.6} \times 100$

$\boxed{CV = 22.7880}$

Section B

$\bar{x} = \Sigma fm/n$

$= 2180/60$

$\boxed{\bar{x} = 36.3333}$

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

$= \sqrt{\frac{116900}{60} - \left(\frac{2180}{60}\right)^2}$

$= \sqrt{1948.3333 - (36.3333)^2}$

$= \sqrt{1948.3333 - 1320.1087}$

$= \sqrt{628.2246}$

$\boxed{\sigma = 25.0644}$

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

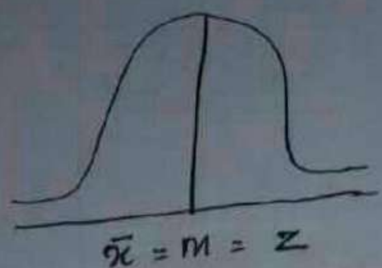
$= \frac{25.0644}{36.3333} \times 100$

$\boxed{CV = 68.9847}$

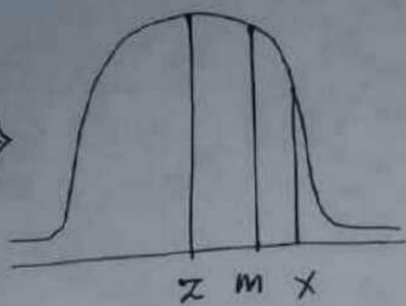
Skewness

Skewness is the degree of asymmetry or departure from symmetry of a distribution.

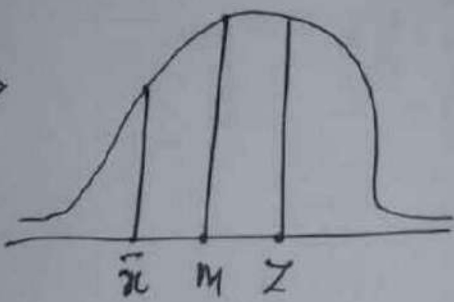
symmetric curve →



positively skewed →



negatively skewed →



Types:

1. Karl Pearson Co-eff of Skewness
2. Bowley " " "
3. Kelly " " "
4. Moment " " "
   (read, beta one)
5. Moment " " "
   (read, gamma one)

Karl Pearson co-eff of skewness.

$$Sk_p = \frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}}$$

(or)

$$\frac{\bar{x} - z}{\sigma}$$

$$Sk_p = \frac{3\,(\text{mean} - \text{median})}{\text{Standard deviation}}$$

(or)

$$= \frac{3\,(\bar{x} - M)}{\sigma}$$

1) from the marks secured by 120 students in section A and 120 students in section B. of class, the following measures are obtained.

Section A: $\bar{x} = 46.83$; SD $= 14.8$; mode $= 51.67$

Section B: $\bar{x} = 47.83$; SD $= 14.8$; mode $= 47.07$

Determain which distribution of marks is more skewed.

Section A:

$$Sk_p = \frac{\bar{x} - z}{\sigma}$$

$$= \frac{46.83 - 51.67}{14.8} = \frac{-4.84}{14.8}$$

$$\boxed{Sk_p = -0.3270}$$

For Section B:

$$Skp = \frac{\bar{x} - Z}{\sigma}$$

$$= \frac{47.83 - 47.07}{14.8}$$

$$= \frac{0.76}{14.8}$$

$$\boxed{= 0.0514}$$

2) From a moderately skewed distribution of retail prices for men's shoes, it is found that the mean price is ₹. 20 and the median price is ₹. 17. If the co-efficient of variation is 20%. find the pearson's co-eff of co. skewness?

$\bar{x} = 20$

$m = 17$

$CV = 20.$

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

$$\cancel{20} = \cancel{\frac{20}{\sigma} \times 100}$$

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

$$20 = \frac{\sigma}{20} \times 100$$

$$\frac{20 \times 20}{100} = \sigma$$

$$\boxed{4 = \sigma}$$

$$Sk_p = \frac{3(\bar{x} - m)}{\sigma}$$

$$= \frac{3(20-17)}{4}$$

$$= \frac{3 \times 3}{4}$$

$$= 9/4$$

$$\boxed{Sk_p = 2.25}$$

3) The sum and the sum of squares of 60 items are 1860 and 67100 respectively. mode is 28.49 find Pearson's Co-efficient of Skewness?

$N = 60$

the sum $\Sigma x = 1860$

the sum of square $\Sigma x^2 = 67100$.

$$Skp = \frac{\bar{x} - z}{\sigma}$$

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

$$= 1860/60$$

$$\boxed{\bar{x} = 31}$$

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

$$= \sqrt{\frac{67100}{60} - \left(\frac{1860}{60}\right)^2}$$

$$= \sqrt{1118.3333 - (31)^2}$$

$$= \sqrt{1118.3333 - 961}$$

$$= \sqrt{157.3333}$$

$$\boxed{\sigma = 12.54}$$

$$Skp = \frac{\bar{x} - Z}{\sigma}$$

$$= \frac{31 - 28.49}{12.54}$$

$$= \frac{2.51}{12.54}$$

$$\boxed{Skp = 0.2002}$$

for a moderately skewed distribution, the mean is 40, the Co-efficient of variation is 5 and Karl pearson Co-efficient of skewness is -0.45. find mode and median.

