**Syllabus**

|  |  |
| --- | --- |
| UNIT – I | Origin, Meaning, Scope and Limitations of Statistics, Relationship with business and industry. |
| UNIT – II | Collection of data-Collection, Classifications and Tabulations of Statistical data, Pie diagrams, Graphic Representation. |
| UNIT – III | Measure of central tendency-Mean, Median and Mode-Meaning and computation. Measure of dispersion standard deviation coefficient of variation. |
| UNIT – IV | Simple Correlation and regression- Meaning, Karl Pearson correlation, Rank correlation, computations, uses, Regression Equations- Index Numbers- Concepts and uses, constructions index numbers, limitations, Test of Index numbers. |
| UNIT – V | Probability Theory-Basic concepts in probability. Statistical dependence and independence, Bayes theorem. |

**UNIT — I**

**STATISTICS**

The word “Statistics” of English language has either been derived from the Latin word status or Italian word statistics and meaning of this term is “An organised political state.

**Meaning:** The science of collecting, analysing and interpreting such data or Numerical data relating to an aggregate of individuals.

E.g:- Statistics of National Income, Statistics of Automobile Accidents, Production Statistics, etc.

**Definition:** - “The classified facts relating the condition of the people in a state specially those facts which can be stated in members or in tables of members or in any tabular or classified arrangements.”

**-Webster**

“Statistics may be regarded as (i) the study of population (ii) The study of variation (iii) The study of method of reduction of data”

**-R.A. Fisher.**

**Nature /Features /Characteristics of statistics**

* It is an aggregate of facts.
* Analysis of multiplicity of causes.
* It is numerically expressed.
* It is estimated according to reasonable standard of accuracy.
* It is collected for pre-determined purpose.
* It is collected in a systematic manner.

**Division of Statistics**

Theoretical Statistical Methods Applied

**Theoretical:** Mathematical theory which is the basis of the science of statistics is called theoretical statistics.

**Statistical Methods:** By this method we mean methods specially adapted to the elucidation of quantitative data affected by a multiplicity of causes.

**Few Methods are:-**

1. Collection of Data (2) Classification (3) Tabulation (4) Presentation (5) Analysis

(6) Interpretation (7) Forecasting.

**Applied: -** It deals with the application of rules and principles developed for specific problem in different disciplines.

**Eg: -** Time series, Sampling, Statistical Quality control, and design of experiments.

**Functions of Statistics:**-

* It presents facts in a definite form.
* It simplifies mass of figures
* It facilitates comparison
* It helps in prediction
* It helps in formulating suitable & policies.

**Scope of Statistics:-**

1. Statistics and state or govt.
2. Statistics and business or management.

* Marketing
* Production
* Finance
* Banking
* Control
* Research and Development
* Purchases

1. Statistics and Economics
   * Measures National Income
   * Money Market analysis
   * Analysis of competition, monopoly, oligopoly,
   * Analysis of Population etc.
2. Statistics and science
3. Statistics and Research

**Limitations:-**

1. It is not deal with items but deals with aggregates.
2. Only on expert can use it
3. It is not the only method to analyze the problem.
4. It can be misused etc.

**Statistical Investigation**

**Meaning:** In general it means as a statistical survey.

In brief. Scientific and systematic collection of data and their analysis with the help of various statistical methods and their interpretation.

**Stages of Statistical Investigation:-**

* Planning of Investigation
* Collection of Data
* Editing of Data
* Presentation of Data

1. Classification
2. Tabulation
3. Diagrams
4. Graphs
   * Analysis of Data
   * Interrelation of Data or Report Preparation

**Types of Statistical Investigation:-**

1. Experiment or survey investigation
2. Complete or sample investigation
3. Official, semi-official, Non official investigation
4. Confidential or open investigation
5. General purpose and specific purpose investigation
6. Original or repetitive investigation.

**Relationship of Statistics with Business and Industry**

The use of statistical methods in solution of business problems is very important. It has a role to play every area of activity in business and industry such as production, financial analysis, market research, man power planning and accounting.

Due to growing size and increasing competition in the market, the activities of most of the business enterprises have crossed the local and national market to international market. In this situation, it is very difficult for a single person to look after all the business activities like sales,

purchase, production, marketing, control, finance, advertisement, etc.

These business activities are managed by specialized and efficient management team.

For example, suppose a businessman wants to manufacture starting with the production process, he must have an overall idea as to ‘how many garments are to be manufactured’, how much raw material, labour and amount is needed for that’, what is the quality, shape, colour, size, etc. of the garments to be manufactured’, and ‘which types of garments are in fashion’, etc.

For this purpose, the role of statistical data and statistical tools like probability, expectation, sampling procedures theory of estimation and theory of forecasting and many others are of highest importance. A successful businessman analyzes the present position and future impacts of seasonal and cyclical variations, consumer’s testes, customs and fashion, standard of living, purchasing power, etc. For deciding the sales and marketing strategies of the firm and the analysis of demand of the commodities, accurate and sufficient statistical data are required. The time series, demand analysis, index numbers, business cycles, which are supposed to be economic barometers are prepared y accurate statistical data, are of great importance for business. According to **Boddingtons,** ‘The successful businessman is the one whose estimates most closely approach to accuracy”.

Following are some points which indicate that how statistics help, in taking management decisions:

**1. Marketing:** Statistical analysis is frequency used in providing information for marketing decisions. In the field of marketing, it is necessary first to find out what can be sold and then to evolve a suitable strategy so consumer. A skillful analysis of data on population, purchasing power, habits of people, competition, transportation cost, etc. should precede any attempt to establish a new market.

Analysis of sales in relation to distribution and maintenance of accurate record is specially important for making sales strategy.

**2. Production:** In the field of production, statistical data and statistical methods play a very important role. The decision about what to produce, how much to produce, when to produce, for whom to produce, is based largely on facts analyzed statistically.

Statistical tools are helpful in quality control and dealing with labour problems.

**3. Finance:** The financial Managers in discharging their finance functions efficiently depend heavily on statistical analysis of facts and figures. Financial forecasts, break-even analysis and investment decisions are under uncertainty but these are also then part of their management activities. In the last three decades, sophisticated models dealing with inventories, cash balances and so on have been developed and applied. These models involve application of several statistical concepts. The area of security analysis is also highly quantitative.

**4. Investment:** Very clear and valuable judgments are to be taken in the field of investment. The investor is required to determine whether to buy or to sale. With the help of statistical data, investor can build fortunes with relatively less risk. The investor must distinguish between good and bad securities. For this purpose, he should have a clear understanding of the present situation of money market and various industries. It is possible only with statistical data.

**5. Quality Control:** The quality of a product can be judged by statistical methods. There are two broad types of controls: (i) Process control and (ii) Product control.

The process control can be done by comparing the product sample with standard measures. It will show the quality in conformity with the standards.

**6. Management Decisions:** For better management decision, the consumer preference studies are very useful because they help in forecasting future sales. The decisions relating to dropping an existing product or introducing a new product depend on the statistical results. Similarly for pricing decisions, the reference to the competing brands is also significant.

Management of big industries today is required to take various decisions about purchasing, credit control, recruitment, research and development. For this purpose, statistical analysis is of great value. Many big organizations have research and development departments which are primarily concerned with out how existing products can be improved and how the optimum use of resources can be made. In the absences of statistical data, it is almost impossible to carry out fruitful research development programmes.

**UNIT-II**

**PROCESS OF DATA COLLECTION**

**Data:** - A bundle of Information or bunch of information.

**Data Collection:** Collecting Information for some relevant purpose & placed in relation to each other.

**Types of Data:-**

1. **Raw Data:-** When we collect data through schedules and questionnaires or some other method eg:- Classification, tabulation etc.
2. **Processed Data: -** When we use the above raw data for application of different methods of analyzing of data. Like using correlation, Z-test, T-test on data. That will be known as processed data.

**Sources of Data Collection:-**

1. **Internal Data: -** When data is collected by problem the internal source for any specific

It purpose.

1. **External Data: -** This type of data collected by the external source.
2. **Primary Data: -** It is original and collected first time. it is like raw material and it is required large sum of money, energy and time.
3. **Secondary Data: -** Secondary data are those already in existence and which have been collected for some other purpose than answering of the question at hand.
4. **Qualitative Data: -** Which cannot be measurable but only their presence and absence in a group of individual can be noted are called qualitative data.
5. **Quantitative Data: -** The characteristics which can be measured directly are known as quantitative data.

**Collection of Data: -** It means the methods that are to be employed for obtaining the required information from the units under investigations.

**Methods of Data Collection:- (Primary Data)**

* Direct Personal Interviews
* By observation
* By Survey
* By questionnaires

**Difference between Primary and secondary data:-**

|  |  |  |
| --- | --- | --- |
| **Points** | **Primary Data** | **Secondary Data** |
| 1. **Originality** | Primary data are original i.e., collected first time. | Secondary data are not original, i.e.., they are already in existence and are used by the investigator. |
| 1. **Organisation** | Primary data are like raw material. | Secondary data are in the form of finished product. They have passed through statistical methods. |
| 1. **Purpose** | Primary data are according to the object of investigation and are used without correction. | Secondary data are collected for some other purpose and are corrected before use. |
| 1. **Expenditure** | The collection of primary data requires large sum, energy and time. | Secondary data are easily available from secondary sources (published or unpublished). |
| 1. **Precautions** | Precautions are not necessary in the use of primary data. | Precautions are necessary in the use of secondary data. |

**Preparation of Questionnaires:-**

This method of data collection is quite popular, particularly in case of big enquires, it is adopted by individuals, research workers. Private and public organization and even by government also.

A questionnaire consists of number of question printed or type in a definite order on a form or set of forms. The respondents have to answer the question on their own.

**Importance:-**

1. Low cost and universal
2. Free from biases.
3. Respondents have adequate time to respond
4. Fairly approachable

**Demerits:-**

(i) Low rate of return

(ii) Fill on educated respondents

(iii) Slowest method of Response

**Preparation of Questionnaires: -** It is considered as the heart of a survey operation. Hence it should be very carefully constructed. If it is not properly set up and carefully constructed.

|  |  |  |
| --- | --- | --- |
| Step I | :- | Prepare it in a general form. |
| Step II | :- | Prepare sequence of question. |
| Step III | :- | Emphasize on question formulation and wordings |
| Step IV | :- | Ask Logical and not misleading questions. |
| Step V | :- | Personal questions should be left to the end. |
| Step VI | :- | Technical terms and vague expressions should be availed classification and Tabulation of Data |

**Classification & Tabulation of Data**

After collecting and editing of data an important step towards processing that classification. It is grouping of related facts into different classes.

