**Basics of Statistics**

Topics covered

* What is statistics?
* Importance & Scope of Statistics
* Limitations of Statistics
* Branches of Statistics
  + Descriptive Statistics
  + Inferential Statistics
* Measures of Central Tendency
* Dispersion or Variation
* Skewness
* Kurtosis
* Correlation

1. What is Statistics?

* Statistics has been defined differently by different authors from time to time but none of them is adequate. Some of the popular definitions are:
* Statistics is the science which deals with the method of collecting, classifying, presenting, comparing and interpreting numerical data collected to throw some light on any sphere of enquiry.
* Modern statistics refers to a body of methods and principles that have been developed to handle the collection, description, summarization and analysis of numerical data. Its primary objective is to assist the researcher in making decisions or generalizations about the nature and characteristics of all the potential observations under consideration of which the collected data form only a small part.

2. Importance & Scope of Statistics:

* In modern times, Statistics is viewed not as a mere device for collecting numerical data but as a means of developing sound techniques for their handling and analysis and drawing valid inferences from them.
* It is now finding wide applications in almost all sciences - social as well as physical- such as biology, psychology, education, economics, business management, and etcetera.
* We now discuss briefly the importance of Statistics in some different sectors and disciplines.
* **Statistics and Planning**: It helps in planning and formulating policy decisions to almost all organizations in the government or managements of business
* **Statistics and Mathematics**: Statistics is intimately related to and essentially dependent upon mathematics.
* **Statistics and Economics:** Statistical data and technique of statiistical analysis have' proved immensely useful in solving a variety of economic problems, such as wages, prices, analysis of time series and demand analysis.
* **Statistics and Business:** Business executives are relying more and more on statistical techniques for studying the needs and the desires of the consumers and for many other purposes.
* **Statistics and Biology, Astronomy and Medical Science:** It is widely used in Biology, Astronomy and Medical field as well.
* **Statistics and War**: In war, the theory of **'Decision Functions'** can be of great assistance to military and technical personnel to plan 'maximum destruction with minimum effort'.
* Thus, we see that the science of Statistics is associated with almost all the sciences - social as well as physical.

3. **Limitations of Statistics**: Statistics, with its wide applications in almost every sphere of human activity is not without limitations. The following are some of its important limitations:

* **Statistics is not suited to the study of qualitative phenomenon**: Statistics, being a science dealing with a set of numerical data, is applicable to the study of only those subjects of enquiry which are capable of quantitative measurement. As such; qualitative phenomena like honesty, poverty, culture, etc., which cannot be expressed numerically, are not capable of direct statistical analysis. However,

Statistical techniques may be applied indirectly by first reducing the qualitative expressions to precise quantitative terns. For example, the intelligence of a group of candidates can be studied on the basis of their scores in a certain test.

* **Statistics does not study individuals:** Statistics deals with an aggregate of objects and does not give any specific recognition to the individual items of a series. Individual items, taken separately, do:not constitute statistical data and are meaningless for any statistical enquiry. For example, the individual figures of agricultural production, industrial output or national income of ~y. country for a particular year are meaningless unless, to facilitate comparison, similar figures of other countries or of the same country for different years are given. Hence, statistical analysis is suited to only those problems where group characteristic are to be studied
* **Statistical laws are not exact:** Unlike the laws of physical and natural sciences, statistical laws are only approximations and not exact. On the basis of statistical analysis we can talk only in terms of probability and chance and not in terms of certainty. Statistical conclusions are not universally true - they are true only on an average. For example, let us consider the statement:" It has been found that 20 % of-a certain surgical operations by a particular doctor are successful. The statement does not imply that if the doctor is to operate on 5 persons on any

day and four of the operations have proved fatal, the fifth must bea success. It may happen that fifth man also dies of the operation or it may also happen that of the five operations on any day, 2 or 3 or even more may be successful. By the statement we mean that as number of operations becomes larger and larger we should expect, on the average, 20 % operations to be successful.

* **Statistics is liable to be misused:** Perhaps the most important limitation

of Statistics is that it must be used by experts. As the saying goes," Statistical

methods are the most dangerous tools in the hands of the inexperts. Statistics is

one of those sciences whose adepts must exercise the self-restraint of an artist."

The use of statistical tools by inexperienced and untrained persons might lead to

very fallacious conclusions. One of the greatest shortcomings of Statistics is that

they do not bear on their face the'label of their quality and as such can be moulded and manipulated in any manner to support one's way of argument and reasoning. As King says," Statistics are like clay of which one can make a god or devil as one pleases." The requirement of experience and skin for judicious use of statistical methods restricts their use to experts only and limits the chances of the mass popularity of this useful and important science.