$\bar{x} = 40$

$CV = 5$

$Skp = -0.45$

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

$$5 = \frac{\sigma}{40} \times 100$$

$$\frac{5 \times 40}{100} = \sigma$$

$$\boxed{2 = \sigma}$$

$$Skp = \frac{\bar{x} - Z}{\sigma}$$

$$-0.45 = \frac{40 - Z}{2}$$

$$-0.45 \times 2 = 40 - Z$$

$$-0.9 = 40 - Z$$

$$-0.9 - 40 = -Z$$

$$\cancel{-}40.9 = \cancel{-}Z$$

$$\boxed{Z = 40.9}$$

$$Z = 3M - 2\bar{x}$$

$$40.9 = 3m - 2(40)$$

$$40.9 = 3m - 80$$

$$40.9 + 80 = 3m$$

$$120.9 = 3m$$

$$\frac{120.9}{3} = m$$

$$\boxed{40.3 = m}$$

OR

$$Skp = \frac{3(\bar{x} - m)}{\sigma}$$

$$-0.45 = \frac{3(40 - m)}{2}$$

$$-0.45 \times 2 = 120 - 3m$$

$$-0.9 = 120 - 3m$$

$$-0.9 - 120 = -3m$$

$$-120.9 = -3m$$

$$\frac{-120.9}{-3} = m$$

$$\boxed{40.3 = m}$$

Solved Ans:

Mode = 40.9

Median = 40.3

Calculate karl Pearson co-eff of skewness
25, 15, 23, 40, 27, 25, 23, 25, 20.

$$Skp = \frac{\bar{x} - Z}{\sigma}$$

x	x^2
25	625
15	225
23	529
40	1600
27	729
25	625
23	529
25	625
30	900
223	5887

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

$= 223/9$

$\boxed{= 24.78}$

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

$$= \sqrt{\frac{5887}{9} - \left(\frac{223}{9}\right)^2}$$

$$= \sqrt{654.1111 - (24.78)^2}$$

$$= \sqrt{654.1111 - 614.0484}$$

$$= \sqrt{40.0627}$$

$\boxed{\sigma = 6.33}$

$\boxed{\text{mode} = 25}$

$$Skp = \frac{\bar{x} - Z}{\sigma}$$

$$= \frac{24.78 - 25}{6.33}$$

$$= \frac{-0.22}{6.33}$$

$\boxed{Skp = -0.0348}$

Calculate Karl Pearson Co-efficient of skewness.

x	12	15	20	25	30	40	50
f	10	25	40	70	32	13	10

x	x^2	f	fx	fx^2	cf
12	144	10	120	1440	10
15	225	25	375	5625	35
20	400	40	800	16000	75
25	625	70	1750	43750	145
30	900	32	960	28800	177
40	1600	13	520	20800	190
50	2500	10	500	25000	200
		200	5025	141415	

$$\bar{x} = \Sigma fx/n$$

$$= 5025/200$$

$$\boxed{\bar{x} = 25.125}$$

$$M = \frac{N+1}{2}$$

$I = 25$

$$= \frac{200+1}{2}$$

OR

$$= 201/2$$

$$= 100.5$$

$$\boxed{M = 25}$$

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

$$= \sqrt{\frac{141415}{200} - \left(\frac{5025}{200}\right)^2}$$

$$= \sqrt{707.075 - (25.125)^2}$$

$$= \sqrt{707.075 - 631.2656}$$

$$= \sqrt{75.8094}$$

$$\boxed{\sigma = 8.7069}$$

$$Sk_p = \frac{\bar{x} - z}{\sigma}$$

$$= \frac{25.125 - 25}{8.7069}$$

$$= \frac{0.125}{8.7069}$$

$$\boxed{Sk_p = 0.0144}$$

(OR)

$$Sk_p = \frac{3(\bar{x} - m)}{\sigma}$$

$$= \frac{3(25.125 - 25)}{8.7069}$$

$$= \frac{3(0.125)}{8.7069}$$

$$= 0.375 / 8.7069$$

$$\boxed{Sk_p = 0.0431}$$

x	m	f	m^2	fm	fm^2
10-20	15	18	225	270	4050
20-30	25	20	625	500	12500
30-40	35	30	1225	1050	36750 ~~42875~~
40-50	45	22	2025	990	44550
50-60	55	10	3025	550	30250
		100		3360	128100

