Testing of hypothesis: Modern theory of probability plays an important role in decision making and the branch of statistics which helps us in arriving at the criterion for such decision is known as Testing of hypothesis. It employs statistical techniques to arrive at decision in certain situations where there is an element of uncertainity on the basis of sample, whose size is fixed in advance.

Hypothesis: A hypothesis is a statement about the population parameter.
ie a hypothesis is a conclusion which is tentatively drawn on logical basis.

Statistical hypothesis: It is some assumption or statement, which may or may not be true, about a population or about the probability distribution characterising the given population which we want to test on the basis of the evidence from a random sample.

Test of hypothesis: The testing of hypothesis is a procedure that helps us to ascertain the likelihood of hypothesised population parameter being correct by making use of the sample statistics.

ie It is a process of test of significance which concerns with the testing of some hypothesis regarding a parameter of the population on the basis of statistics from the sample.

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 complemen-tary to the null hypothesis is called an alternative hypothesis and is denoted by H_1.

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

Type-I Error: It is the error of
Reject H_0 when it is true.

Type-II error: It is the error of
Accept H_0 when it is false.

Level of significance:
The level of significance is the maximum probability of making a type I error and it is denoted by 'α'.
ie P(Rejecting H_0 when H_0 is true) $= \alpha$.
The probability of making a correct decision is then $1-\alpha$.

Note: The commonly used level of significance in practice are 5% (0.05) and 1% (0.01). If we use 5% LoS, probability of making type I error is 5%.

Critical Region (or) Rejection Region:

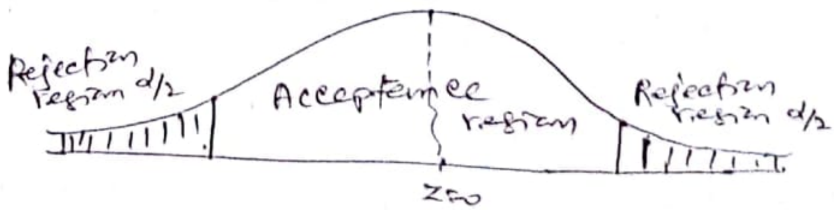
The rejection region or critical region is the region of the standard normal curve corresponding to a predetermined level of significance 'α'.

Acceptance region:-

The region under the normal curve which is not covered by the rejection region is known as acceptance region.

Two tailed test: when the test of hypothesis is made on the basis of rejection region represented by both sides of the standard normal curve, it is called a two tailed test.

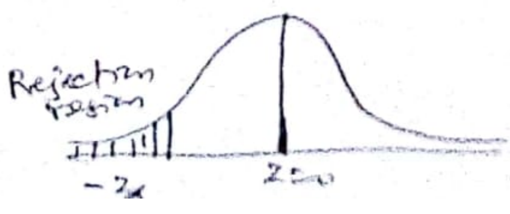
Ex. $H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$
$(\mu > \mu_0$ or $\mu < \mu_0)$



one tailed test: A test of statistical hypothesis where the alternative hypothesis is one sided is called as one tailed test.

Right tailed test: In this test, the rejection region lies entirely on the right tail of the normal curve

Left tailed test: In this test, the rejection region lies entirely on the left tail of the normal curve.



Critical values or Table values

Nature of Test / LOS (α)	1% (0.01)	5% (0.05)
Two tailed	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 1.96$
Right tailed	$z_\alpha = 2.33$	$z_\alpha = 1.645$
Left tailed	$z_\alpha = -2.33$	$z_\alpha = -1.645$

Procedure for Testing of Hypothesis

1. Set up null hypothesis H_0.
2. Set up alternative hypothesis H_1 after careful study of the problem and decide the nature of the test (whether one tailed or two tailed test)
3. Fix 'LOS'. (if not given in the problem) or take from the problem if specified and note z_α (Table value)
4. Compute the test statistics.

$$z = \frac{t - E(t)}{S.E(t)}$$

5. Compare the values $|z|$ and z_α

If $|z| < z_\alpha$, accept H_0, ie it is concluded that the difference bet. t and $E(t)$ is not significant

Otherwise Reject H_0

Test-I: Test of significance of the difference between sample mean and population mean.

Test statistic $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

where $\bar{x}$ – sample mean σ – Population S.D

μ – Population mean n – Sample size.

Prob-1: A sample of 400 students is found to have a mean height of 171.38 cms. Can it be reasonably regarded as a sample from a large population with mean height 171.17 cms and s.d 3.30 cms.

Soln:- Given that $n = 400$, $\bar{x} = 171.38$ cms, $\mu = 171.17$ cms

$\sigma = 3.30$ cms

1. $H_0: \bar{x} = \mu$;
2. $H_1: \bar{x} \neq \mu$ (Two tailed test) is used
3. Let the Los be 1%, $z_\alpha = 2.58$
4. $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{171.38 - 171.17}{3.30/\sqrt{400}} = 1.27$
5. Calculated value of $z = 1.27 <$ Tab. Value $z_\alpha = 2.58$

Accept H_0

Prob-2: A random sample of 200 tins of coconut oil gave an average weight of 4.95 kgs with a s.d of 0.21 kg. Do we accept the hypothesis of net weight 5 kgs per tin at 5% Los.