4. **Branches of Statistics**:

5. **Measures of Central Tendency:** The tendency to concentrate at certain values, usually somewhere in the centre of the distribution. Measure of this tendency are called measures of central tendency or averages.

Plainly speaking, an average of a statistical series is the value of the variable which is representative of the entire distribution.

The following are the five measures of central tendency that are in common use:

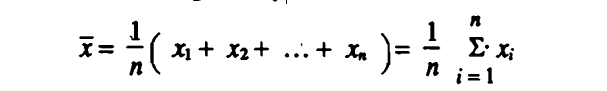
1. Arithmetic Mean or Simple Mean

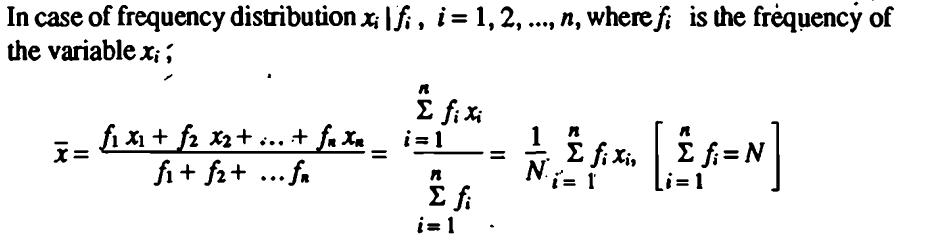
2. Median

3. Mode

1. **Arithmetic Mean or Simple Mean**:

Arithmetic mean of a set of observations is their sum divided by the number of observations.



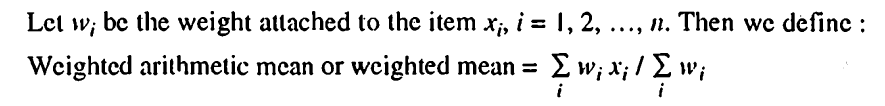


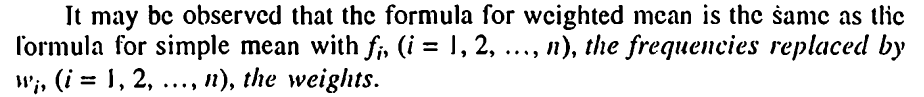
Note: In case of grouped or continuous frequency distribution. *X* is taken as the mid Value of the corresponding class.

Question: I want to compute the average price of below commodities such that **wheat, Rice and pulse are more important than Cigarettes, Tea and Sweets**. Can I use Arithmetic Mean ?

|  |  |
| --- | --- |
| Commodities | Price |
| Wheat | 30 |
| Rice | 50 |
| Pulse | 90 |
| Cigarettes | 200 |
| Tea | 150 |
| Sweets | 300 |

1.1 **Weighted Mean:** In calculating arithmetic mean we suppose that all the items in the distribution have equal importance. But in practice this may not be so. If some items in a distribution are more important than others, then we calculate weighted mean. Weights attached to each item are proportional to the importance of the item in the distribution.





Note: Weighted mean gives the result equal to the simple mean if the weights assigned to each of the variate values arc equal. It results in higher value than the simple mean if smaller weights are given to smaller items and larger weights to larger items. If the weights attached to larger items are smaller and those attached to smaller items are larger, then the weighted mean results in smaller value than the simple mean.

Merits and Demerits of Arithmetic Mean:

Merits:

1. Arithmetic mean rigidly defined by algebraic formula

2.  It is easy to calculate and simple to understand

3. It is based on all observations and it can be regarded as representation of the given data

4. It is capable of being treated mathematically and hence it is widely used in statistical analysis.

5. Arithmetic mean can be computed even if the detailed distribution is not known but some of the observation and number of the observation are known.

Demerits:

1. It cannot be determined by inspection nor it can be located graphically.

2. Arithmetic mean cannot be used if we are dealing with qualitative characteristics which cannot be measured quantitatively; such as, intelligence, honesty, beauty, etc. In such cases median (discussed later) is the only average to be used.

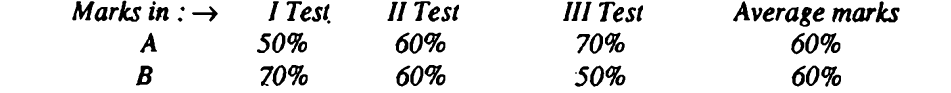
3. Arithmetic mean cannot be obtained if a single observation is missing or lost or is illegible unless we drop it out and compute the arithmetic mean of the remaining values.

4. Arithmetic mean is affected very much by extreme values. In case of extreme items, arithmetic mean gives a distorted picture of the distribution and no longer remains representative of the distribution.