$$\bar{x} = \Sigma fm / n$$

$$= 3360/100$$

$$\boxed{\bar{x} = 33.6}$$

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

$$= \sqrt{\frac{128100}{100} - \left(\frac{3360}{100}\right)^2}$$

$$= \sqrt{1281 - (33.6)^2}$$

$$= \sqrt{1281 - 1128.96}$$

$$= \sqrt{152.04}$$

$$\boxed{\sigma = 12.3305}$$

$$Z = L + \frac{i(D_1)}{D_1 + D_2}$$

$L = 30$, $i = 10$; $f_0 = 20$, $f_1 = 30$, $f_2 = 22$

$D_1 = f_1 - f_0$

$= 30 - 20$

$D_1 = 10$

$D_2 = f_1 - f_2$

$= 30 - 22$

$D_2 = 8$

$$= 30 + \frac{10(10)}{10+8}$$

$$= 30 + \frac{100}{18}$$

$$= 30 + 5.56$$

$$\boxed{Z = 35.56}$$

$$Sk_p = \frac{\bar{x} - Z}{\sigma}$$

$$= \frac{33.6 - 35.56}{12.3305}$$

$$= \frac{-1.96}{12.3305}$$

$$\boxed{Sk_p = -0.1589}$$

Bowley's co-efficient

$$Sk_B = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

calculate Sk_B. 3, 9, 7, 4, 12, 15, 19, 6, 5

3, 4, 5, 6, 7, 9, 12, 15, 19

$Q_1 = \frac{N+1}{4}$

$= 9+1/4$

$= 10/4$

$Q_1 = 2.5$

$\boxed{Q_1 = 5}$

$Q_3 = 3\left(\frac{N+1}{4}\right)$

$= 3\left(\frac{9+1}{4}\right)$

$= 3(10/4)$

$= 30/4$

$= 7.5$

$\boxed{Q_3 = 15}$

$m = \frac{N+1}{2}$

$= \frac{9+1}{2}$

$= 10/2$

$= 5$

$\boxed{m = 7}$

$$Sk_B = \frac{15+5-2(7)}{15-5}$$

$$= \frac{20-14}{10}$$

$$= 6/10$$

$$\boxed{Sk_B = 0.6}$$

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

x	f	cf	
0	7	7	
1	10	17	
2	16	33	Q_1
3	25	58	m
4	18	76	Q_3
5	11	87	
6	8	95	
	95		

$$Q_1 = \frac{N+1}{4}$$

$$= 95+1/4$$

$$= 96/4$$

$$= 24$$

$$\boxed{Q_1 = 2}$$

$$Q_3 = 3\left(N+1/4\right)$$

$$= 3(24)$$

$$= 72$$

$$\boxed{Q_3 = 4}$$

$$m = N+1/2$$

$$= 96/2$$

$$= 48$$

$$\boxed{m = 3}$$

$$Sk_B = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

$$= \frac{4+2-(2\times 3)}{4-2}$$

$$= \frac{6-6}{2}$$

$$= 0/2$$

$$\boxed{Sk_B = 0}$$

x	0 - 200	200 - 250	250 - 300	300 - 350	350 - 400	400 - 450
f	10	25	145	220	70	30

x	f	cf	
0 - 200	10	10	
200 - 250	25	35	
250 - 300	145	180	Q_1
300 - 350	220	400	Q_3 m.
350 - 400	70	470	
400 - 450	30	500	
	500		

$$Q_1 = L + \frac{i(N/4 - cf)}{f}$$

$$= \frac{N+1}{4}$$

$$= \frac{500+1}{4}$$

$$= 501/4.$$

$$= 125.25$$

$L = 250$; $i = 50$; $N/4 = 500/4 = 125$

$cf = 35$; $f = 145$.

$= 250 + \left(\frac{50\,(125-35)}{145}\right)$

$= 250 + \frac{50\,(90)}{145}$

$= 250 + \frac{4500}{145}$

$= 250 + 31.03$

$\boxed{Q_1 = 281.03}$

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

$= 3(N+1/4)$

$= 3(125.25)$

$= 375.75$

$L = 300;\ i = 50\ ;\ cf = 180\ ;\ f = 220$

$3\,(N/4) = 3 \times 125 = 375.$

$= 300 + \frac{50\,(375-180)}{220}$

$= 300 + \frac{50\,(195)}{220}$

$= 300 + 44.32$

$\boxed{Q_3 = 344.32}$

$$m = L + \left[\frac{i\,(N/2 - cf)}{f}\right]$$

$= N+1/2$

$= 500+1/2$

$= 501/2$

$= 250.5$

$L = 300$; $i = 50$; $cf = 180$; $f = 220$; $N/2 = 250$

$= 300 + \dfrac{50\,(250 - 180)}{220}$

$= 300 + \dfrac{50\,(70)}{220}$

$= 300 + \dfrac{3500}{220}$

$= 300 + 15.91$

$\boxed{m = 315.91}$

$$Sk_B = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

$$= \frac{344.32 + 281.03 - (2 \times 315.91)}{344.32 + 281.03}$$

$$= \frac{-6.47}{63.29}$$

$$\boxed{Sk_B = -0.1022}$$

x	0-20	20-50	50-100	100-250	250-500	500-1000
f	20	50	69	30	25	19

x	f	cf	
0-20	20	20	
20-50	50	70	Q_1
50-100	69	139	m
100-250	30	169	Q_3
250-500	25	194	
500-1000	19	213	
	213		

$$Q_1 = l + i\frac{(N/4 - cf)}{cf.}$$

$$= N+1/4.$$

$= 213 + 1/4$

$= 214/4$

$= 53.25$

$L = 20 \; ; \; i = 50 - 20 = 30 \; ; \; f = 50 \; ; \; cf = 20$

$N/4 = 213/4 = 53.25$

$$Q_1 = 20 + \left[\frac{30\,(53.25 - 20)}{50}\right]$$

$$= 20 + \left[\frac{30 \times 33.25}{50}\right]$$

$$= 20 + 19.95$$

$$\boxed{Q_1 = 39.95}$$

$$Q_3 = 3\left(\frac{N+1}{4}\right)$$

$$= 3\,(53.25)$$

$$= 159.75$$

$L = 100 \; ; \; i = 250 - 100 = 250 \; ; \; f = 30 \; ; \; cf = 139$

$$= 100 + \left[\frac{150\,(159.75 - 139)}{30}\right]$$

$$= 100 + \left(\frac{150 \times 20.75}{30}\right)$$

$$= 100 + 103.75$$

$$\boxed{Q_3 = 203.75}$$

$$m = L + i\left(\frac{N/2 - cf}{f}\right)$$

$$= N+1/2$$

$$= 214/2 = 107$$

$L = 50$; $i = 100 - 50 = 50$; $f = 69$; $cf = 70$

$$m = 50 + \left(\frac{50\,(106.5 - 70)}{69}\right)$$

$$= 50 + \left(\frac{50 \times 36.5}{69}\right)$$

$$= 50 + 26.45$$

$$\boxed{m = 76.45}$$

$$Sk_B = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

$$= \frac{203.75 + 39.95 - (2 \times 76.75)}{203.75 - 39.95}$$

$$\boxed{Sk_B = 0.5543}$$

Measures	Place A	Place B
mean	256.5	240.8
median	201	201.6
SD	215.4	181.1
Third Quaetile	260	242
~~fi~~ first Quaetile	157	164.2

Calculate Sk_P and Sk_B.