**Types of classification:-**

1. **Geographical: -** On the basis of location difference between the various items. E.g. Sugar Cave, wheat, rice, for various states.
2. **Chronological:-** On the basis of time

e.g.-

|  |  |
| --- | --- |
| **Year** | **Sales** |
| 1997 | 1,84,408 |
| 1998 | 1,84,400 |
| 1999 | 1,05,000 |

1. **Qualitative classification: -** Data classified on the basis of some attribute or quality such as, color of hair, literacy, religion etc.

**Population**

1. **Quantitative Classification: -** When data is quantify on some units like height, weight, income, sales etc.

**Tabulation of Data**

A table is a systematic arrangement of statistical data in columns and Rows.

**Part of Table:-**

1. Table number
2. Title of the Table
3. Caption
4. Stub
5. Body of the table
6. Head note
7. Foot Note

**Types of Table:-**

(i) Simple and Complex Table:-

**(a) Simple or one-way table:-**

|  |  |
| --- | --- |
| **Age** | **No. of Employees** |
| 25 | 10 |
| 30 | 7 |
| 35 | 12 |
| 40 | 9 |
| 45 | 6 |

**(b) Two way Table**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **Males** | **Females** | **Total** |
| 25 | 25 | 15 | 40 |
| 30 | 20 | 25 | 45 |
| 35 | 24 | 20 | 44 |
| 40 | 18 | 10 | 28 |
| 45 | 10 | 8 | 18 |
| **Total** | **97** | **78** | **175** |

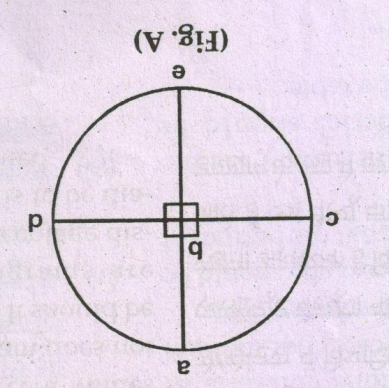
**2) General Purpose and Specific Purpose Table:-** General purpose table, also known as the reference table or repository tables, which provides information for general use or reference.

Special purpose are also known as summary or analytical tables which provides information for one particular discussion or specific purpose.

**PIE DIAGRAM**

**Meaning of Pie Diagram**

Pie Diagram is also known as circular diagram. Pie Diagram is circular divided into various segments showing percent values of a series. This diagram does not shown absolute values. It should be noted that Pie Diagrams are drawn only when Percentage distribution of the value is to be diagrammatically presented.



**Construction of Pie Diagram**

Construction of a Pie Diagram involves the following steps:

1. Under the first step, absolute values of the series are converted into percentage values. **For example:** If a class of 55 students, 15 students are first divisioners, 209 students are second divisioners and 20 students are third divisioners, then in terms of percent values it would be  are first divisioners,  are second divisioners; and  are third divisioners.
2. Under the second step, a circle is drawn. Only 4 angles of 900 can be drawn within a circle, so that it comprises 900 × 4 = 3600

[See. fig.A] ∠abc + ∠abd + ∠ebd + ebc = 3600

900 + 900 + 900 + 900 = 3600

1. Under the third step, divide 3600 in the proportion of 27.27%, 36.36% and 36.36%. We get.

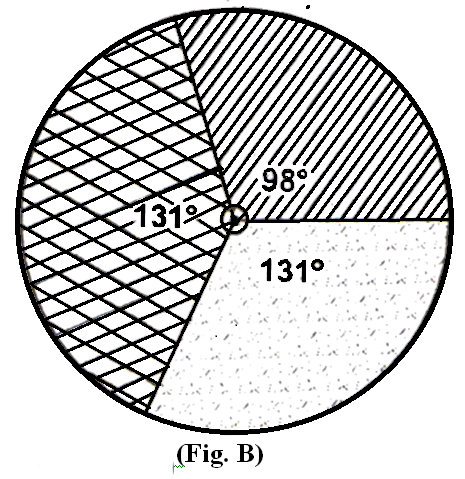






**Test:** 980 + 1310 + 1310 = 3600

Accordingly 27.27% of the first divisoners would mean 980 out of 3600, 36.36% of the second divisoners would mean 1310 out of 3600 and 36.36% of the third divisoners would mean 1310 out of 3600 [See Fig.B].

****

**Examples:**

In 2015-16, Net Domestic Product by industry of origin (at 2010-2011) is as given below. Present this information in the form of a pie Diagram.

|  |  |
| --- | --- |
| **Sectors** | **% Share** |
| Primary  Secondary  Transport  Administration  Banking and Finance | 20  23  25  14  18 |
| **Total** | **100** |

**GRAPHICAL REPRESENTATION OF DATA**

**Graphical Presentation of Data**

One important function of statistics is to simplify the complexity of quantitative data and to make them easily understandable.

One more important method of making the data understandable is to represent them by way of graphs. Figures are generally avoided by the common man but pictures and diagrams always attract and impress him. Diagrams give a bird’s eye-view of the whole mass of statistical data.

In the words of **M.J. Moroney,** “Cold figures are uninspiring to most people. Diagrams help us to see the pattern and shape of any complex situation. Just as a map wide stretch of a country, so diagram help us visualise the whole meaning of a numerical complex at a single glance”.  
**Advantages of Graphic Presentation**

Graphic presentation of data has several advantages over the tabular presentation of data. The major advantages of graphic presentation of data are as follows:

1. Graphs are always attractive and impressive than table of figures. In graph, all the features of data become visible at a glance, which may help in studying the trends and function of data accurately.
2. It is the simplest method of presenting statistical data.
3. It saves time and energy of both, the statistician as well as the observers.
4. There is no much mathematical knowledge is required to understand the graphs.
5. It is possible to determine the positional averages like median, mode, etc. with the help of graph.
6. By observing the graphs of the data, one can study the correlation between two sets of data.
7. The forecasting of data can easily be done with the help of graphs.
8. Comparisons can be made very easily between two or more phenomena with the help of graphs.

**Disadvantages of Graphic Presentation**

Graphic presentation is not completely free from defects. The major disadvantages are as follows:

1. Graphs cannot be used to test the accuracy of the data.
2. Graphs cannot be used to show more than two features. It is not possible to have solutions of all the problems from graph. Hence, they furnish incomplete and inadequate information.
3. A small error in measurement scale may create large difference in the shape of graph.

**UNIT-III**

**Measure of Central tendency**

The point around which the observations concentrate in general in the central part of the data is called central value of the data and the tendency of the observations to concentrate around a central point is known as Central Tendency.

**Objects of Statistical Average:**

* To get a single value that describes the characteristics of the entire group
* To facilitate comparison

**Functions of Statistical Average:**

* Gives information about the whole group
* Becomes the basis of future planning and actions
* Provides a basis for analysis
* Traces mathematical relationships
* Helps in decision making

**Requisites of an Ideal Average:**

* Simple and rigid definition
* Easy to understand
* Simple and easy to compute
* Based on all observations
* Least affected by extreme values
* Least affected by fluctuations of sampling
* Capable of further algebraic treatment

**ARITHMETIC MEAN ()**

Arithmetic Mean of a group of observations is the quotient obtained by dividing the sum of all observations by their number. It is the most commonly used average or measure of the central tendency applicable only in case of quantitative data. Arithmetic mean is also simply called “mean”. Arithmetic mean is denoted by .

**Merits of Arithmetic Mean:**

* It is rigidly defined.
* It is easy to calculate and simple to follow.
* It is based on all the observations.
* It is readily put to algebraic treatment.
* It is least affected by fluctuations of sampling.
* It is not necessary to arrange the data in ascending or descending order.

**Demerits of Arithmetic Mean:**

* The arithmetic mean is highly affected by extreme values.
* It cannot average the ratios and percentages properly.
* It cannot be computed accurately if any item is missing.
* The mean sometimes does not coincide with any of the observed value.
* It cannot be determined by inspection.
* It cannot be calculated in case of open ended classes.

**Methods of Calculating Arithmetic Mean:**

* Direct Method
* Short cut method
* Step deviation method

**Use of Arithmetic Mean:**

Arithmetic Mean is recommended in following situation:

* When the frequency distribution is symmetrical.
* When we need a stable average.
* When other measures such as standard deviation, coefficient of correlation are to be computed later.

**MEDIAN (M)**

The median is that value of the variable which divides the group into two equal parts, one part comprising of all values greater and other of all values less than the median. For calculation of median the data has to be arranged in either ascending or descending order. Median is denoted by **M**.

**Merits of Median:**

* It is easily understood and easy to calculate.
* It is rigidly defined.
* It can sometimes be located by simple inspection and can also be computed graphically.
* It is positional average therefore not affected at all by extreme observations.
* It is only average to be used while dealing with qualitative data like intelligence, honesty etc.
* It is especially useful in case of open end classes since only the position and not the value of items must be known.
* It is not affected by extreme values.

**Demerits of Median:**

* For calculation, it is necessary to arrange data in ascending or descending order.
* Since it is a positional average, its value is not determined by each and every observation.
* It is not suitable for further algebraic treatment.
* It is not accurate for large data.
* The value of median is more affected by sampling fluctuations than the value of the arithmetic mean.

**Uses of Median:**

The use of median is recommended in the following situations:

* When there are open-ended classes provided it does not fall in those classes.
* When exceptionally large or small values occur at the ends of the frequency distribution.
* When the observation cannot be measured numerically but can be ranked in order.
* To determine the typical value in the problems concerning distribution of wealth etc.

**MODE (Z)**

Mode is the value which occurs the greatest number of times in the data. The word mode has been derived from the French word **‘La Mode’** which implies fashion. The Mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values. Mode is denoted by **Z.**

**Merits of Mode:**

* It is easy to understand and simple to calculate.
* It is not affected by extreme large or small values.
* It can be located only by inspection in ungrouped data and discrete frequency distribution.
* It can be useful for qualitative data.
* It can be computed in open-end frequency table.
* It can be located graphically.

**Demerits of Mode:**

* It is not well defined.
* It is not based on all the values.
* It is suitable for large values and it will not be well defined if the data consists of small number of values.
* It is not capable of further mathematical treatment.
* Sometimes, the data has one or more than one mode and sometimes the data has no mode at all.