5. Arithmetic mean cannot 'be calculated if the extreme class is open, *e.g.,* below 10 or above90. Moreover, even if a single observation is missing mean cannot be calculated.

6. In extremely asymmertrical (skewed) distribution, usually arithmetic mean is not a suitable measure of location.

7. Arithmetic mean may lead to wrong conclusions if the details of the data from which it is computed are not given. Let us consider the following Marks obtained by two students A and *B* in three tests, *viz.,* terminal test, half-yearly examination and annual examination respectively.



Thus average marks obtained by each of the two students at the end of the year are 60%. If we are given the average marks alone we conclude that the level of intelligence of both the students at the end of the year is same. This is a fallacious conclusion since we find from the data that student A has improved consistently while student B has deteriorated consistently.

**2. Median:**

The median is that value of the series which divides the group into two equal parts, one part comprising all values greater than the median value and the other part comprising all the values smaller than the median value. The median is called positional average.

**How to calculate**:

In case of ungrouped data, if the number of observations is odd then median is the middle value after the values haye been arranged in ascending or descending order of magnitude.

In case of even number of observations, there are two middle terms and median is obtained by taking the arithmetic mean of the middle terms.

For example, the median of the values 25, 20,15,35,18, *i.e.,* 15, 18, 20, 25, 35 is 20 and the median of 8, 20, 50, 25, 15, 30, *i.e.,* of 8, 15, 20, 25, 30, 50' is ( 20 + 25 )/2 = 22·5 .

**Remark:** In case of even number of observations, in fact any value lying between the two middle values can be taken as median but conventionally we take it to be the mean of the middle terms.

Merits and Demerits of Median:

Merits:

1. It is rigidly defined.

2. It is easily understood and is easy to calculate. In some cases it can be located merely by inspection.

3. It is not at all affected by extreme values.

4. It can be calculated for distributions with open-end classes.

Demerits:

1. In case of even number of observations median cannot be determined exactly. We ,merely estimate it by taking the mean of two middle terms.

2. It is not based on all the observations. For example, the median of 10, 25, 50,60 and' 65 is 50. We can replace the observations 10 and 25 by any two values which are smaller than 50 and the observations 60 and 65 by any two values greater than 50 without affecting the value of median. This property is sometimes described by saying that median is *insensitive.*

3. It is not amenable to algebraic treatment. For example, multiplying the median with the number of items in the series will not give us the sum total of the values of the series.

**Uses of Median:**

**1.** Median is the only average to be used while dealing with qualitative data which cannot be measured quantitatively but still can be arranged in ascending or descending order of magnitude, *e.g.,* to ind the average intelligence or average honesty among a group of people.

2. It is to be used for determining the typical value in problems concerning wages, distribution of wealth, etc.

**3. Mode:**

**Let us cosider the following statements:**

***(i)*** The average height of an Indian (male) is 5'-6".

*(ii)* The average size of tile shoes sold in a shop is 7.

*(iii)* An average student in a hostel spends Rs.1000 p.m.

In all the above cases, the average referred to is mode. Mode is the value which occurs most frequently in a set of observations and around which the other items

of the set cluster densely. In other words, mode is the value of the variable which is predominant in the series. Thus in the case of discrete frequency distribution mode is thevalue of X corresponding to maximum frequency. For example, in the following frequency distribution:

****

the value of X corresponding to the maximum frequency, *viz.,* 25 is 4. Hence mode is 4**.**

**Merits and Demerits:**

**Merits:**

1. Mode is readily comprehensible and easy to calculate. Like median, mode can be located in some cases merely by inspection**.**

2. Mode is not at all affected by extreme values**.**

3. Mode can be conveniently located even if the frequency distribution has class-intervals of unequal magnitude provided the modal class and the classes preceding and succeeding it are of the same magnitude. Open-end classes also do not pose any problem in the location of mode.

**Demerits:**

1. Mode is ill-defined. It is not always possible to find a clearly defined mode. In some cases, we may come across distributions with two modes.

2. It is not based upon all the observations.

3. It is not capable of further mathematical treatment.

**Uses:**

1. Mode is the average to be used to find the ideal size, examplein business forecasting, In the manufacture of ready-made garments, shoes, etcetera.

**6. Dispersion or Variation:**

* **Why do we need Dispersion:** Averages or the measures of central tendency give us an idea of tile concentration of the observations about the central part of the distribution. If we know the average alone we cannot form a complete idea about the distribution as will be clear from the following example.