Place A

$$Sk_p = \frac{3(\bar{x} - m)}{\sigma}$$

$$= \frac{3(256.5 - 201)}{215.4}$$

$$\boxed{Sk_p = 0.7730}$$

$$Sk_B = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

$$= \frac{260 + 157 - (2 \times 201)}{260 - 157}$$

$$= 15/103$$

$$\boxed{Sk_B = 0.1456}$$

Place B

$$Sk_p = \frac{3(\bar{x} - m)}{\sigma}$$

$$= \frac{3(240.8 - 201.6)}{181.1}$$

$$\boxed{Sk_p = 0.6494}$$

$$Sk_B = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

$$= \frac{242 + 164.2 - (2 \times 201.6)}{242 - 164.2}$$

$$= 3/77.8$$

$$\boxed{Sk_B = 0.0386}$$

Calculate Sk_p and Sk_B

x	0	1	2	3	4	5	6
f	2	8	3	4	7	6	5

x	f	fx	cf	x^2	fx^2
0	2	0	2	0	0
1	8	8	10	1	8
2	3	6	13	4	12
3	4	12	17	9	36
4	7	28	24	16	112
5	6	30	30	25	150
6	5	30	35	36	180
	35	114			498

Karl Pearson

$$Sk_p = \frac{\bar{x} - M}{\sigma}$$

$$\bar{x} = \Sigma fx / n$$

$$= 114/35$$

$$\boxed{\bar{x} = 3.26}$$

$$m = N+1/2$$

$$= \frac{114+1}{2} \quad \frac{35+1}{2}$$

$= 115/2 \qquad = 36/2$

$= 57.5 \qquad = 18$

$$\boxed{M = 4}$$

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

$$= \sqrt{\frac{498}{35} - \left(\frac{114}{35}\right)^2}$$

$$= \sqrt{14.23 - (3.26)^2}$$

$$= \sqrt{14.23 - 10.63}$$

$$= \sqrt{3.6}$$

$$\boxed{\sigma = 1.8974}$$

$$Sk_p = \frac{3.26 - 4}{1.8974}$$

$$= \frac{-0.74}{1.8974}$$

$$\boxed{Sk_p = -0.39}$$

$$Sk_B = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

$$= \frac{5 + 1 - 2(4)}{5 - 1}$$

$$= \frac{6 - 8}{4} = -2/4$$

$$\boxed{Sk_B = -0.5}$$

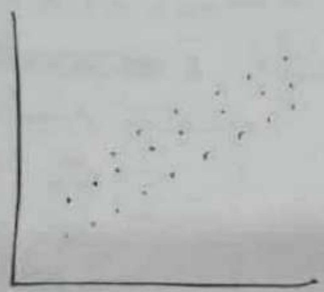
UNIT - 4

CORRELATION

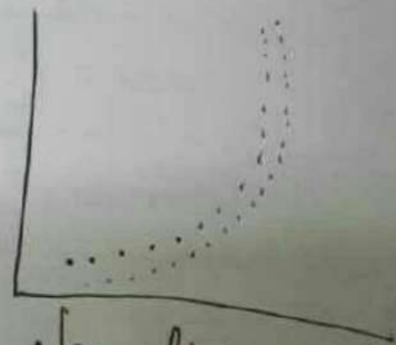
It means relationship between the variables.

Range is b/w -1 to +1.

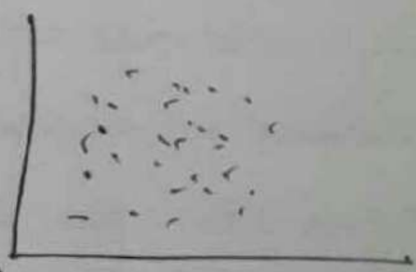
Scatter diagram:



linear Correlation

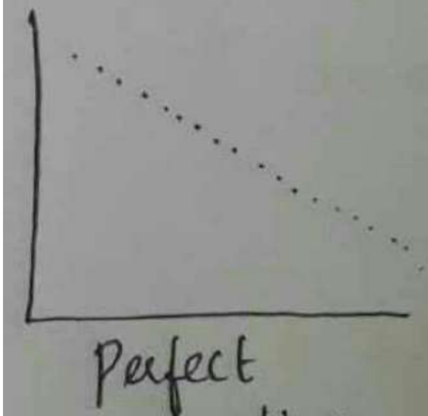


Non - linear Correlation



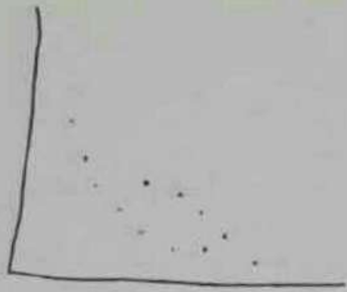
No correlation

Perfect positive Correlation

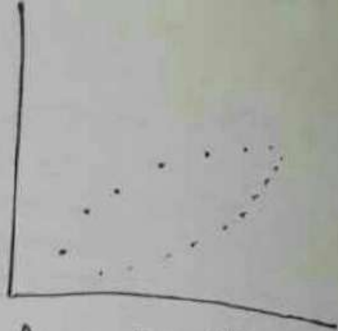


Perfect negative Correlation

High positive Correlation.



High negative Correlation



low positive Correlation

low negative Correlation

Karl pearson co-eff of correlation

$$\boxed{r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\sqrt{N\Sigma y^2 - (\Sigma y)^2}}}$$

(OR)

$$= \frac{N\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{N\Sigma u^2 - (\Sigma u)^2}\sqrt{N\Sigma v^2 - (\Sigma v)^2}}$$

(OR)

$$= \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$

Compute the co-efficient of correlation b/w x - Advertisement expenditure and y - sales.

x	10	12	18	8	13	20	22	15	5	17
y	88	90	94	86	87	92	96	94	88	85

x	y	xy	x^2	y^2
10	88	880	100	7744
12	90	1080	144	8100
18	94	1692	324	8836
8	86	688	64	7396
13	87	1131	169	7569
20	92	1840	400	8464
22	96	2112	484	9216
15	94	1410	225	8836
5	88	440	25	7744
17	85	1445	289	7225
140	900	12718	2224	81130

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{(10 \times 12718) - (140 \times 900)}{\sqrt{10 \times 2224 - (140)^2}\sqrt{(10 \times 81130) - (900)^2}}$$

$$= \frac{1180}{\sqrt{2640}\sqrt{1300}}$$

$$\boxed{r = 0.6370}$$

Calculate Co-eff of Correlation.