**Uses of Mode:**

The use of mode is recommended in the following situations:

* When a quick approximate measure of central tendency is desired.
* When the measure of central tendency should be the most typical value.

**GEOMETRIC MEAN (G.M)**

The geometric mean also called geometric average is the nth root of the product of n non-negative quantities. Geometric Mean is denoted by **G.M**.

**Properties of Geometric Mean:**

* The geometric mean is less than arithmetic mean, G.M<A.M
* The product of the items remains unchanged if each item is replaced by the geometric mean.
* The geometric mean of the ratio of corresponding observations in two series is equal to the ratios their geometric means.
* The geometric mean of the products of corresponding items in two series.

**Merits of Geometric Mean:**

* It is rigidly defined and its value is a precise figure.
* It is based on all observations.
* It is capable of further algebraic treatment.
* It is not much affected by fluctuation of sampling.
* It is not affected by extreme values.

**Demerits of Geometric Mean:**

* It cannot be calculated if any of the observation is zero or negative.
* Its calculation is rather difficult.
* It is not easy to understand.
* It may not coincide with any of the observations.

**Uses of Geometric Mean:**

* Geometric Mean is appropriate when:
  + Large observations are to be given less weight.
  + We find the relative changes such as the average rate of population growth, the average rate of interest etc.
  + Where some of the observations are too small and/or too large.
* Also used for construction of Index Numbers.

**HARMONIC MEAN (H.M)**

Harmonic mean is another measure of central tendency. Harmonic mean is also useful for quantitative data. Harmonic mean is quotient of “number of the given values” and “sum of the reciprocals of the given values”. It is denoted by **H.M.**

**Merits of Harmonic Mean:**

* It is based on all observations.
* It not much affected by the fluctuation of sampling.
* It is capable of algebraic treatment.
* It is an appropriate average for averaging ratios and rates.
* It does not give much weight to the large items and gives greater importance to small items.

**Demerits of Harmonic Mean:**

* Its calculation is difficult.
* It gives high weight-age to the small items.
* It cannot be calculated if any one of the items is zero.
* It is usually a value which does not exist in the given data.

**Uses of Harmonic Mean:**

* Harmonic mean is better in computation of average speed, average price etc. under certain conditions.

**Dispersion**

The Dispersion (Known as Scatter, spread or variations) measures the extent to which the items vary from some central value. The measures of dispersion is also called the average of second order (Central tendency is called average of first order).

The two distributions of statistical data may be symmetrical and have common means, median or mode, yet they may differ widely in the scatter or their values about the measures of central tendency.

**Significance/ objectives of Dispersion-**

* To judge the reliability of average
* To compare the two an more series
* To facilitate control
* To facilitate the use of other statistical measures.

**Properties of good Measure of Dispersion**

* Simple to understand
* Easy to calculate
* Rigidly defined
* Based on all items
* Sampling stability
* Not unduly affected by extreme items.
* Good for further algebraic treatment

Dispersion

Based on selected Items

Graphic Method

Based on all items

1. Mean Deviation (coefficient of M.D)
2. Standard Deviation
3. Range (coefficient of Range)
4. Inter-quartile, coefficient of Range (IQR), (IQR)

Lorenz Curve

1. **Range: -** Range (R) is defined as the difference between the value of largest item and value of smallest item included in the distributions. Only two extreme of values are taken into considerations. It also does not consider the frequency at all series.
2. **Quartile Deviation: -** Quartile Deviation is half of the difference between upper quartile (Q3) and lower quartile (Q1). It is very much affected by sampling distribution.
3. **Mean Deviation: -** Mean Deviation or Average Deviation (δAlpha) is arithmetic average of deviation of all the values taken from a statistical average (Mean, Median, and Mode) of the series. In taking deviation of values, algebraic sign + and – are also treated as positive deviations. This is also known as first absolute moment.
4. **Standard Deviation:-** The standard deviation is the positive root of the arithmetic mean of the squared deviation of various values from their arithmetic mean. The S.D. is denoted as σ Sigma.

**Method of calculating standard Deviation-**

1. Direct Method 2. Short-cut-Method 3. Step deviations Method

**Properties**

Fixed Relationship among measures of dispersion in a normal distribution there is a fixed relationship between quartile Deviation, Mean Deviation and Standard Deviation Q.D = 2/3 σ, Mean Deviation = 4/5σ.

**Distinction between mean deviation and standard deviation**

|  |  |  |
| --- | --- | --- |
| **Base** | **Mean Deviation** | **Standard Deviation** |
| 1. Algebraic Sign | Actual +, - Signs are ignored and all deviation are taken as positive | Actual signs +, - are not ignored whereas they are squared logically to be ignored. |
| 2. Use of Measure | Mean deviation can be computed from mean, median, mode | Standard deviation is computed through mean only |
| 3. Formula | M.D or δ = | S.D or σ = |
| 4. Further algebraic  Treatment | It is not capable of further algebraic treatment. | It is capable of further algebraic treatment |
| 5. Simplicity | M.D is simple to understand and easy to calculate | S.D is somewhat complex than mean deviation. |
| 6. Based | It is based on simple average of sum of absolute deviation | It is based on square root of the average of the squared deviation |

**Variance**

The square of the standard deviation is called variance. In other words the arithmetic mean of the squares of the deviation from arithmetic mean of various values is called variance and is denoted as σ2. Variance is also known as second movement from mean. In other way, the positive root of the variance is called S.D.

Coefficient of Variations- To compare the dispersion between two and more series we define coefficient of S.D. The expression isx 100 = known as coefficient of variations.

**Interpretation of Coefficient of Variance-**

|  |  |
| --- | --- |
| **Value of variance** | **Interpretation** |
| Smaller the value of σ2 | Lesser the variability or greater the uniformity/ stable/ homogenous of population |
| Larger the value of σ2 | Greater the variability or lesser the uniformity/ consistency of the population |

**Dispersion**

**Range = R**

|  |  |  |
| --- | --- | --- |
| **Individual Series** | **Discrete Series** | **Continuous Series** |
| Range = L-S  Where L=Largest,  S=Smallest Observation |  |  |
| Coefficient of Range |  |  |

**Quartile Deviation - Q.D.**

|  |  |  |
| --- | --- | --- |
| **Individual Series** | **Discrete Series** | **Continuous Series** |
|  |  |  |
| Coefficient of Q.D. |  |  |

**Mean Deviation - M.D. δ (“Through actual mean, mode, median)**

|  |  |  |
| --- | --- | --- |
| **Individual Series** | **Discrete Series** | **Continuous Series** |
|  |  |  |
| Coefficient of |  |  |
| Mean |  |  |
| Coefficient of |  |  |
| (Mode) |  |  |
| Coefficient of |  |  |

**Standard Deviation =  can be calculated through mean only**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Individual Series** | **Discrete Series** | **Continuous Series** |
| Direct  (Through actual mean) |  |  |  |
| Indirect (Through assumed mean) |  |  |  |

**UNIT-IV**

**Correlation**

**Introduction**

1. Correlation is a statistical tool & it enables us to measure and analyse the degree or extent to which two or more variable fluctuate/vary/change w.e.t. to each other.
2. For example – Demand is affected by price and price in turn is also affected by demand. Therefore we can say that demand and price are affected by each other & hence are correlated. the other example of correlated variable are –
3. While studying correlation between 2 variables use should make clear that there must be cause and effect relationship between these variables. For e.g. – when price of a certain commodity is changed (↑ or ↑) its demand also changed (↑ or ↑) so there is case & effect relationship between demand and price thus correlation exists between them. Take another eg. Where height of students; as well as height of tree increases, then one cannot call it a case of correlation because neither height of students is affected by height of three nor height of tree is affected by height of students, so there is no cause & effect relationship between these 2 so no correlation exists between these 2 variables.
4. In correlation both the variables may be mutually influencing each other so neither can be designated as cause and the other effect for e.g. –

Price ↑→ Demand ↓

Demand ↓→ Price ↑

So, both price & demand are affected by each other therefore use cannot tell in real sense which one is cause and which one is cause and which one is effect.

**definitions of Correlation**

1. “If 2 or more quantities vary is sympathy, so that movements is one tend to be accompanied by corresponding movements in the other(s), then they are said to be correlated”. **Connor.**
2. “Correlation means that between 2 series or groups of data there exists some casual correction”. **WI King**
3. “Analysis of Correlation between 2 or more variables is usually called correlation.” **A.M. Turtle**
4. “Correlation analysis attempts to determine the degree of relationship between variables.

**Ya Lun chou**

**types of correlation**

Correlation

Positive Negative Correlation

Simple & Multiple Correlation

Partial & Total Correlation

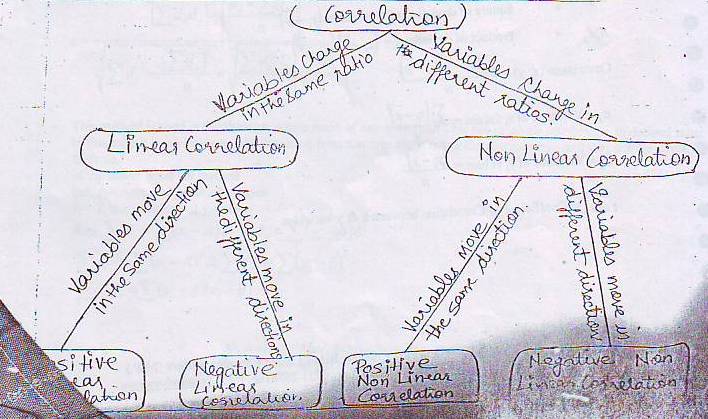
Liner & Non Linear Correlation

|  |  |  |
| --- | --- | --- |
|  | **Positive Correlation** | **Negative Correlation** |
| 1 | Value of 2 variables move in the same direction i.e. when increase/decrease in value of one variable will cause increase or decrease in value of other variable. | Value of 2 variables move in opposite direction i.e. when one variable increased, other variable decreases when one variable is decreased, other variable increase. |
| 2 | E.g. Supply & Price  So, supply and price are …….correlated  P = Price/Unit  Q = quantity Supplied | E.g. Demand & Price  So, Demand & Price vely correlated  P = Price/Unit  Q = quantity Supplied |

