SAMPLING THEORY

Introduction :

Sampling theory deals with inductive inference which is " the process by which we draw a conclusion about some measure of a population based on a sample value. The measure might be a variable, such as the average or mean . . . . The purpose of sampling is to estimate some characteristics of the population from which the sample is selected ".

Definitions :

Population :

The population or the Universe in Statistics is used to refer any collection of individuals. The population may be finite or infinite. Eg : Students of SNS

Sample :

A part or small section selected from the population is called sample and the process of such selection is called sampling. Er : Students of mechanical Engg

Sample Size :

The number of individuals in a sample is called the sample size.

Objectives of Sampling :

The main objectives of sampling are :

1. To obtain information about the population on the basis of sample drawn from such population.
2. To set up the limits of accuracy of the estimates of the population parameters computed on the basis of sample statistics.
3. To test the significance about the population characteristics on the basis of sample statistic

Parameter and Statistic :

The various Statistical measures used in Statistics like mean, median and Standard deviation etc., may be computed from population data and also from sample data.

Any Statistical measure computed from Population data is known as parameter and any Statistical measure computed from sample data is known as statistic.

Thus, population mean, population median, population Standard deviation, Population Proportion, etc., are parameters but Sample mean,

Sample median, Sample Standard deviation, Sample Proportion, etc., are called Statistic.

The usual notations used for the parameter in case of population and the statistic in case of sample are given below:

Measures	Population	Sample
Size	N	n
Mean	μ	$\bar{x}$
Standard Deviation	σ	s
Proportion	P	p
Variance	σ^2	s^2

Sampling Distribution:

Sampling distribution of a statistic refers to the distribution of the various values, which can be assumed by that Statistic, computed from the various samples of the same size randomly drawn from the population. It may be noted that from a population of fixed size a large number of samples of the same size may be drawn and each of such a sample may contain different population members, though the samples are all of the same size.

Standard Error :

The statistical measure of standard deviation may be computed both from all observations in a population or from the observations of a sampling distribution. When an average amount of variability of the observation of a population is computed, it is known as standard deviation but an average amount of variability of the observations of a sampling distribution is computed, it is known as Standard error.

Hence the standard deviation computed from the observations of a sampling distribution of a statistic is known as Standard error of the statistic and is denoted by S.E

Utility of Standard error :

1. To determine the precision of the sample estimate of some population parameter, which is given by the reciprocal of the SE of the sampling distribution of the estimate.
2. To test if the sample statistic differ significantly from the corresponding hypothetical value in the population.

TEST OF HYPOTHESIS

Statistical hypothesis :

When we attempt to make decisions about the population on the basis of sample information, we have to make assumptions about the nature of the population involved or about the value of some Parameter of the population. Such assumptions, which may or may not be true, are called Statistical hypothesis.

Null hypothesis :

We set up a hypothesis which assumes that there is no significant difference between the sample Statistic and the corresponding population parameter or between two sample statistics. Such a hypothesis of no difference is called a null hypothesis and is denoted by H_0.

Alternative hypothesis :

A hypothesis that is complementary to the null hypothesis is called an alternative hypothesis and is denoted by H_1.

A procedure for deciding whether to accept or reject the null hypothesis is called the test of hypothesis.

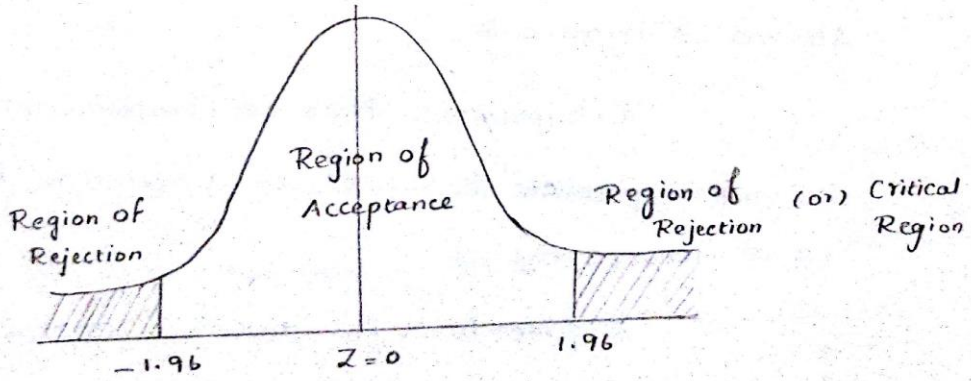
TEST OF SIGNIFICANCE

If θ_0 is the parameter of the population and θ is the corresponding sample statistic, the procedure of testing whether the difference between θ_0 and θ is significant or not is called the test of significance.

Critical Region and Level of Significance:

If we are prepared to accept that the difference between a sample statistic and the corresponding parameter is significant, when the sample statistic lies in a certain region, then that region is called the critical region or region of rejection.

The region complementary to the critical region is called the region of acceptance



Errors in Hypothesis testing :

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 to reject the lot after examining a sample from it. As such we have two type of errors.

(i) **Type I error :**

The error committed in rejecting H_0, when it is really true is called Type I error.

(ii) **Type II error :**

The error committed in accepting H_0, when it is false is called Type II error.

The probabilities of committing Type I and II errors are denoted by α and β respectively.

The sizes of type I (α) and type II (β) are also known as producer's risk and consumer's risk respectively.

One-Tailed and Two-Tailed Tests :

If θ_0 is a population parameter and θ is the corresponding sample statistic and if we set up the null hypothesis $H_0 : \theta = \theta_0$, then the alternative hypothesis H_1 can be any one of the following :

(i) $H_1 : \theta \neq \theta_0$ i.e., $\theta > \theta_0$ or $\theta < \theta_0$

(ii) $H_1 : \theta > \theta_0$

(iii) $H_1 : \theta < \theta_0$

Hence (i) corresponds to 2 tailed alternative hypothesis, (ii) corresponds to right-tailed alternative hypothesis, (iii) corresponds to left-tailed alternative hypothesis.

Critical values or Significant values :

The value of the test statistic z for which the critical region and acceptance region are separated is called the critical value or the significant value of z and is denoted by z_α, where α is the level of significance.

Procedure for Testing of Hypothesis :

1. Null hypothesis H_0 is defined.
2. Alternative hypothesis H_1 is also defined after a careful study of the problem and also the nature of the test is decided.
3. LOS 'α' is fixed or taken from the problem if specified and Z_α is noted.
4. The test-statistic $Z = \dfrac{t - E(t)}{S.E(t)}$ is computed.
5. Comparison is made between $|Z|$ and Z_α.

   If $|Z| < Z_\alpha$, H_0 is accepted or H_1 is rejected i.e., it is concluded that the difference between t and E(t) is not significant at α % LOS.

   On the other hand, if $|Z| > Z_\alpha$, H_0 is rejected or H_1 is accepted, i.e., it is concluded that the difference between t and E(t) is significant at α % L.O.S