x	40	45	47	50	53	60	57	51
y	65	64	70	71	75	83	90	92

x	y	xy	x^2	y^2
40	65	2600	1600	4225
45	64	2880	2025	4096
47	70	3290	2209	4900
50	71	3550	2500	5041
53	75	3975	2809	5625
60	83	4980	3600	6889
57	90	5130	3249	8100
51	92	4692	2601	8464
403	610	31097	20593	47340

$$= \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{8 \times 31097 - (403)(610)}{\sqrt{8 \times 20593 - (403)^2}\sqrt{8 \times 47340 - (610)^2}}$$

$$= \frac{2946}{\sqrt{2335}\sqrt{6620}}$$

$$\boxed{r = 0.7493}$$

---

find co-off of correlation; if $N = 11$; $\Sigma x = 117$; $\Sigma y = 260$; $\Sigma x^2 = 1313$; $\Sigma y^2 = 6580$; $\Sigma xy = 2827$

$$= \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{11 \times 2827 - 117 \times 260}{\sqrt{11 \times 1313 - (117)^2}\sqrt{11 \times 6580 - (260)^2}}$$

$$= \frac{677}{\sqrt{754}\sqrt{4780}}$$

$$\boxed{r = 0.3566}$$

Spearman's Rank Correlation Co-efficient

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

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

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

x	y	d	d^2
1	6	-5	25
6	8	-2	4
3	3	0	0
9	7	2	4
5	2	3	9
2	1	1	1
7	5	2	4
10	9	1	1
8	4	4	16
4	10	-6	36
			100

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

$$= 1 - \left[\frac{6 \times 100}{10(10^2-1)}\right]$$

$$= 1 - \left[\frac{600}{10(99)}\right]$$

$$= 1 - \left[\frac{600}{990}\right]$$

$$= 1 - 0.6061$$

$$\boxed{\rho = 0.3939}$$

x 21 36 42

y 47 40 37 42 43

X	Y	x	y	d	d^2
21	47	5	1	4	16
36	40	3	4	-1	1
42	37	1	5	-4	16
37	42	2	3	-1	1
25	43	4	2	2	4
					38

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

$$= 1 - \left[\frac{6 \times 38}{5(5^2-1)}\right]$$

$$= 1 - \left[\frac{6 \times 38}{5 \times 24}\right]$$

$$= 1 - \left[\frac{228}{120}\right]$$

$$= 1 - 1.9$$

$$\boxed{\rho = -0.9}$$

x	50	60	65	70	75	40	70	80
y	80	71	60	75	90	82	70	50

X	Y	x	y	d	d^2
50	80	7	3	4	16
60	71	6	5	1	1
65	60	5	7	-2	4
70	75	3.5	4	-0.5	0.25
75	90	2	1	1	1
40	82	8	2	6	36
70	70	3.5	6	-2.5	6.25
80	50	1	8	7	49
					113.5

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

$$= 1 - \left[\frac{(6)(113.5) + \frac{2(2^2-1)}{12}}{8(8^2-1)}\right]$$

$$= 1 - \left[\frac{6(113.5 + 0.5)}{8 \times 63}\right]$$

$$= 1 - 1.3571$$

$$\boxed{P = -0.3571}$$

x	15	20	28	12	40	60	20	80
y	40	30	50	30	20	10	30	60

x	y	x	y	d	d^2
15	40	7	3	4	16
20	30	5.5	5.	0.5	0.25
28	50	4	2	2	4
12	30	8	5	3	9
40	20	3	7	−4	16
60	10	2	8	−6	36
20	30	5.5	5.	0.5	0.25
80	60	1	1	6	0
					81.5

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

$$= 1 - \left[\frac{6\left\{81.5 + \frac{2(2^2-1)}{12} + \frac{3(3^2-1)}{12}\right\}}{8(8^2-1)}\right]$$

$$= 1 - \left[\frac{6\{81.5 + 0.5 + 2\}}{8\ (63)}\right]$$

$$= 1 - \left[\frac{\cancel{6}^{2} \times \cancel{84}^{4}}{\cancel{8}_{2} \times \cancel{63}_{21}}\right]$$

$$= 1 - 1$$

$$\boxed{\rho = 0}$$

Concurrent deviation method

$$r_c = \pm\sqrt{\pm\left(\frac{2C-N}{N}\right)}$$

Calculate co-eff of correlation

yr	1959	60	61	62	63	64	65	66	67	68	69
import	85	82	89	95	104	108	112	100	99	93	96
price	110	115	112	118	120	109	98	102	130	105	107

x	y	D_x	D_y	D_{xy}
85	100			
82	115	−	+	−
89	112	+	−	−
95	118	+	+	+
104	120	+	+	+
108	109	+	−	−
112	98	+	−	−
100	102	−	+	−
99	103	−	+	−
93	105	−	+	−
96	107	−	+	−

$N = 10$

$C = 2$

$$r_c = \sqrt{\frac{2C-N}{N}}$$

$$= \sqrt{\frac{2(2)-10}{10}}$$

$$= \sqrt{\frac{4-10}{10}}$$

$$= -\sqrt{\frac{6}{10}}$$

$$\boxed{= -0.7746}$$

x	17	12	25	41	32	51
y	12	15	23	32	28	26

x	y	D_x	D_y	D_{xy}
17	12			
12	15	−	+	−
25	23	+	+	+
41	32	+	+	+
32	28	−	−	+
51	26	+	−	−

$n = 5$

$c = 3$

$$r_c = \sqrt{\frac{2C - N}{N}}$$

$$= \sqrt{\frac{(2 \times 3) - 5}{5}} = \sqrt{\frac{6 - 1}{5}}$$

$$= \sqrt{\frac{1}{5}} = \boxed{0.4472}$$

Karl Pearson co-eff Correlation

x	-3	-2	-1	0	1	2	3
y	9	4	1	0	1	4	9

x	y	xy	x^2	y^2
-3	9	-27	9	81
-2	4	-8	4	16
-1	1	-1	1	1
0	0	0	0	0
1	1	1	1	1
2	4	8	4	16
3	9	27	9	81
0	28	0	28	196

$$= \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{(7\times 0) - (0\times 28)}{\sqrt{7\times 28 - (0)^2}\sqrt{7\times 196 - (28)^2}}$$

$$= \frac{0}{\sqrt{152}\sqrt{1372}}$$

$$\boxed{r = 0}$$

Regression

y on x

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

$$b_{yx} = \frac{r\sigma_y}{\sigma_x}$$

$$= \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\Sigma x^2 - (\Sigma x)^2}$$

x on y

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

$$b_{xy} = \frac{r\sigma_x}{\sigma_y}$$

$$= \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\Sigma y^2 - (\Sigma y)^2}$$

1) from the following information on variables of 2 values of x and y. find 2 regression lines and the correlation co-efficient.