|  |  |  |
| --- | --- | --- |
|  | **simple Correlation** | **multiple Correlation** |
| 1 | In simple correlation, the relationship is confined to 2 variables only, i.e. the effect of only one variable is studied | The relationship between more than 2 variables is studied. |
| 2 | E.g. Demand & Price  Demand depends on → Price  This is case of simple correlation because relationship is confined to only one factor (that affects demand) i.e. price so we have to find correlation between demand & price.  If, demand = Y  If, demand – X  Then, Correlation between Y & X | E.g. Demand & Price  Demand depends on → Price  Demand on → income  This is case of multiple correlations because 2 factors (Price & Income) that affects demand are taken. We have to find correlation between demand & price.  Demand & Price  If, demand = Y  Price = X1  Price = X2  Then  Correlation between Y & X1  Correlation between Y & X2 |

|  |  |
| --- | --- |
| **simple Correlation** | **multiple Correlation** |
| In partial correlation though more than 2 factors are involved but correlation is studies only between to be constant.  E.g.  X1 Y = Demand  Y X1 = Price  X2 X2 = Income | In total correlation relationship between all the variables is studied i.e., none of item is assumed to be constant  E.g.  X1 Y = Demand  Y X1 = Price  X2 X2 = Income |
| If we study correlation between Y & X1 & assume X2 to be constant it is a case of partial correlation. this is what we do in law of demand – assume factors other than price as constant (Ceteris paribus – Keeping other things constant) | If we assume that income is not constant i.e. we study the effect of both price & income on demand, it is a case of total correlation.  In other words, cataris paribus assumption is relaxed in this case. |

|  |  |  |
| --- | --- | --- |
|  | **linear Correlation** | **non-linear Correlation** |
| 1 | In linear correlation, due to unit, change value of one variable there is constant change in the value of other variable. The graph for such a relationship is straight line. E.G. – If in a factory no of workers are doubled, the production output is also doubled, and correlation would be linear. | In non linear or curvilinear correlation, due to unit, change value of one variable, the change in the value of other variable is not constant. the graph for such a relationship is a curve. E.G. – The amount spent on advertisement will not bring the change in the amount of sales in the same ratio, it means the variation. |
| 2 | If the changed in 2 variables are in the same direction and in the constant ratio, it is linear positive correlation  Y  X   |  |  | | --- | --- | | X | Y | | 2 | 3 | | 4 | 6 | | 6 | 9 | | 8 | 12 | | If the change in 2 variables is in the same direction but not in constant ratio, the correlation is non linear positive.  Y  X   |  |  | | --- | --- | | X | Y | | 50 | 10 | | 55 | 12 | | 60 | 15 | | 90 | 30 | | 100 | 45 | |
| 3 | If changes in 2 variables are in the opposite direction but in constant ratio, the correlation is linear negative. For eg. Every 5% ↑ is price of a good is associated with 10% decrease in demand the correlation between price and demand would be linear negative.   |  |  | | --- | --- | | X | Y  Y  X | | 2 | 21 | | 4 | 18 | | 6 | 15 | | 8 | 12 | | 10 | 9 | | If changes in 2 variables are in opposite direction and not in constant ratio, the correlation is non linear negative. For eg: - every 5% ↑in price of good is associated with 20% to 10%↓in demand, the correlation between price & demand would be non linear negative.   |  |  | | --- | --- | | X | Y  Y  X | | 80 | 50 | | 55 | 60 | | 50 | 75 | | 90 | 130 | |

****

**Type – 1 [Based on karl Pearson’s cofficient of correlation]**

Before use move to numerical, use understand the basic notions & concepts –

dx = Deviations of xi value from mean = (xi - )

x = Mean of x value [Average of X values] =

n = No. of observations

dy = Deviation of y value from mean = (y - )

= Mean of y values =

d2x = Square of deviation of x values = (xi - )2

d2y = Square of deviation of x values = (yi - )2

dxdy = Product of deviations = (xi - ) (yi - )

Covariance (x,y) =

σx = Variance of xi values =

σy = Variance of yi values =

r or rxy = coefficient of correlation between x 7 y variables.

Direct Method for Karl Pearson’s Coefficient of correlation



Deviation from actual mean method



Put 

Deviation from assumed mean method (Short Cut Method)



This method is used in the situation where mean of any series (x or y) is not in whole number, i.e. in decimal value. in this case it is advisable to take deviation from assumed mean rather than actual mean and then use the above formula.

In the above short cut method

Let, A = Assumed mean of X series

B = Assumed mean of y series

then Σdx = Σ(xi – A) & Σdy = Σ(yi – B) &

Σdx 2= Σ(xi – A)2 & Σdy2= Σ(yi – B)2

Σdxdy= Σ(xi – A)(xi – B)

**Regression Analysis**

The dictionary meaning of regression is “Stepping Back”. The term was first used by a British Biometrician” Sir Francis Galton 1822 – 1911) is 1877. He found in his study the relationship between the heights of father & sons. In this study he described “That son deviated less on the average from the mean height of the race than their fathers, whether the father’s were above or below the average, son tended to go back or regress between two or more variables in terms of the original unit of the data.

**Meaning**

Regression Analysis is a statistical tool to study the nature extent of functional relationship between two or more variable and to estimate the unknown values of dependent variable from the known values of independent variable.

**Dependent Variables –** The variable which is predicted on the basis of another variable is called dependent or explained variable (usually devoted as y)

**Independent variable –** The variable which is used to predict another variable called independent variable (denoted usually as X)

**Definition**

Statistical techniques which attempts to establish the nature of the relationship between variable and thereby provide a mechanism for prediction and forecasting is known as regression Analysis.

**– Ya-lun-Chon”**

**Importance/uses of Regression Analysis**

* Forecasting
* Utility in Economic and business area
* Indispensible for goods planning
* Useful for statistical estimates.
* Study between more than two variable possible
* Determination of the rate of change in variable
* Measurement of degree and direction of correlation
* Applicable in the problems having cause and effect relationship
* Regression Analysis is to estimate errors
* Regression Coefficient (bxy & byx) facilitates to calculate of determination ® & coefficient or correlation (r)

**Regression Lines**

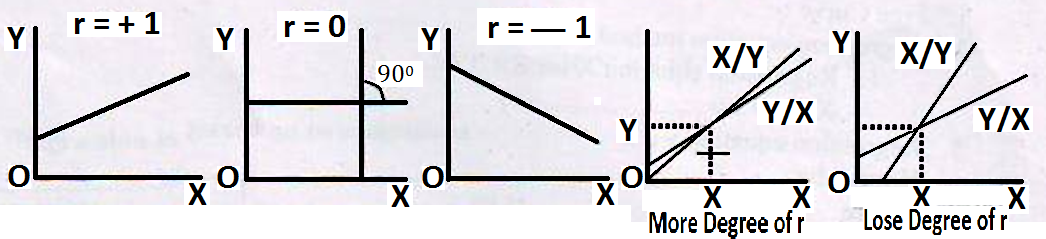
The lines of best fit expressing mutual average relationship between two variables are known as regression lines – there are two lines of regression

**Why are two Regression lines –**

1. While constructing the lines of regression of x on y is treated as independent variables where as ‘x’ is treated as treated as dependent variable. This gives most probable values of ‘X’ for gives values of y. the same will be there for y on x.

**Relationship between correlation & REgression**

1. When there is perfect correlation between two series (r = 1) the regression with coincide and there will be only one regression line.
2. When there is no correction (r = o)> both the lines will cut each other at point.
3. Where there is more degree of correction, say (r = 70 or more the two regression line with be next to each other whereas when less degree of correction. Say (r= 10 on less) the two regression line will be a parted from each other.

**regression lines and degree of correlation** 

**difference between Correlation and REgression Analysis**

the correlation and regression analysis, both, help us in studying the relationship between two variables yet they differ in their approach and objectives. The choice between the two depends on the purpose of analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Base** | **Correlation** | **Regression** |
| 1 | meaning | Correlation means relationship between two or more variables in which movement in one have corresponding movements in other | regression means step ping back or returning to the average value, i.e., it express average relationship between two or more variables. |
| 2 | relationship | correlation need not imply cause and effect relationship between the variables under study | regression analysis clearly indicates the cause and effect relationship. The variable(s) constituting causes(s) is taken as independent variables (s) and the variable constituting the variable consenting the effect is taken as dependent variable. |
| 3 | object | correlation is meant for co-variation of the two variables. the degree of their co-variation is also reflected in correlation. but correlation does not study the nature of relationship. | regression tells use about the relative movement in the variable. We can predict the value of one variable by taking into account the value of the other variable. |
| 4 | nature | there may be nonsense correlation of the variable has no practical relevance | There is nothing like nonsense regression. |
| 5 | measure | correlation coefficient is a relative measure of the linear relationship between X and Y. It is a pure number lying between 1 and +1 | the regression coefficient is absolute measure representing the change in the value of variable. We can obtain the value of the dependent variable. |
| 6 | application | correlation analysis has limited application as it is confined only to the study of linear relationship between the variables. | Regression analysis studies linear as well as non linear relationship between variables and therefore, has much wider application. |

**Why least square is the Best?**

When data are plotted on the diagram there is no limit to the number of straight lines that could be drawn on any scatter diagram. Obviously many lines would not fit the data and disregarded. If all the points on the diagram fall on a line, that line certainly would the best fitting line but such a situation is rare and ideal. Since points are usually scatters, we need a criterion by which the best fitting line can be determined.

