UNIT – 3 – TESTING OF HYPOTHESIS

Introduction :

Every statistical investigation aims at collecting information about some aggregate or collection of individuals or of their attributes, rather than the individuals themselves. In statistical language, such a collection is called a population or universe.

A population is finite or infinite, according as the number of element is finite or infinite. In most situations, the population may be considered infinitely large.

A finite subset of a population is called a sample and the process of selection of each samples is called sampling.

The basic objective of the theory of sampling is to draw inference about the population using the information of the sample.

Parameters and Statistics :

Statistical measures such as mean, std, etc are calculated on the basis of population values of x are called Parameters. corresponding measures computed on the basis of sample observations are called Statistics.

Sampling Distribution :

The probability distribution of the statistic that would be obtained if the number of samples, each of same size were infinitely large is called the sampling distribution of the statistic.

The standard deviation of the sampling distribution of a statistic is of particular importance in tests of hypothesis and is called the Standard Error of the statistic

Estimation and Testing of Hypothesis

Some charecteristic or feature of the population in which we are interested may be completely unknown to us and we may like to make a guess about the characteristic entirely on the basis of a random sample drawn from the population. This type of problem is known as the problem of estimation.

Some information regarding the characteristic or feature of the population may be available to us and we may like to know whether the information is tenable in the light of the random sample drawn from the population and if it can be accepted with what degree of Confidence it can be accepted. This type of problem is known as the problem of testing of hypothesis.

Tests of ~~Hypothesis~~ significance

When we attempt to make decisions about the population on the basis of sample information, we have to make assumptions or guesses about the nature of the pop. 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 ~~of~~ population parameter or between two sample statistics.

Such a hypothesis of no difference is called null hypothesis, and is denoted by H_0.

Alternative Hypothesis:

A hypothesis that is different from or (complementary to) the null hypothesis is called an alternative hypothesis, and is denoted by H_1.

Testing of Hypothesis:

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

Critical Region or Region of Rejection:

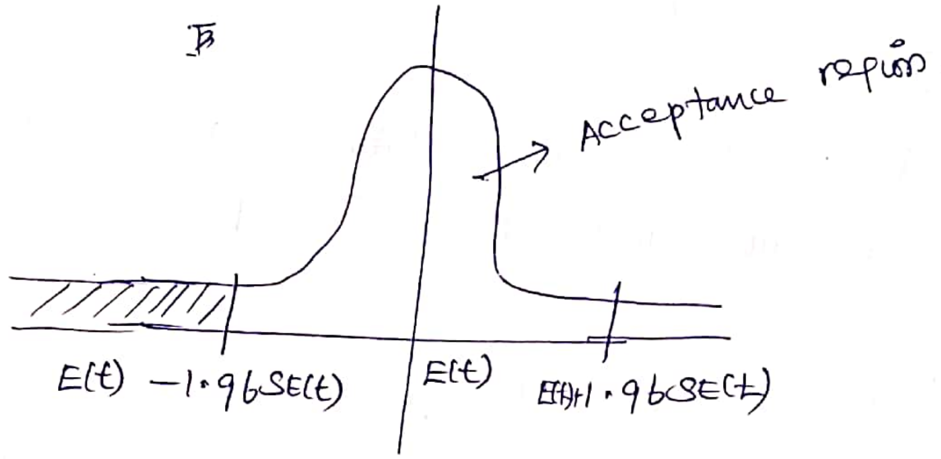
If we are prepared to reject a null hypothesis when it is true or if we are prepared to accept that the difference between a sample statistic and the corresponding parameter is significant, when the sample statistic lies in a certain region or interval, then that region is called critical region or region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

The region complementary to the critical region is called the region of Acceptance.

If 't' is a statistic in large samples, then t follows a normal distribution with mean E(t) and S.D equal to S.E(t). Hence $z = \dfrac{t - E(t)}{S.E(t)}$ is a standard normal variate.

It is known from the study of normal distribution, that the area under the standard normal curve b/w $t=-1.96$ and $t=+1.96$ is 0.95.

Equivalently the area under the general normal curve of 't' between $[E(t) - 1.98SE(t)]$ and $[E(t) + 1.98SE(t)]$ is 0.95.



If we prepared to accept that the difference between t and E(t) is significant when 't' lies in ~~the~~ either of the regions ~~−∞, −1.96~~ $(-\infty, E(t) - 1.96SE(t)]$ and $[E(t) + 1.96SE(t), \infty)$ then these two regions constitute the critical region of 't'.

Level of significance:

The maximum probability with which are prepared to reject Ho when it is true is called LOS. In other words, the total area of the region of region rejection expressed as a percentage is called the LOS.

Type I and Type II errors

Decision	Condition	
	H_0: True	H_0: False
Accept H_0	Correct decision	Type II Error
Reject H_0	Type I Error	Correct decision

One Tailed and Two-Tailed Tests:

$H_0 : \theta = \theta_0$ where $\theta \rightarrow$ Sample statistic

$\theta_0 \rightarrow$ population parameter

$H_1 : \theta \neq \theta_0$ (Two tailed)

$H_1 : \theta > \theta_0$ (Right Tailed)

$H_1 : \theta < \theta_0$ (left Tailed)

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, when α is the LOS.

When $z = \dfrac{t - E(t)}{SE(t)}$,

$P[|z| < 1.96] = 0.95$ and $P[|z| > 1.96] = 0.5$

$z = \pm 1.96$ Separate the critical region and the acceptance region at 5% LOS for a two-tailed test.

LOS / Nature of test	1%	2%	5%	10%
Two tailed	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 2.33$	$\lvert z_\alpha \rvert = 1.96$	$\lvert z_\alpha \rvert = 1.645$
Right tailed	$z_\alpha = 2.33$	$z_\alpha = 2.055$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Left tailed	$z_\alpha = -2.33$	$z_\alpha = -2.055$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

Procedure for Testing of hypotheses

1. Set up a null hypotheses H_0
2. Set up the alternative hypotheses H_1 (Two tailed / left tailed / Right tailed)
3. Set up a suitable significance level (α).
4. The test statistic $z = \dfrac{t - E(t)}{SE(t)}$ is computed
5. Conclusion:

   compare the computed value of $|z|$ with the tabulated value z_α at the LOS α.

   If $|z| < z_\alpha$, H_0 is accepted

   If $|z| > z_\alpha$, H_0 is rejected.