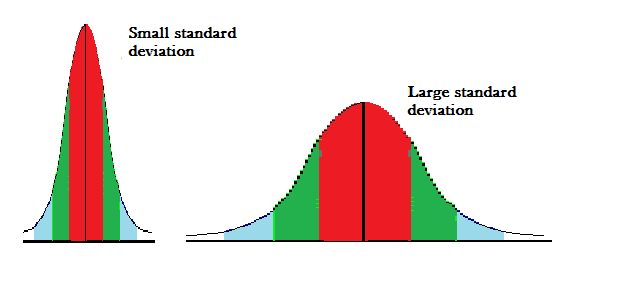
Consider the series *(i)* 7, 8, 10, 11, *(ii)* 3, 6, 9, 12, 15, *(iii)* I, 5, 9, 13, 17.

In all these cases we see that n, the number of observations is 5 and the mean is 9. If we are given that the mea n of 5 observations is 9, we cannot forma an idea as to whether it is the average of first series or second series or third series or of any other series of 5 observations whose sum is 45. Thus we see that the measures of central tendency are inadequate to give us a complete idea of the distribution. They must be supported and supplemented by some other measures, One such measure is *Dispersion.*

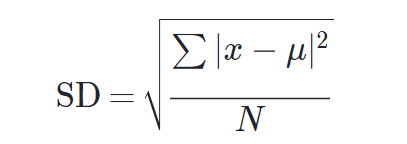
* Literal meaning of dispersion is *·scatteredness'.*
* We study dispersion to have an idea about the homogeneity or heterogeneity of the distribution.
* In the above case we say that series *(i)* is more homogeneous (less dispersed) than the series *(ii)* or *(iii)* or we say that series *(iii)* is more heterogeneous (more scattered) than the series *(i)* or *(ii),*

**6.1 Standard Deviation:**

* Square root of variance is called standard deviation
* Standard deviation, usually denoted by the Greek letter small sigma (σ), is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean.
* What does it actually mean?
* It tells you how tightly your data is clustered around the mean. When the bell curve is flattened (your data is spread out), you have a large standard deviation — your data is further away from the mean. When the bell curve is very steep, your data has a small standard deviation — your data is tightly clustered around the mean. For example, the graph on the left might represent an abnormally high number of students getting scores close to the average, while the graph on the right represents more students getting scores *away* from the average.

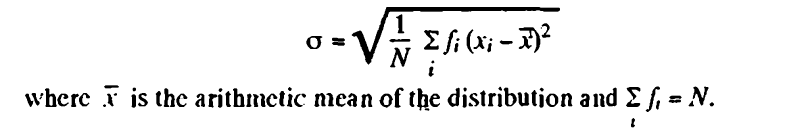


* How to calculate?
* The formula for standard deviation (SD) is for an ungrouped data:



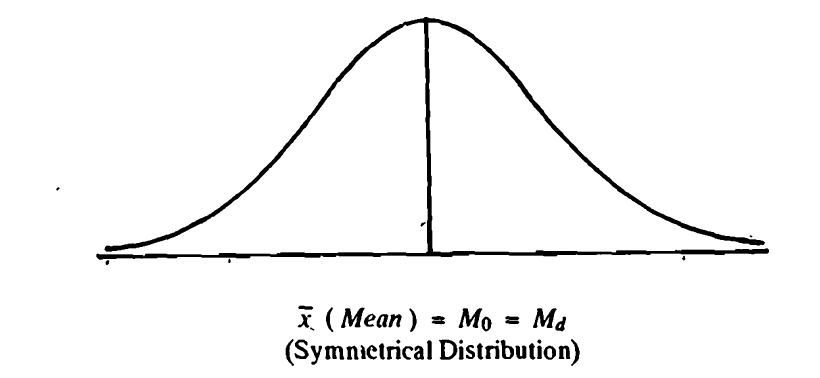
Where ∑ means "sum of", X is a value in the data set, *μ* is the mean of the data set, and N is the number of data points in the population.

* For the frequency distribution *Xi/*fi;, i = 1,2, ... , *n, is calculates a*

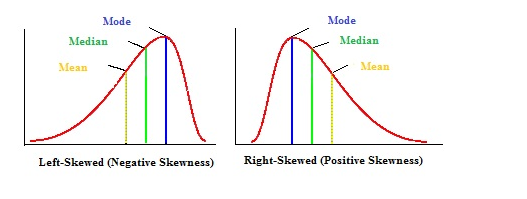


**7. Skewness:**

* Literal meaning of skewness is 'lack of symmetry·.
* We study to have an idea about the shape of the curve which we can draw with the help of the given data, A distribution is said to be skewed if
  + Mean, median and mode fall at different points. *i.e.,* Mean not equal to Median not equal to Mode,
  + Quartiles are not equidistant from median. And
  + The curve drawn with the help of the given data is not symmetrical but stretched more to *one* side than to the oher.
* Symmetrical Distribution:

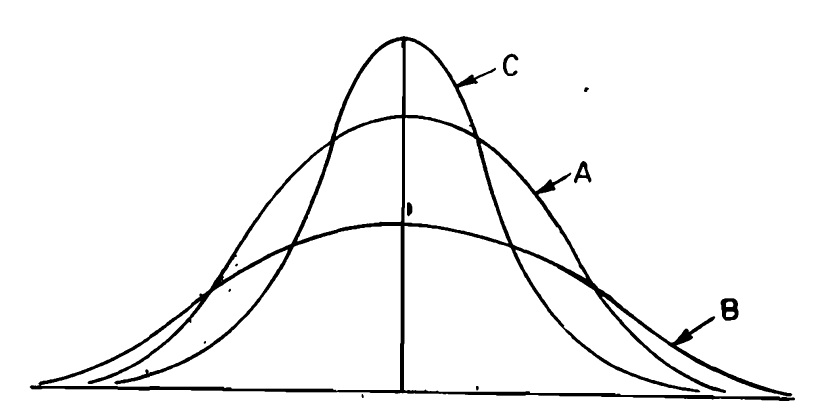


Asymmetrical or Skewed Distribution:



**8. Kurtosis:**

* If we know the measures of central tendency, dispersion and skewness, we still cannot form a complete idea about the distribution' as will be clear from the following figure in which all the three curves *A,* B and Care symmetrical about the meanand have the same range.

****

* Kunosis enables us to have an idea about the flatness or peakedness of the curve.
* Curve of the type *'A'* which is neither flat nor peaked is called the *normal curve or mesokurtic curve.* Curve of the type' *B'* which is flatter than the normal curve is known as *platykurtic* . Curve of the type 'C' which is more peaked than the normal curve is called *leptokurtic.*

9. **Correlation:**

Often we come across situations in which our focus is simultaneously on two or more variables and invariably, we observe that movements in one variable are accompanied by movements in other variable. For example Income and expenditure on households or price and demand of commodities, exhibit accompanying movements of two variables.

**Meaning of Correlation:**

In a bivariate distribution we may be interested to find out if there is any correlation or covariation between the two variables under study. It the change in one variable affects a change in the other variable, the variables are said to be correlated. If the two variables deviate in. the same direction, *i.e .,* if the increase (or decrease) in one results in a corresponding increase (or decrease) in the other, correlation is said to be *direct* or *positive.* But if they constantly deviate in the opposite directions, *i.e.,* if increase (or decrease) in one results in corresponding decrease (or increase) in the other, correlation is said to be *diverse* Of *negative.* For example, the correlation between *(l)* the heights and weights of a group of persons, *(ii)* the income and expenditure is positive and the correlation between *(i)* price and demand of a commodity, *(ii)* the volume and pressure of a perfect gas, is negative. Correlation is said to be *perfect* if the deviation in one variable is followed by a corresponding and proportional deviation in the other.

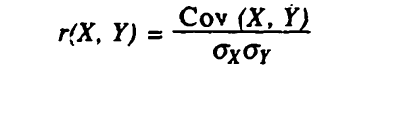
**Scatter Diagram:**

It is the simplest way of the diagrammatic representation of bivariate data. Thus for the' bivariate distribution *(Xi, y;); i* = I, 2, ... , *n.* if the values of the variables *X* and Y are plotted along the x-axis and y-axis respectively in the *xy* plane, the diagram of dots so obtained is known as *scatter diagram.* From the scatter diagram, we can form a fairly good, though vague, idea whether the variables are correlated or not, *e.g..* if the points are very dense, *i.e.,* very close to each other, we should expect a fairly good amount of correlation between the variables and if the ,points are widely scattered, a poor correlation is expected. This method however is not suitable if the number of observations is fairly large.

**Karl Pearson Coefficient of Correlation:**

It is a measure of intensity or degree of linear relationship between two variables.