(Ans. $|z| = 3.37$; $z_\alpha = 1.96$)

Test-2: Test of significance of the difference between the means of two samples.

Let $\bar{x}_1$ and $\bar{x}_2$ be the means of two large samples of sizes n_1 and n_2 drawn from two populations (normal or non normal) with the same mean μ and variances σ_1^2 and σ_2^2 respectively.

$$\text{Test Statistics} \quad Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Case (i): If the samples are drawn from the same population ie if $\sigma_1 = \sigma_2 = \sigma$ then

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Case (ii): If σ_1 and σ_2 are not known and $\sigma_1 \neq \sigma_2$, σ_1 and σ_2 can be approximated by the sample s.d's s_1 and s_2

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Case (iii): If σ_1 and σ_2 are equal and unknown, then $\sigma_1 = \sigma_2 = \sigma$ is approximated by

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\left(\frac{n_1 + n_2}{n_1 n_2}\right)}}$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}}$$

Problem:- Test the significance of the difference between the means of the samples, drawn from two normal populations with the same S.D from the following data.

	size	mean	S.D
Sample-1	100	61	4
Sample-2	200	63	6

Soln:- Given that $\bar{x}_1 = 61$, $\bar{x}_2 = 63$, $n_1 = 200$, $n_2 = 200$
$s_1 = 4$, $s_2 = 6$.

1. $H_0: \bar{x}_1 = \bar{x}_2$
2. $H_1: \bar{x}_1 \neq \bar{x}_2$
3. Let the LOS be 5%, $Z_\alpha = 1.96$
4. $Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}} = -3.02$

5. Cal. value of $|z| = 3.02$ > Tab. value of $z_\alpha = 1.96$

Reject H_0.

Prob. - 2 : The average marks scored by 32 boys is 72 and s.d of 8, while that for 36 girls is 70 with a s.d of 6. Test at 1% level of significance whether the boys perform better than girls.

Soln. - Given that

$n_1 = 32, \quad \bar{x}_1 = 72, \quad s_1 = 8$

$n_2 = 36, \quad \bar{x}_2 = 70, \quad s_2 = 6.$

1. $H_0 : \bar{x}_1 = \bar{x}_2$

2. $H_1 : \bar{x}_1 > \bar{x}_2$

3. Given that LOS is 1%, $\quad z_\alpha = 2.33.$

4. Compute $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{72 - 70}{\sqrt{\frac{8^2}{32} + \frac{6^2}{36}}} = 1.15$$

5. Cal. value $|z| = 1.15 < z_\alpha = 2.33$

Accept H_0.

Prob. 3 : A sample of heights of 6400 English men has a mean of 170 cm and a s.d of 6.4 c.m, while a sample of heights of 1600 Americans has a mean of 172 cm and a s.d of 6.3 cm. Do the data indicate that Americans are, on the average, taller than the Englishmen?

Soln.:- $n_1 = 6400$, $\bar{x}_1 = 170$, $s_1 = 6.4$

$n_2 = 1600$, $\bar{x}_2 = 172$, $s_2 = 6.3$

1. $H_0 : \bar{x}_1 = \bar{x}_2$

2. $H_1 : \bar{x}_1 < \bar{x}_2$

3. Let the LOS be 1%, $Z_\alpha = -2.33$

4. $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = -11.32$$

5. Cal. value of $|Z| = 11.32 >$ Tab. value of $|Z_\alpha| = 2.33$

Reject H_0.

Test-3 - Test of significance of the difference between sample proportions and population proportion.

Test statistics $$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

where p - sample proportion ; $Q = 1 - P$

P - population proportion ; n - sample size.

Prob.-1 : In a sample of 600 parts manufactured by a factory, the number of defective parts was found to 45. The company however claimed that only 5% of their product is defective. Is the claim tenable.?

Soln:- 1. $H_0: P = 0.05$

2. $H_1: P > 0.05$ (one tailed test) is used

3. LOS is 5%, $z_\alpha = 2.33$

4. $$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}},$$ $n = 600$, $p = \frac{45}{600}$

$P = \frac{5}{100}$, $Q = \frac{95}{100}$

$$= \frac{\frac{45}{600} - \frac{5}{100}}{\sqrt{\frac{5/100 \times \frac{95}{100}}{600}}} = 2.8$$

5. Cal. value of $z = 2.8 >$ Tab. value of $z = 2.33$

Reject H_0

Prob-2 In a certain city 380 men out of 800 are found to be smokers. Discuss whether this information supports the view that majority of men in this city are non-smokers?

Soln:- Given $n = 800$,

Sample proportion of non smokers $= \frac{420}{800}$.

population proportion $P = \frac{1}{2}$, $Q = 1 - P = \frac{1}{2}$.

1. $H_0: P = \frac{1}{2}$

2. $H_1: P > \frac{1}{2}$ (majority of men are non smokers, one tailed test)

3. Let the LOS be 5%, $z_\alpha = 1.645$