$N = 10$; $\Sigma x = 20$; $\Sigma y = 40$; $\Sigma x^2 = 240$; $\Sigma y^2 = 410$; $\Sigma xy = 200$.

y on x:

$$\bar{y} = \Sigma y/n$$
$$= 40/10$$
$$\bar{y} = 4$$

$$\bar{x} = \Sigma x/n$$
$$= 20/10$$
$$\bar{x} = 2$$

$$b_{yx} = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{10(200) - (20)(40)}{10(240) - (20)^2}$$

$$= \frac{2000 - 800}{2400 - 400}$$

$$= \frac{1200}{2000}$$

$$= 0.6$$

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

$$y - 4 = 0.6 (x - 2)$$

$$y - 4 = 0.6x - 1.2$$

$$y = 0.6x - 1.2 + 4$$

$$\boxed{y = 0.6x + 2.8} \quad \text{——— } (1).$$

<u>x on y</u>:

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

$$\cancel{x - 2 =}$$

$$b_{xy} = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{N\Sigma y^2 - (\Sigma y)^2}$$

$$= \frac{10(200) - (20)(40)}{10(410) - (40)^2}$$

$$= \frac{2000 - 800}{4100 - 1600}$$

$$= \frac{1200}{2500}$$

$$= 0.48$$

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

$$x - 2 = 0.48 (y - 4).$$

$= \quad x = 0.48y - 1.4$

$$x = 0.48y + 0.08 \qquad \text{--- (2)}$$

Correlation:

$$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2}\ \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{1200}{\sqrt{2000}\ \sqrt{2500}}$$

$$\boxed{r = 0.5367}$$

(OR)

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

$$= \pm\sqrt{0.48 \times 0.6}$$

$$\boxed{r = 0.5367}$$

Calculate 2 regression equation

x	10	12	13	12	16	15
y	40	38	43	45	37	43

x	y	xy	x^2	y^2
10	40	400	100	1600
12	38	456	144	1444
13	43	559	169	1849
12	45	540	144	2025
16	37	592	256	1369
15	43	645	225	1849
78	246	3192	1038	10136

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

$= 78/6$

$\boxed{\bar{x} = 13}$

$\bar{y} = \Sigma y / n$

$= 246/6$

$\boxed{\bar{y} = 41}$

$$b_{xy} = \frac{N \Sigma xy - (\Sigma x)(\Sigma y)}{N \Sigma y^2 - (\Sigma y)^2}$$

$$= \frac{6 \times 3192 - (78)(246)}{6 \times 10136 - (246)^2}$$

$$= \frac{19152 - 19188}{60816 - 60516}$$

$$= -36/300$$

$$\boxed{b_{xy} = -0.12}$$

$$b_{yx} = \frac{N \Sigma xy - (\Sigma x)(\Sigma y)}{N \Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{\quad}{6 \times 1038 - (78)^2}$$

$$= \frac{-36}{6228 - 6084}$$

$$= \frac{-36}{144}$$

$$\boxed{b_{yx} = -0.25}$$

<u>y on x:</u>

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

$$y - 41 = -0.25\ (x - 13)$$

$$y - 41 = -0.25x + 3.25$$

$$y = -0.25x + 3.25 + 41$$

$$\boxed{y = -0.25x + 44.25} \quad \text{—— (1)}$$

<u>x on y:</u>

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

$$x - 13 = -0.12\ (y - 41)$$

$$x - 13 = -0.12y + 4.92$$

$$x = -0.12y + 4.92 + 13$$

$$\boxed{x = -0.12y + 17.92} \quad \text{—— (2)}$$

UNIT 5

Analysis of time series

Long term ↓ Secular trend

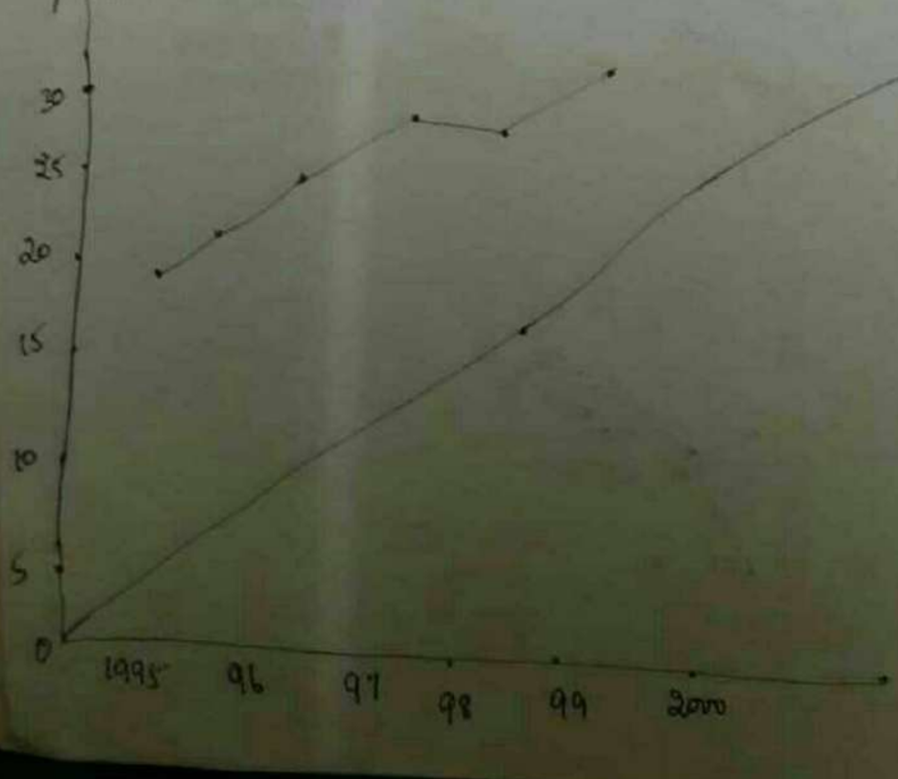
Short term ↓ Seasonal fluctuation, cyclical ", irregular "