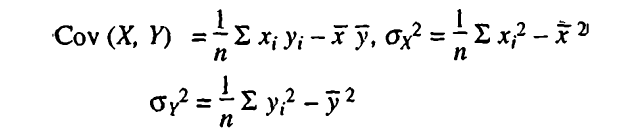
Correlation coefficient between two random variables X and Y, usually denoted by *r(X.* Y) is a numerical measure of *linear relationship* between them and is defined as

Where,

Cov(X,Y) = Covariance of X and Y

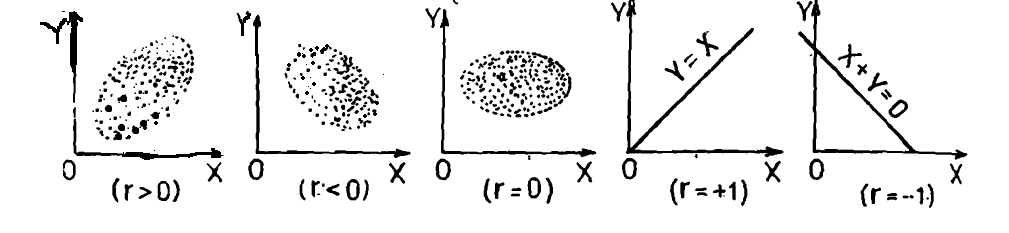
σx = Standard Deviation of X

σy = Standard Deviation of Y



**Some Important Points:**

**1.** Following are the figures of the standard data for *r>* 0, <'0, = 0, and *r* = ± I.

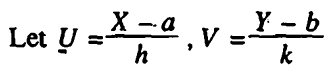


2. It may be noted that r (X, Y) provides a measure of linear relationship between X and Y. For nonlinear relationship, however, it is not very suitable and correlation ration used in that case.

3. Karl Pearson's correlation coefficient is also called product-moment correlation coefficient.

4. Correlation coefficient cannot exceed unity numerically. It always lies between -1 and + 1. If r = + 1, "the c6rtelation is perfect and positive and if *r* = -1, correlation is perfect and negative'.

5*.* Correlation coefficient is independent of change of origin and scale.

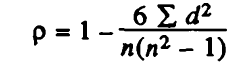
* We can prove that*

**

**Rank Correlation:**

Let us suppose that a group of n individuals is arranged in order of merit or proficiency in possession of two characteristics A and B. These ranks in the two characteristics will in general be different. For example, if we consider the relation between intelligence and beauty. it is not necessary that a beautiful individual is intelligent also. Let (Xi. Yi); i = 1. 2 … n be the ranks of the ith individual in two characteristics A and B respectively. Pearsonian coefficient of correlation between the ranks Xi's and Yi's is called the rank correlation coefficient between A and B for that group of individuals.

**Spearman rank correlation coefficient:**

****

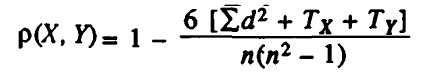
*Where,*

*d = Rank of X – Rank of Y*

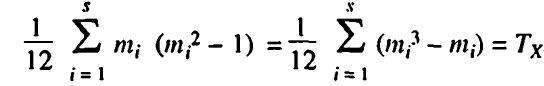
*n = no. of observations*

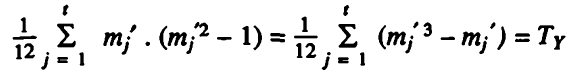
**In case of Tie Rank:**

If some of the individuals receive the same rank in a ranking or merit, they are said to be tied. In this case each of these individuals is assigned a common rank, which is the arithmetic mean of the ranks of tied observations. And the rank correlation coefficient is computed as below:



Where,

 and



**Sampling and Large Sample Theory**

Topics Covered:

1. Introduction or Overview

2. Types of Sampling

3. Parameter and Statistics

4. Tests of Significance

5. Procedure for testing of hypothesis

6. Concept of P - Values

1. **Introduction or Overview of Sampling:**

Before giving the notion of sampling we will first define *population.*

**Population:** in statistics, population is an aggregate of objects, animate or inanimate or group of individuals, under study. The population may be finite or infinite.

**Need For Sampling:**

* It is obvious that, for any statistical investigation complete enumeration of the population is rather impracticable.

For example, if we want to have an idea of the average per capita (monthly) income of the people in India, we will have to enumerate all the earning individuals in the country which is rather a very difficult task.

* If the population is infinite (Very Large Data), complete enumeration is not possible.

**Sample:** A finite subset of statistical individuals in a population is called a *sample* and the number of individuals in a sample is called the sample size.

**Use of Sampling in our day to day life:**

* It is quite often used in our day-to-day practical life. For example, in a shop we assess the quality of sugar, wheat or any other commodity by taking a handful of it from the bag and then decide to purchase it or not.
* A housewife normally tests the cooked products to find, if they are properly cooked and contain the proper quantity of salt.