**Methods of Drawing Regression Lines –**

1. Free curve –
2. regression equation x on y,

X = a + by ……………………………. (1)

1. regression equation y on x

y = a + bx

where

‘a’ is that point where regression lines touches y axis (the value of dependent variable value when value or independent variable is zero)

‘b’ is the slop of the said line (The amount of change in the value of the dependent variable per unit change)

Change in independent variable)

A and b constants can be calculated through –

Σ(x = a + by) (by multiplying ‘Σ’)

Σx = Na + bΣy (1)

Σx (y = a + bx) (by multiplying Σx)

Σxy = Σxa + bΣx2 (2)

**kinsds of regression analysis**

1. linear and Non- Linear Regression
2. Simple and Multiple Regression

**functions of regression lines –**

1. to make the best estimate –
2. to indicate the nature and extent of correlation

**regression equations –**

the regression equation’s express the regression lines, as there are two regression lines there are two regression equations –

Explanation is given in formulae –

**regression lines**

1. regression equation of x on y

x – X = bxy (y – y)

where bxy = regression coefficient of X on Y

1. Regression euation of y on x

y – Y = bxy (x – x) where bxy = regression coefficient of Y on X

and also and

Relation between r, byx and bxy r = 

**REGRESSION ANALYSIS**

Regression is based on two equations –

|  |  |  |
| --- | --- | --- |
| **Equations** | **X on y** | **y on x** |
| After elaborating them |  |  |
| Coefficient of Regression |  |  |
| To fin out coefficient of regression through actual mean |  |  |
| through assumed mean |  |  |
|  | | |

**regression coefficient –** There are two regression coefficient like regression equation, they are (bxy and byx)

Properties of regression coefficients –

* Same sign – Both coefficient have the same either positive on negative
* Both cannot by greater than one – If one Regression is greater than “One” or unity. Other must be less than one.
* Independent of origin – Regression coefficient are independent of origin but not of scale.
* A.M.> ‘r’ – mean of regression coefficient is greater than ‘r’
* R is G.M. – Correlation coefficient is geometric mean between the regression coefficient
* R, bxy and bxy – They all have same sign

**Index numbers**

Index numbers are devices which measure the change in the level of a phenomenon with respect to time, geographical location or some other characteristic. The first index number was constructed in the year 1764 by an Italian named Carli to compare the changes in the price for the year 1750 with the price level of the year 1500. In present day situation changes in production, consumption, exports, imports, national income, cost of living, incidence of crimes, number of road accidents, business failures and a very wide variety of other phenomena are studied with the help of index numbers. Index numbers are supposed to be barometers which measure the change in the level of a phenomenon.

“An index number is a statistical measure designed to show changes in variable or a group of related variables with respect to time, geographical location or other characteristics.”

**Spiegel**

**Characteristics of index numbers**

1. **Index numbers are a specialised type of average.** Averages can be used to compare only those series which are expressed in the same units. However the device of index number helps us in comparing change in series which are in different units.
2. **Index numbers study the effects of such factors which cannot be measured directly.** Index numbers are meant to study the changes in the effects of such factors which cannot be measured directly.
3. **Index numbers being out the common characteristics of a group items.**
4. **Index number measure only relative changes in the values of a phenomenon.**

**Uses of index numbers**

1. **Help in Studying Trends.** Index numbers helps to find out the trend of exports, imports, balance of payments, industrial production, prices, national income and a variety of other phenomena.
2. **Help in policy formulation.** Index numbers help us in studying trends of various phenomena and these trends and tendencies are the bases on which may policy decisions are taken index number are used by the government in deciding the rates of D.A. and levy of excise duties.
3. **Help in measuring the Purchasing Power of Money.** Index numbers are helpful in finding out the intrinsic worth of money as contrasted with its nominal worth.
4. **Helps in deflating various values.** Index numbers are very helpful in deflating national income on the basis of constant prices.
5. **Act as economic barometers.** Index numbers measure the pulse of an economy and act as barometers to find the ups and down in the general economic condition of a country.

**Types of index numbers**

(a) Price Index Number (Wholesale and Retail)

(b) Quantity Index Numbers.

(c) Value Index Numbers.

(d) Special Purpose Index numbers.

**Problems in the construction of index numbers**

1. **The selection of item-** The first problem which the marker of an index number of wholesale prices has to face is that of the selection of items from which the index number is to be constructed.
2. **The selection of the base year-** Second problem in the construction of index numbers is the selection of the base year and the conversion of current prices to price relatives based on the prices of the base year.
3. **The selection of the average-** The next step in the construction of wholesale price index number is to average the prices relatives of the various commodities.
4. **Selecting suitable weights.** All the items used in the construction of an index number are not of equal importance and as such if the index number is to be a representative one, weights should be assigned to various items in relation to their importance.

**Methods of constructing index numbers**

Broadly speaking various methods of constructing index numbers can be classified in two groups viz.

1. Unweighted Index Numbers
2. Weighted Index Numbers
3. Simple Aggregative Method
4. Simple Average of Relatives Method.

**A Unweighted Index Numbers**



Where

P01 = Index number of the current year

= Total of the current year; price of all commodities.

= Total of the base year’s price of all commodities.

**Simple Average of Relatives Method.** 

**B. Weighted Index Numbers**

1. **Laspeyres Method -**  **2. Passche’s Method** 

**3. Drobish and Bowleys Method** 

**4.** **Fisher’s Ideal Index**. 

**5.** Marshall-Edgeworth formula

 or 

**6. Walsch Formula.  7. Kelly’s Method. **

**Quantity Index Number**

Quantity index number measure the changes in the volume of production, construction or employment over a period of years.

Formula for simple or unweighted quantity index;

 Here 

**Base Shifting:- Base shifting is generally required due to following reasons**

1. The base year is too old to compare the current year.
2. If different series of index numbers are based on different base years and they are to be compared from each other.

**Deflation of index numbers**

Computation of real wages from money income with taking the effect of price level changes is called as deflating of index numbers.

****

**Splicing :** Sometimes series of index number based on a certain year is discontinued and a new series of index number is prepared by taking another year as base. Thus two series of index number would result. In this situation index number of these two series are not comparable because both are based on different years. If these are to be compared then new series will be covered on the basis of old series or vice-versa; this conversion/shifting is called as spicing. Splicing may be taken as another form of base shifting.

Formula for splicing :-

1. **Splicing of new series in old series (Forward splicing):**

****

Spliced Index Number

1. **Splicing of old series in new series (backward splicing):**

****

**Tests of adequacy of index number formulae**

We have discussed a large number of formulae for the construction of both simple and weighted index numbers. We formula should be chosen for the construction of an index number is a question which cannot be satisfactorily answered. However some tests have been suggested to determine the adequacy of an index number formula. These tests are:

1. **Unit Test –** This test requires that the formula for the construction of index numbers should be such which is not affected by the unit in which prices or quantities have been quoted. This test is satisfied by all index number formulae discussed above except the simple (unweighted) aggregative index formula. In this index as we have discussed earlier the units play an important part in determining the value of the index. If only the unit is changed (say from kg to quintal) the value of the index would change.
2. **Time Reversal Test-** In the worlds of Fisher: “The test is that the formulae for calculating an index number should be such that it will give the same ratio between one point of comparison and the other no matter which of the two is taken as base.” This mean that the index number should work both backwards as well as forwards. Thus, if the index number of the current year is 4000 then the index number of the base year (based on the current year) should be 25. In other words, the two index numbers thus calculated (without the figure 100) should be reciprocals of each other. The reciprocal of 4 is .25 and the reciprocal of .25 is 4. The product of these two ratios would always be equal to one. Thus, if P­10 represents the price change in the current year and P10 the price change of the base year (based on the current year) the following equation should be satisfied:-
3. **Factor Reversal Test-** In the words of Fisher: “Just as each formula should permit inter-changing the price and quantities without giving inconsistent result, i.e., the two results multiplied together should give the trust value ratio.” It means that the changes in the prices multiplied by the changes in quantity should be equal to the total change in value. Change in value is the result of changes should represent the total change in value. Thus, if the price of a commodity has doubled during a certain period and if in this period the quantity has trebled the total change in the value should be six time the former level. In the other words, if p1 and p0 represent the prices and q1 and q0 the quantities in the current and the base years respectively, and if p01 represent the change in price in the current year and q01 the change in the quantity in the current year then



The factor reversal test is satisfied only by the Fisher’s Ideal Index Number.

The proof of it is given below:

**Circular Test**

Another test applied in index number studies is the circular test. It is a short of extension of the time reversal test. Suppose an index number is constructed for the year 1983 with the base of 1982 and another index number for 1982 on the base of 1981, then it should be possible for us to directly get an index number for 1983 on the base of 1981. If the index number calculated directly does not give an inconsistent value, the circular test is said to be satisfied. If p­01 represent the price change of the current year on the base year and P12 the price change of the base year on some other base and p20 the price change of the current year on this second base then the following equation should be satisfied.

**UNIT-V**

**Probability Theory**

**Meaning of Probability**

Uncertainty is part of life. If the decisions could have been made under conditions of certainty, one could make a perfect forecast of the future. Unfortunately, we all live in a world of uncertainty, Weather, stock-market prices, product quality are some areas where it is not possible to comment on futures with certainty. Decision making in such areas is very crucial. For this purpose, the concept of probability is of great importance. Mathematicians have always been interested to quantify uncertainties associated with an event so that one may take a better decision when a situation arises. Probability is a mathematical measure of uncertainty. Probability theory provides us the means and ways to make some precise expressions for uncertainty.

The word ‘probability’ or ‘chance’ is very commonly used in day-to-day conversation. **For** **example**, we come across statements like.

1. Probably it may rain tomorrow.
2. Mr. X may not come in the meeting.
3. The chances of winning of both teams are equal, etc.

In general, the word ‘probability’ is used to express that there is uncertainty about something. But in mathematics and statistics, we try to present conditions under which we can make numerical statements about uncertainty and apply some methods of calculating numerical values of probability.

**Basics Concepts of Probability**

The theory of probability employs certain concepts which are as follows:

1. **Experiment:** An activity that produces some result and which can be repeated in identical environment is called an experiment. Thus, tossing a coin, throwing a dice or circulating an advertisement, etc. is some of the examples of experiments, because these activities can be repeated as many times as one wants producing some results. If an activity can be termed as experiment.

The experiment may be further divided into two parts:

1. **Deterministic experiment:**

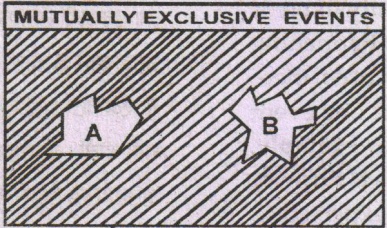
The experiments whose outcomes can be predicted are called deterministic experiments.

1. **Random or Stochastic Experiments:** If all the possible outcomes of an experiment are known but it can not be predicted with certainty which will take place, it is called random or stochastic experiment. **For example,** when a coin is tossed, we know that either a head or a tail will appear but we can not say with certainty that which will turn up. It is called a random or stochastic experiment.
2. **Events:** In probability theory, an event constitutes one or more possible outcomes of an experiment. Thus, when a dice is thrown, it would be called experiment and the result (any number 2, 3, 4, 5 or 6) is called an event.

Following are the different types of events:

1. **Simple of Compound Events:** Events could be either **simple** or **compound** (also called simple if there is only one possible outcome. **For** **example,** if a dice is thrown, the chance of getting 2 is a simple events because there is only one 2 in the dice but the chance of getting an even number is a compound event because even numbers are more than one(2, 4, and 6).
2. **Mutually Exclusive Event**: Two events are said to be mutually exclusive when both can not happen simultaneously in a single trial. In other words, if one event happens, then another can not happens and vice versa. **For example,** if a single coin is tossed; either head or tail can be up. Both can not be up at the same time. Similarly, if a single dice is thrown, the occurrence of 5 and 6 are mutually exclusive because any one of these two can occur at a time. Symbolically, if A and B are mutually exclusive events, then P(AB) = 0 i.e. probability of both A and B is zero.