① Secular trend:

1. Graphic method
2. Semi-Average "
3. Moving " "
4. Least square "

ⓐ Graphic method

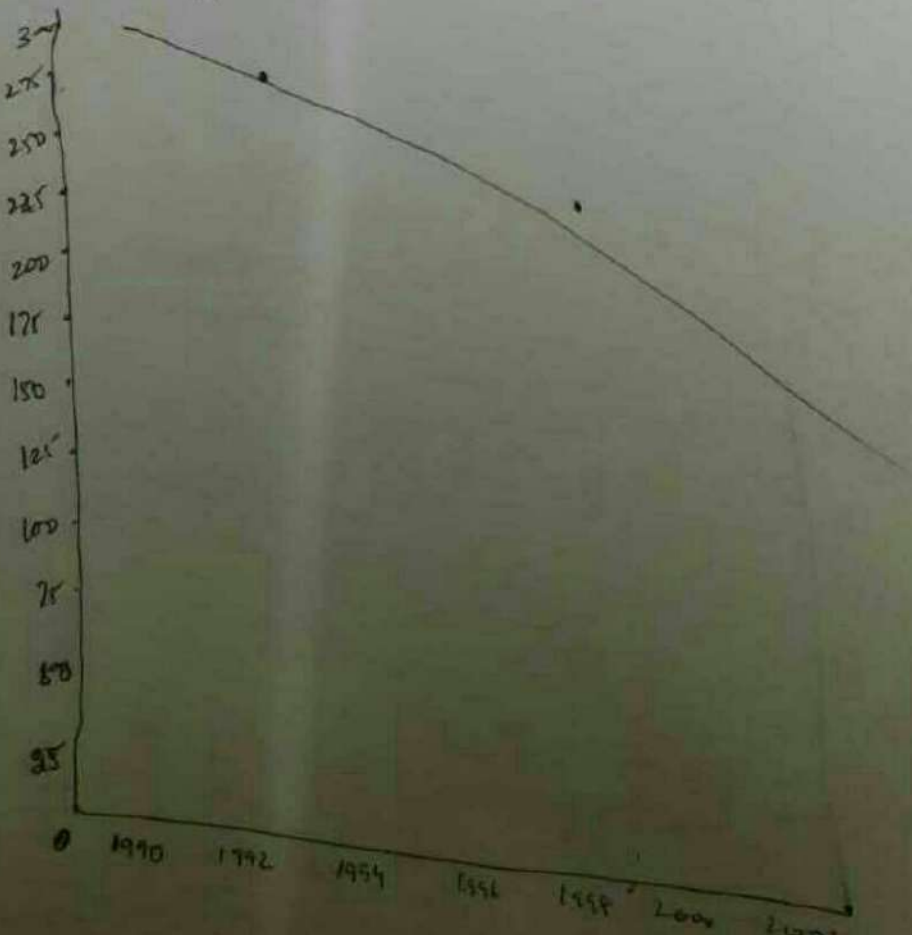
Year	1995	96	97	98	99	2000	2001
Production	20	22	25	26	25	27	30



(b) Method of Semi - Average

Sales is from 1990 to 2001 are as follows 280, 300, 280, 280, 270, 240, 230, 230, 220, 200, 210, 200

year	Sales	middle yr	mean score
1990	280		
1991	300		
92	280		
93	280	1992.5	1650/6 = 275
94	270		
95	240		
96	230		
97	230		
98	220	1998.5	1290/6 = 215
99	200		
'00	210		
'01	200		



2) yr 1987 88 89 90 91 92 93

Sale 90 110 130 150 100 150 200

yr	Sale	middle yr	mean
1987	90		
88	110	1988	$330/3 = 110$
89	130		
90	150		
91	100		
92	150	1992	$450/3 = 150$
93	200		

(E) Moving average.

yr	87	88	89	90	91	92	93	94	95	96
Sale	332	311	357	392	402	405	410	427	405	438

find 5 yr moving avg.

yr	Sale	5 yr moving total	5 yr moving avg.
1987	332	-	-
88	311	-	-
89	357	1794	358.8
90	392	1867	373.4
91	402	1966	393.2
92	405	2036	407.2
93	410	2049	409.8
94	427	2085	417.
95	405	-	
96	438	-	

2)

yr	1983	84	85	86	87	88	89	90	91	92
Production	21	22	23	25	24	22	25	26	27	26

find 3 yr moving avg.

yr	Production	3 yr moving total	3 yr moving avg
1983	21	-	-
84	22	66	22
85	23	70	23.33
86	25	72	24
87	24	71	23.67
88	22	71	23.67
89	25	73	24.33
90	26	78	26
91	27	79	26.33
92	26	-	—

3) 4 yr moving avg.

yr	production	4 yr moving total	2 period moving total	4 yr moving avg
1981	464			
82	515	1964		
83	518	2002	3966	495.75
84	467	2027	4029	503.63
85	502	2066	4093	511.63
86	540	2170	4236	529.5
87	557	2254	4424	553
88	571	2326	4580	572.
89	586			
90	612			

4) 6 yr moving avg.