**2. Types of Sampling:** Some of the commonly known and frequently used types, of sampling are

* Purposive sampling
* Simple Random sampling
* Stratified sampling
* *Systematic Sampling.*
* **Purposive Sampling:**
* Purposive sampling is one in which the sample units are selected with definite purpose in view.
* For example, if we want to give the picture that the standard of living has increased in the city of New Delhi, we may take individuals in the sample from rich and posh localities like Defence Colony, South Extension, Golf Links, Jor Bagh, Chanakyapuri, Greater Kailash etc. and ignore the localities where low income group and the middle class families live.
* This sampling suffers from the drawback of favouritism and nepotism and does not give a representative sample of the population.
* Simple Random sampling:
* A random sampling in which each unit of the population has an equal chance, say *p,* of being included in the sample and that this probability is independent of the previous drawings.
* Thus a simple random sample of size n from a population may be identified with a series of n independent trials with constant probability *'p'* of success for each trial.
* Stratified sampling:
* In this type of sampling entire heterogeneous population is divided into a number of homogeneous groups usually termed as *strata,* which differ from one another but each of these groups is homogenous within itself.
* Then units are sampled/Selected at random from each of these stratum, the sample size in each stratum varies according to the relative importance of the stratum in the population.
* Such a sample is by far the best and can safely be considered as representative of the population from which it has been drawn.
* Systematic Sampling:

**3. Parameter and Statistic:**

* Statisticalconstants of the population are called parameters**.** For example: Population mean and population variance.
* Statistic may be regarded as an estimate of parameter.
* It is obtained from the sample that it is a function of the sample values only. For example: Sample mean and sample variance.

****

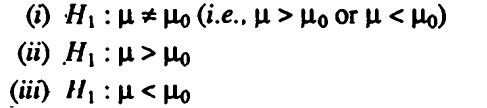
**4. Tests of Significance**: A very important aspect of the sampling theory is the study of the *tests of significance,* which enable us to decide on the basis of the sample results, if

(i*)* The deviation between the observed sample statistic and the hypothetical parameter value, or

(ii) The deviation between two independent sample statistics; is significant or might be attributed to chance or the fluctuations of sampling.

(iii) Since for large sample almost all the distributions can be approximated very closely by a normal probability curve. So we will study here the Normal Tests of Significance.

* **Null Hypothesis:** For applying the test of significance we first set up a hypothesis. Null hypothesis is a definite statement about the population parameter. Such a hypothesis, which is usually a hypothesis of no difference, is called *null hypothesis* and is usually denoted by ***Ho****.* For Example: the null hypothesis that the population has a specified mean μ0.
* **Alternative Hypothesis:** Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by ***H1.*** The alternative hypothesis could be

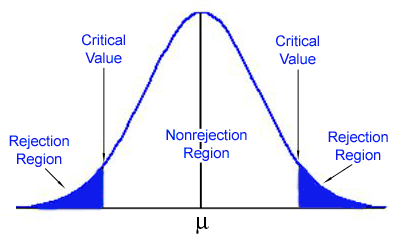


The alternative hypothesis in (i) is known as a two *tailed alternative* and the alternatives in (ii) and *(iii)* are known as *right tailed* and *left-tailed alternatives* respectively.

The setting of alternative hypothesis is very important since it enables us to decide whether we have to use a single-tailed (right or left) or two tailed test.

**Note:** In a particular problem, whether one tailed or two tailed test is to be applied depends entirely on the nature of the alternative hypothesis. If the alternative hypothesis is two-tailed we apply two-tailed test and if alternative hypothesis is one-tailed, we apply one tailed test.

* **Errors in Sampling:** The main objective in sampling theory is to draw valid inferences about the population parameters on the basis, of the sample results. In practice we decide to accept or reject the lot after examining a sample from it As such we are liable to commit the following two types of errors:
* **Type I Error:**
* It is denoted by alpha α
* In practice, type 1 error amounts to rejecting a tot when it is good.
* It is referred to as producer’s risk
* Mathematically, type 1 error is P(Reject *Ho* when it is true}
* **Type** II **Error:**
* It is denoted by alpha β
* In practice, type 2 error amounts to accepting the lot when it is bad.
* It is referred to as consumer’s risk
* Mathematically, type 1 error is *P* {Accept *Ho* when it is wrong}
* **Critical Region or Rejection Region and Level of Significance:**
* A region (corresponding to a statistic *t)* in the sample space which amounts to rejection of *null hypothesis* is termed as *critical region* or *region of rejection:*
* The level of significance is the size of the type I error **α** (or the maximum producer's risk). The levels of significance usually employed in testing of hypothesis are *5%* and 1%.
* The level of significance is always fixed in advance before collecting the sample information.