The following diagram will clearly illustrate the meaning of mutually exclusive events:



1. **Independent Events:** Two or more events are said to be independent when the result of one does not affect and is not affected by the other. **For example,** if a coin is tossed two times, the result of the second toss would not be affected by the result of first toss. In the same way, if a dice is thrown repeatedly, coming up of 3 in the first thrown is independent of coming up of 3 in the second thrown.
2. **Dependent Events:** Dependent events are those in which the occurrence or non-occurrence of one event in any one trial affects the probability of other events in other trials. **For example,** the chance of getting a queen on the first draw from a pack of card is 4/52 (because there are 4 queens in a pack). If the first card is queen and the card is not replaced before the second draw, the chance of getting a queen again is 3/51 (because there are 51 card only and they contain only 3 queens).
3. **Equally Likely Events:** Events are said to be equally likely if the chance of their happening is equal. In other words, there is no preference of one event over the other. **For example,** if an unbiased dice is thrown, each face may be expected to be occurred the same number of times in the long run.

Similarly, the cards of a pack are so closely alike that the chance of each card to appear is equal when a large number of draws are made with replacement. However, if the coin or dice is biased, the case would not be called equally likely.

**Definition**

“It is the ratio of favourable events to the total number of equal likely events.” **– La place**

In simple words, probability is the ratio of possibility of the happening of any one of the several equally likely events.

**Main Features of Probability**

Following are the main features of probability:

1. Probability is the chance that something will happen.
2. Probabilities are expressed as fractions like 1/2, 5/6, 2,5, etc. or as decimals like 0.5, 0.33, etc.
3. The probability varies between 0 and 1. Zero probability means something will never happen and probability of one indicates that something will always happen.
4. In probability theory, an event is one or more of the possible outcomes of doing something. If we toss a coin getting a head would be an event and getting a tail would be another event.
5. The activity that produces such an event is referred in probability theory as an experiment. **For example,** tossing a coin or drawing a card from a pack of cards are known as experiments.

**Application of Probability in Decision Making**

Probability theory is used to analyze data for decision making. There are various business decisions in real life when the decision maker is very uncertain as to what will happen after the decisions are made. Probabilistic models help the businessman in taking decisions under such situations. Probability theory is successfully used in solving the problems of industries like inventory control problem, replacement problem, waiting line problem, etc. As a result, probability theory has become an essential which involve uncertainty. **For example,**

1. The insurance company uses probability theory to calculate premium rates.
2. An investor estimates about returns of the stock and economic scenarios.
3. A project manager estimates the probability of the success of his project.

**Mutually Exclusive as well as independent Event**

Can an event be mutually exclusive and independent simultaneously? The answer is yes. Mutually exclusive means any one event can occur at a time and independent means which is not affected and does not affect the other event. So, an event can be mutually exclusive as well as independent at a time. **For example,** if an unbiased dice is thrown, occurrence of any number (say 5) is mutually exclusive because at a time any one number can occur. At the same time, it is independent also because occurrence of any number does not affect and is not affected by the result of second throw.

**Types of Probability**

There are three basic ways of classifying probability. These three represent different conceptual approaches to the study of probability theory. Experts have different opinions about which approach is proper to use. These are:

1. Classical or Mathematical Approach
2. Relative Frequency Approach
3. Subjective Approach

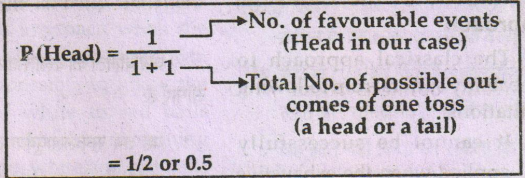
**Classical or Mathematical Approach**

Classical approach of probability was suggested by Bernoulli. This approach was developed to meet the gambling problems in 18th century. The basic assumption of classical approach is that the outcomes of a random experiment are ‘Equally likely’, ‘Mutually exclusive’ and ‘Collectively exhaustive’. The ‘event’ under consideration possesses one or more possible outcomes. **For example,** when a dice is thrown, any one of the six outcomes (1, 2, 3, 4, 5 or 6) can occur.

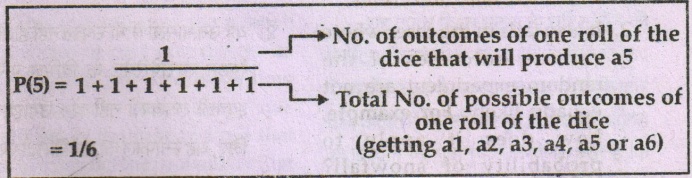
According to Classical approach, the probability of an event may be calculated as follows:

|  |
| --- |
| Probability of an event = |

Here, it is assumed that the outcomes of the experiment are equally likely, mutually exclusive. **For example,** the probability of getting a head on a toss of a fair coin will be determined as:



Similarly, if we calculate the probability of getting 5 by rolling a dice.



Classical probability is often called a ‘priory’ probability because if we use fair coins or unbiased dice, we can find the answer in advance (priory) without tossing a coin or rolling a dice. We need not to perform experiments to make our probability statement about fair coins or unbiased dice.

This approach to probability is useful when we deal with card games or tossing coins and dice. But it can not be used in solving decision problems which one encounter in management. **The classical approach to probability assumes a world that does not exist. It is not suitable for the real life situations which are often disorderly and unlikely.**

**Limitations of the Classical Approach**

The classical approach to probability suffers from following limitations.

1. It cannot be successfully applied when the exhaustive number (n) of outcomes of the random experiment is unlimited and un-certain.
2. It also fails in the cases where various outcomes of the random experiment are not equally likely. **For example,** how does it apply to probability of snowfall? Nobody will agree at any time that the two contingencies-‘Snowfall’ and ‘no-snowfall’ have same likelihood i.e. they are equally likely. Likewise if a person jumps from the top of a twenty storied building, then the probability of his survival will not be 50%, the two mutually exhaustive outcomes, viz, survival and death in this case are not equally likely.
3. The concept of probability based on logic are not always in statistics. It is difficult to predict that how many bulbs will be defective from 2,000 bulbs unless one really observes. Real life situations unlikely and disorderly as often they are, make it difficult and at times impossible to apply classical approach.
4. It is also difficult to apply classical approach when the actual value of n is not known. It is impossible to know the ratio of white to red balls drawn from a bag containing unknown white and red balls unknown white and red balls without actual experiment.

**Relative Frequency Approach**

This method uses the relative frequencies of past occurrences as probability. We determine how many times something has happened in the past and use that figure to predict the probability that it will happen again in the future.

During 17th Century, British Statisticians, who were interested in calculating risk of losses in life insurance and commercial insurance, started defining probabilities from statistical data collected on births and deaths. Today, this approach is called the ‘relative frequency of occurrence’. It defines probability as:

1. The observed ‘relative frequency of an event in a very large number of trials or
2. The proportion of times that an event occurs in the long run when conditions are stable.

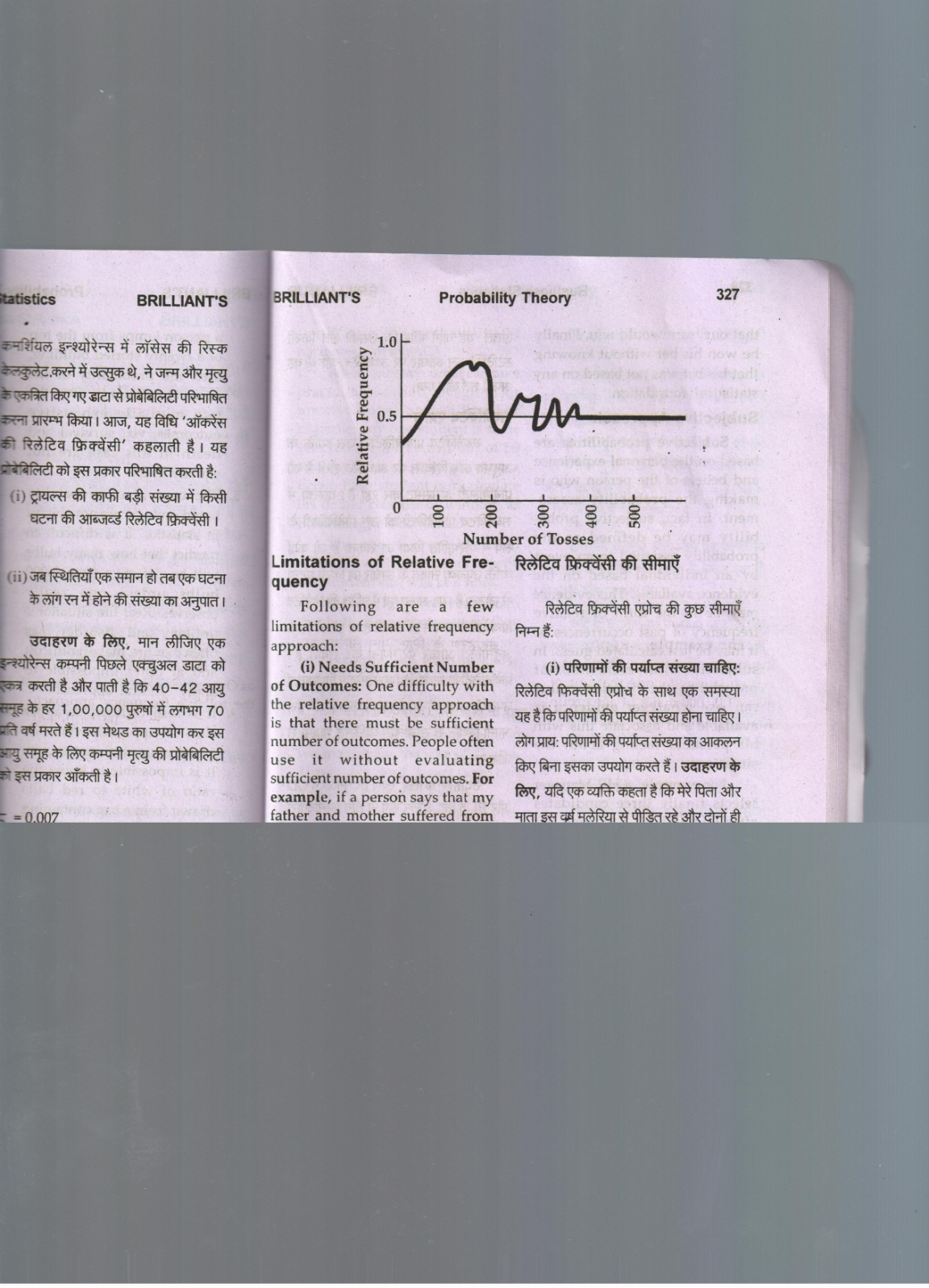
**For example,** suppose an insurance company collects past actual data and finds that

Out of ever 1,00,000 males of age group 40–42, about 70 die within one year. Using this method, the company estimates the probability o9f death for that age group as:

****

When we use the relative frequency approach to establish probabilities, it will be

more accurate if we increase the number of observations. If we toss a fair coin 500 times, we may observe that the proportion of heads was away from 0.5 in the first 100 tosses, but it seemed to establish and approach 0.5 as the number of tosses increased. The following figure illustrates the outcomes of these 500 tosses:



**Limitation of Relative Frequency**

Following are a few limitations of relative frequency approach.