yr	Sale	6 yr moving total	2 period moving total	6 yr moving avg
1985	10			
86	12			
87	13	78		
88	15	84	162	13.5
89	14	90	174	14.5
90	14	99	189	15.75
91	16	108	207	17.25
92	18	120	228	19
93	22	135	255	21.25
94	24	144	279	23.25
95	26	147	291	24.25
96	25	150	297	24.75
97	25	153	303	25.25
98	21			
99	25			
00	27			

Method of least square:

$$\Sigma y = NA + B\Sigma x$$

$$\Sigma xy = N\Sigma x + B\Sigma x^2$$

$$y = a + bx$$

yr	1979	1980	1981	82	83
Sale	100	120	140	160	180

estimate the value for 1985.

yr (X)	Sale (y)	x	xy	x^2	y_t
1979	100	-2	-200	4	100
80	120	-1	-120	1	120
$\boxed{81}$	140	0	0	0	140
82	160	1	160	1	160
83	180	2	360	4	180
	700	0	200	10	700

$$\Sigma y = NA + B\Sigma x$$

$$= 5A + B(0)$$

$$\Sigma y = 5A + 0B \quad \text{——} \quad (1)$$

$$\Sigma xy = N\Sigma x + B\Sigma x^2$$

$$= 50 + B10$$

$$\Sigma xy = 0 + 10B \quad \text{——} \quad (2)$$

$\therefore$ $5A + 0B = 700$

$0 + 10B = 200$

$A = 140$

$B = 20$

$y = a + bx$

$(x = 1979)$

$= y = 140 + 20(-2)$

$= 140 - 40$

$\boxed{y = 100}$

$(x = 1980)$

$y = 140 + 20(-1)$

$= 140 - 20$

$\boxed{y = 120}$

$(x = 1981)$

$y = 140 + 20(0)$

$= 140 + 0$

$\boxed{y = 140}$

$(x = 1982)$

$y = 140 + 20(1)$

$= 140 + 20$

$\boxed{y = 160}$

$y = 140 + (20 \times 2)$

$= 140 + 40$

$= 180$

$(x = 1985)$

$y = 140 + 20(4)$

$= 140 + 80$

$\boxed{y = 220}$

Index Number

$$P = \frac{P_1}{P_0} \times 100$$

$$Q = \frac{Q_1}{Q_0} \times 100$$

P_0 / Q_0 = Base year

P_1 / Q_1 = Current year

P - Price

Q. Quantity

I

1.1 IN

leader

sustai

peopl

easy

solu

equ

coı

re

g

Methods

- unweighted (simple)
  - Simple aggregatives method
  - simple averages of relatives method
- weighted
  - weighted aggressive method
  - weighted averages of relatives method

Simple agressive method:

$$P_{01} = \frac{\Sigma P_1}{\Sigma P_0} \times 100$$

$$Q_{01} = \frac{\Sigma q_1}{\Sigma q_0} \times 100$$

simple averages of relatives method:

$$Am\ P_{01} = \Sigma P / n$$

$$GM\ P_{01} = \text{Antilog}\left(\frac{\Sigma \log P}{N}\right)$$

1) From the following data construct an index for 1995 taking 1994 as base

1994	50	40	80	110	20
1995	70	60	90	120	20

P_0	P_1	$P = P_1/P_0 \times 100$	$\log P$
50	70	140	2.1461
40	60	150	2.1761
80	90	112.5	2.0512
110	120	109.09	2.0378
20	20	100	2
300	360	611.59	10.4112

$$P_{01} = \frac{\Sigma P_1}{\Sigma P_0} \times 100$$

$$= 360/300 \times 100$$

$$= 120$$

using Am:

$$P_{01} = \Sigma P/N$$

$$= \frac{611.59}{5}$$

$$= 122.32$$

using Gm

$$= \text{Antilog}\left(\frac{\Sigma \log P}{N}\right)$$

$$= \text{Antilog}\left(\frac{10.4112}{5}\right)$$

$$= \text{Antilog}(2.0823)$$

$$= 120.84$$

weighted average method.

(1) Laspeyre's $P_{01}L$

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

(2) Paasche's $P_{01}P$

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

(3) fisher's $P_{01}F$

$$= \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$= \sqrt{P_{01}L \cdot P_{01}P}$$

(4) marshall - edgeworth

$$P_{01}^{ME} = \frac{\Sigma p_1 (q_0 + q_1)}{\Sigma p_0 (q_0 + q_1)} \times 100$$

$$= \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$$

(5) Bowley $P_{01}B$

$$\frac{P_{01}L + P_{01}P}{2}$$

(6) Kelly $P_{01}k$

$$\frac{\Sigma P_1 q}{\Sigma P_0 q} \times 100$$

	2000		2001	
Commodity	price	Quantity	price	quantity
A	2	74	3	82
B	5	125	4	140
C	7	40	6	33

find weighted average method

Commodity	P_0	q_0	P_1	q_1	P_0q_0	P_1q_0	P_0q_1	P_1q_1
A	2	74	3	82	148	222	164	246
B	5	125	4	140	625	500	700	560
C	7	40	6	33	280	240	231	198
					1053	962	1095	1004

$$P_{01}L = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100$$

$$= \frac{962}{1053} \times 100$$

$$\boxed{P_{01}L = 91.3580}$$

(2) $P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

$= \frac{1004}{1095} \times 100$

$\boxed{P_{01}P = 91.6895}$

(3) $P_{01}F = \sqrt{P_{01}L \cdot P_{01}P}$ ~~x100~~

$= \sqrt{91.3580 \times 91.6895}$

$\boxed{P_{01}F = 91.524}$

(4) Marshal

$P_{01}ME = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$

$= \frac{962 + 1004}{1053 + 1095} \times 100$

$\boxed{P_{01}ME = 91.527}$

(5) Bowley

$P_{01}B = \frac{P_{01}L + P_{01}P}{2}$

$= \frac{91.3580 + 91.6895}{2}$

$\boxed{P_{01}B = 91.5238}$