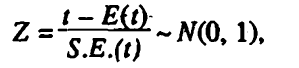
* In the right tailed test the critical region lies entirely in the right tail of the sampling distribution*.*
* In the left tailed test the critical region lies entirely in the left tail of the sampling distribution*.*
* In the two tailed test critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic
* **Critical Values or Significant Values:**

The value of test statistic which separates the critical (or rejection) region and acceptance region is called the *critical value* or *significant* value. It depends upon:

*(I)* The level of significance used, and

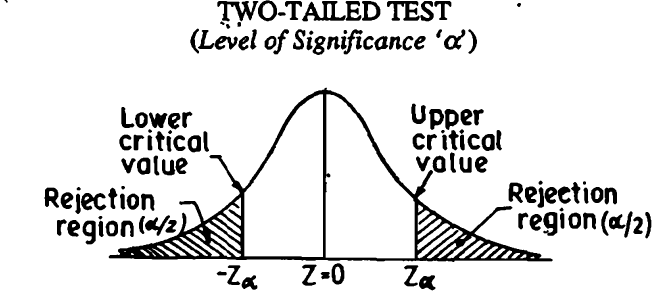
*(ii)* The alternative hypothesis, whether it is two-tailed or single-tailed.

For large sample the standardized variable corresponding to the statistic *t*

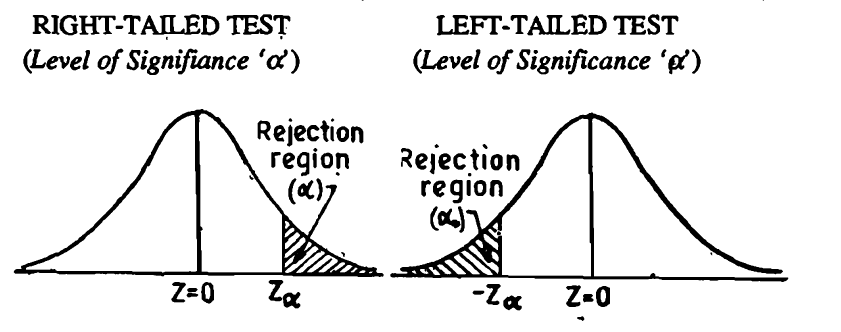


The value of Z under the null hypothesis is known as test statistic. The critical value of the test statistic at level of significance α for a two-tailed test is given by Zα where Zα, is determined by the equation

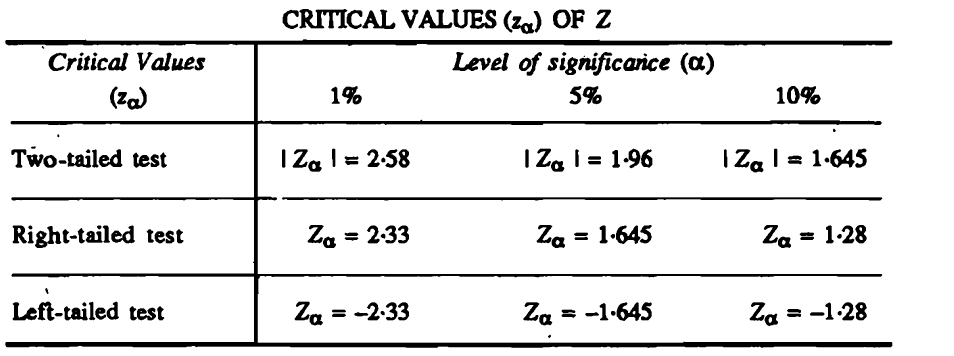




In case of single-tail alternative, the critical value Zα is determined so that total area to the right of it (for right-tailed test) is α and for left-tailed test the total area to the left of -Zα is α.

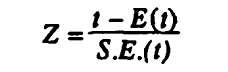


**Commonly used critical value and their values:**



**5. Procedure for testing of hypothesis:** We now summarize below the various steps in testing of a statistical hypothesis in a systematic manner.

* Null Hypothesis. Set up the Null Hypothesis Ho
* Alternative Hypothesis. Set up the Alternative Hypothesis H1 . This will enable us to decide whether we have to use a single-tailed (right or left) test or two-tailed test.
* Level of Significance. Choose the appropriate level of Significance (α) depending on the reliability of the estimates and permissible risk. This is to be decided before sample is drawn, i.e.. α is fixed in advance.
* Test Statistic (or Test Criterion). Compute the test statistic under the null hypothesis



* *Conclusion.* We compare the computed value of Z above with the significant value (tabulated value) Zα *,* at the given level of significance α.
* If I Z 1< Zα*, that is if the calculated value of Z is less than tabulated value of Z we say it is not siginificant.*
* . If I Z I > *l* Zα *that is if the calculated value of Z is greater than tabulated value of Z, we say that it is significant* and the null hypothesis is rejected at level of significance α  *that is* with confidence coefficient (1 - α).

6. **Concept of P – Values:**