**(i) Needs Sufficient Number of Outcomes:** One difficulty with the relative frequency approach is that there must be sufficient number of outcomes. People often issue it without evaluating sufficient number of outcomes. **For example,** if a person says that my father and mother suffered from malaria this year and both are above 60 and so, everyone in that age group will probably get the malaria. We know that this assumption is not based on enough evidence as the data observed were in sufficient.

**(ii) Estimates Which are not Based on Statistics:** Sometimes the estimates get correct without any previous base of statistics. Suppose Indian Cricket team was playing first time in 20–20 world cup and a loyal fan betted Rs. 1,000 that out team would win. Finally he won his bet without knowing that his bet was not based on any statistical foundation.

**Subjective Approach**

Subjective probabilities are based on the personal experience based on the personal experience and beliefs of the person who is making the probability assessment. In fact, subjective probability may be defined as the probability assigned to an event by an individual based on the evidence available. This evidence may be in the form of relative frequency of past occurrences or it may be just calculated guess. In subjective assessment of probability, the decision marker can use whatever evidence is available and associate this with his personal felling about the situations.

**For example,** aHR Manager selects finally three candidates after GD and PI. Each one has confidence, high level of energy and an attractive appearance. Choosing any one among the three will require him to assign a subjective probability to each one’s potential.

**Some Commonly Used Symbols and Rules of Probability**

In probability theory, a few symbols are used to simplify the presentation of ideas. **For example,** the probability of the events A is expressed as:

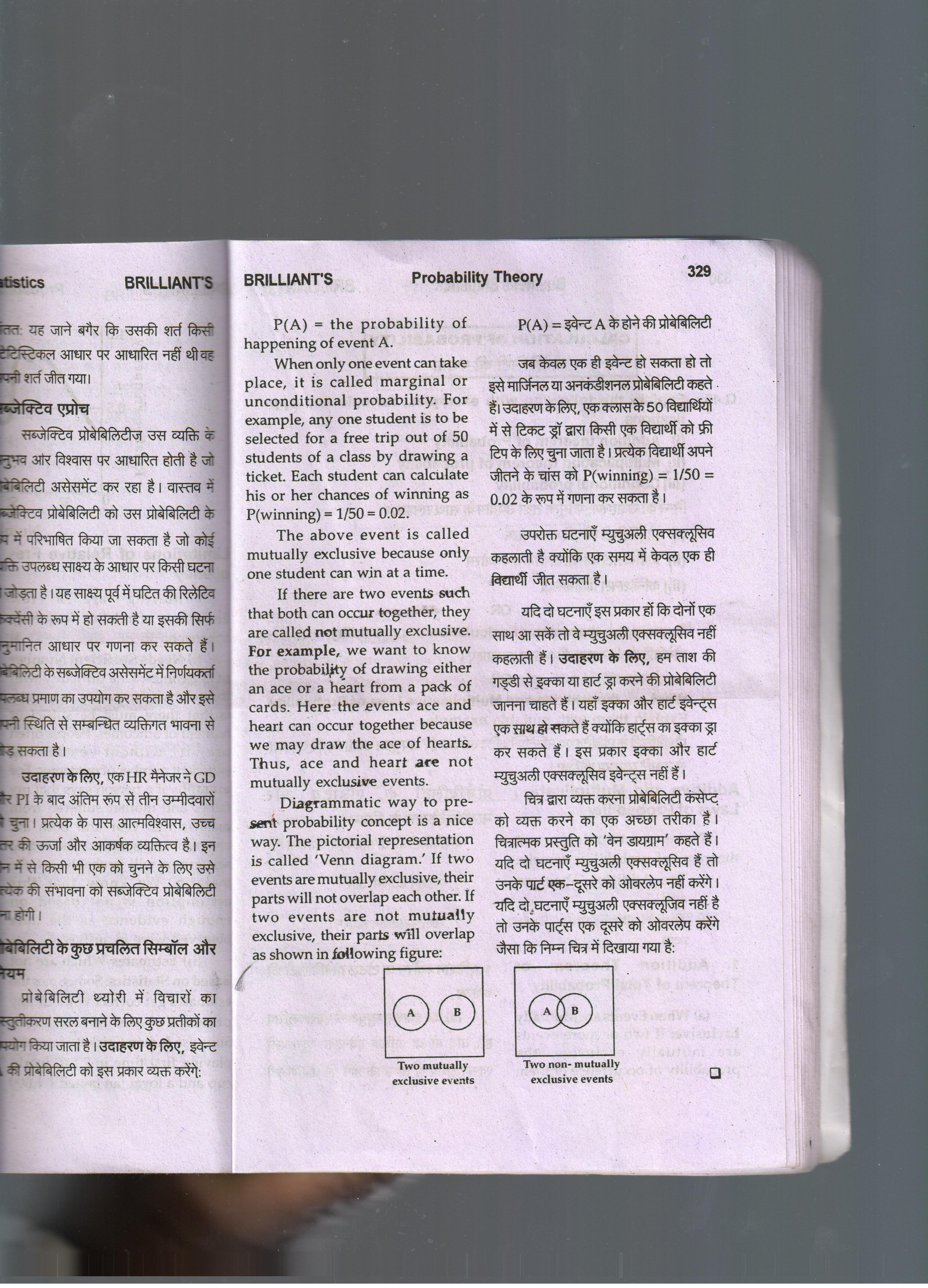
P(A) = the probability of happening of event A.

When only one event can take place, it is called marginal or unconditional probability. For example, any one student is to be selected for a free trip out of 50 students of a class by drawing a ticket. Each student can calculate his or her chances of winning as P(winning) = 1/50 = 0.2.

The above event is called mutually exclusive because only one student can win at a time.

If there are two events such that both can occur together, they are called not mutually exclusive. **For example,** we want to know the probability of drawing either an ace or a heart from a pack of cards. Here the events ace and heart can occur together because we may draw the ace of hearts. Thus, ace and heart are not mutually exclusive events.

Diagrammatic way to present probability concept is a nice way. The pictorial representation is called ‘Venn diagram.’ If the events are mutually exclusive, their parts will not overlap each other. If two events are not mutually exclusive, their parts will overlap as shown in following figure.



**Additive and Multiplication Laws of Probability**

There are two important theorems of probability namely:

1. The Addition Theorem; and

2. The Multiplication Theorem.

1. **Addition Theorem or Theorem of Total Probability**

**(a) When Events Mutually Exclusive:** If two more events are mutually exclusive, the probability of occurrence of them is the addition of the marginal probabilities of the events.

P (A or B or C) = P (A) + P (B) + P (C)

This can be represented by the Venn diagram as follows:



Suppose

A = getting 1 on throw of a dice

B = getting 2 on throw of the dice

C = getting 3 on throw of the dice

As there are six possible equally likely outcomes on throwing the dice.

P (A or B or C) = 

**Proof of the Theorem:** If an event A can happen in a1 ways and B in2 ways, then the number of ways in which either event can happen is a1 + a2. If the total number of possibilities in n, then by definition, the probability of either the first or the second event happing is:



and

Hence P (A or B) = P (A) + P (B).

There theorem can be extended to three or more mutually exclusive events. Thus,

P (A or B or C) = P (A) + P (B) + P (C)

**Examples**

One card is drawn from a standard pack of 52, what is the probability that it is either a king or a queen?

**(b) When Events are not Mutually Exclusive:** The addition theorem discussed above is not applicable where the events are not mutually exclusive. **For example,** if the probability of a mans’s buying a ready-made shirt is 0.65 and that of buying a ready-made trousers is 0.25, we cannot calculate the probability of his buying either a shirt or a trousers by adding the two probabilities because the events are not mutually exclusive.

When events are not mutually exclusive, the addition theorem has to be modified. The probability that at least one of these two events A and B which are not mutually exclusive will occur, is:

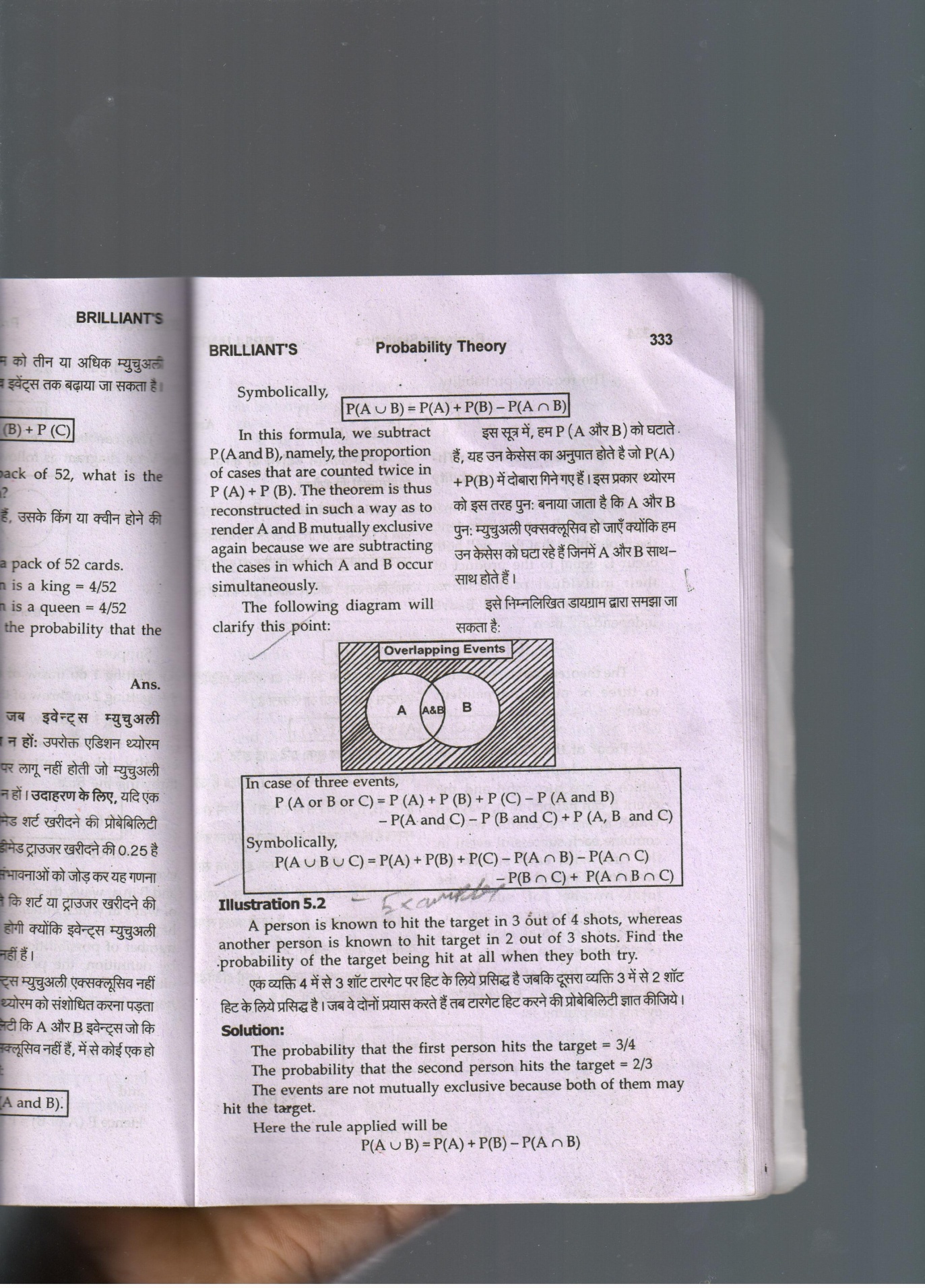
P (A or B) = P (A) + P (B) – P (A and B).

Symbolically,

P (A ∪ B) = P (A) + P(B) – P(A ∩ B)

In this formula, we subtract P (A and B), namely, the proportion of cases that are counted twice in P (A) + P (B). The theorem is thus reconstructed in such a way as to render A and B mutually exclusive again because we are subtracting the case in which A and B occur simultaneously.

The following diagram will clarify this point:



In case of three events,

P (A or B or C) = P (A) + P (B) + P (C) – P (A and B)

– P (A and C) – P (B and C) + P (A ,B and C)

Symbolically,

P (A ∪ B ∪ C) = P(A) + P(B) + P(C) – P(A ∩ B) – P(A ∩ C)

– P(B ∩ C) + P(A ∩ B ∩ C)

Example

A person is known to hit the target in 3 out of 4 shots, whereas another person is known to hit target in 2 out of 3 shots. Find the probability of the target being hit at all when they both try.

1. **Multiplication Theorem of Theorem of Compound Probability**

This theorem states that if two events A and B are independent, the probability that they

will both occur is equal to the product of their individual probabilities. Symbolically, if A and B are independent, then

P (A and B) = P (A) × P (B)

The theorem can be extended to there or more independent events.

P (A, B and C) = P (A) × P (B) × P (C)

**Proof of the Theorem:** If an events A can happen in n1 ways of which a1 are successful and the event B can happen in n2 ways of which a2 are successful, we can combine each successful event in the first with each successful event in the second case, Thus, the total number of successful happening in both cases is a1 × a2. Similarly, the total number of possible cases is n1 × n2.

Then, by definition the probability of both independent events happing is:



But  = P (A) and  = P (B).

P (A and B) = P (A) × P (B).

In a similar way, the theorem can be extended to three or more events.

If we are given independent events. A1, A2, A3 …… An with respective probabilities of occurrences as P1, P2, P3 …..Pn, then the probability of occurrence of at least one of the n events A1, A2, A3…An can be determined as follows:

P (happening of a least one of the events+

= 1 – P (happening of none of the events)

The following example shall illustrate the application of the above principle.

**Example**

A problem in regression analysis is given to four students, A, B, C and D. Their chances of solving it are 1/2, 1/3, ¼. What is the probability that the problem will be solved?

**Conditional Probability in case of independent Event**

If the probability of an event is subject to a restriction on the sample space, the probability is said to be conditional. Conditional probability is the probability of the occurrence of an event, say A, subject to the occurrence of a previous event, say B. We define the conditional probability of event A, given that B has occurred. In case of A and B being independent events, as the probability of event A.

P (A/B) = P (A)

It is so because independent events are those whose probabilities are in no way affected by the occurrence of each other.

**Conditional Probability in case of Dependent Events**

We can define the conditional probability of event A, given that event B occurred when both A and B are dependent events, as the ratio of the number of elements common in both A and B to the number of elements in B.



Where, P (A ∩ B) = P (A) + P(B) – P(A ∪ B)

Because, P (A ∪ B) = P (A) + P(B) – P (A ∩B)

Similarly, 

The general rule of multiplication in its modified form in terms of conditional probabilities becomes:

P (A and B) = P (B) × P (A/B)

or P (A and B) = P (A) × P (B/A)

P (ABC) = P (A) × P (B| A) × P (C | AB)

i.e., the probability of occurrence of A, B and C is equal to the probability of A times the probability of B given that A has occurred, B times the probability of C given that both A and B have occurred.

**Example**

The chances that doctor A will diagnose a disease X correctly is 60%. The chance that a patient will die by his treatment after the correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. A patient of doctor A, who had disease X died. What is the probability that his disease was diagnosed correctly?

**BAYE’S THEOREM/BAYE’S**

**Baye’s Theorem**

One of the most interesting (and controversial) applications of the result of probability theory is to calculate unknown probabilities and making decisions on the basis of new information.

The Baye’s theorem named after **Thomas Baye** who was an English philosopher. It was published in 1763 and it has become one of the most famous theorems in the history of science. His contribution was mainly to provide a unique method of calculating conditional probabilities. Baye’s rule is concerned with determining the probability of an event on the basis of certain sample information. **For example,** if there are 2 defective items in 50 trials (event A), it might be used to estimate the probability that a machine is not working correctly (event Ei) or on the basis of results of any one subject (event A), we can estimate the probability of getting first division in the examination, etc.

Suppose that a personnel administrator wishes to hire one person from among a number of job applicants for a clerical position. The job to be filled is fairly simple. On the basis of past experience, the personnel director feels that there is a 80% probability of an applicant being able to fill the position. This probability is the prior probability.

A personnel administrator usually interviews or tests each applicant, rather than selecting one at random. This procedure supplies additional direct information about the applicant. In light of this additional information, the personnel director may revise the prior probability about an applicant’s chance for success or failure at the job. The revised probability is the posterior probability.

The terms prior and posterior refer to the time when information is collected. Before information is obtained, we have prior probabilities. Baye’s theorem provides a means of calculating posterior probabilities example illustrates the use of Baye’s theorem.

A personnel administrator feels that an applicant has the following chances for success and for failure in a given job opening.

P (success) = 0.8

These probabilities are prior probabilities. They are the probabilities. They are the probabilities of success and failure prior to obtaining specific information about an applicant.

During the test and interview period, the administrator assigns each applicant a rating of 1 (above average), 2(average), or 3 (below average.) The records show that applicants who turn out to be successful have the following chances of receiving 1,2, or 3 rating:

P(1/success) = 0.7

P (2/success) = 0.2

P (3/ success) = 0.1

The records also show the chances of unsuccessful candidates receiving 1, 2, or 3 rating:

P(1/ failure = 0.1

P (2/ failure) 0 .3

P (3/ failure) = 0.6

These last six values are conditional probabilities. The 0.6 represents the probability receiving 3 rating given that a candidate is a potential failure for the position.

The personnel administrator will use Baye’s theorem to find successful candidates for each job on the basis of the applicant’s rating.

In the above example, a person applying for the position can succeed with either 1, 2 or 3 rating. Given a particular rating, the personnel administrator must determine the probability of the applicant being successful in that position.

We use the multiplication rule to find the joint probability of an applicant having a rating of 1 and being successful in the position.

P (1 and success) = P (1/ success)

Computation of similar joint probabilities shows

P (1 and success) = 0.56

P (2 and success) = 0.16

P (3 and success) = 0.08

These results, in turn, allow us to compute the probability of an applicant receiving 1, 2, or 3 rating during an interview. Use of the addition rule shows:

P (1) = P (1 and success) + P (1 and failure)

= 0.56 + 0.02 = 0.58

P (2) = 0.22

and P (3) = 0.20

Suppose that a particular applicant is given 1 rating. Baye’s theorem allows us to calculate the probability that the applicant will be successful. The posterior probability of an applicant being a success given that 1 rating was assigned is calculated by the conditional probability formula as follows:



The formula says that the probability of a success given 1 rating is equal to the probability of 1 rating and a success divided by the probability of 1 rating. By substitution, we have:



Similarly, the posterior probability of a failure given 1 rating is:



Thus, when an applicant has a 1 rating, the probability of being a potential success increases from 0.8 (the prior probability) to 0.97 (the posterior probability). So the rating system appears to be reliable.

Similarly computations show that the posterior probabilities for 2 ratings are:



For 3 ratings, the posterior probabilities are:



Baye’s theorem is a version of the multiplication rule that we studied previously, namely,

P(A and B) = P (A) × P (B/ A)

P (A and B) = P(B) × P (A/ B)

or

Baye’s theorem is merely this rule in the form



Baye’s theorem allows us to revise prior probabilities in light of direct information. The results of the formula are posterior probabilities.

It is clear from the above example that Baye’s theorem provides a powerful method in improving the quality of probability in decision making under uncertainty. The more evidence, we accumulate, the less important are the prior probabilities. The only restriction on the application of Baye’s rule is that all hypotheses must be applicable in a given situation and that none is assigned a prior probability of 0 or 1.