# UNIT-I

# TESTING THE HYPOTHESIS

**Population:**

A population consists of collection of individual units, which may be person's or experimentaloutcomes, whose characteristics are to be studied.

**Sample:**

A sample is proportion of the population that is studied to learn about the characteristics of the population.

**Random sample:**

A random sample is one in which each item of a population has an equal chance of being selected.

**Sampling:**

The process of drawing a sample from a population is called sampling.

**Sample size:**

The number of items selected in a sample is called the sample size and it is denoted by 'n'.

If n $\geq 30$, the sample is called large sample and if n $\leq 30$, it is called small sample.

**Sampling distribution:**

Consider all possible samples of size' n' drawn from a given population at random. We calculate mean values of these samples.

If we group these different means according to their frequencies, the frequency distribution so formed is called sampling distribution.

The statistic is itself a random variate. Its probability distribution is often called sampling distribution.

All possible samples of given size are taken from the population and for each sample, the statistic is calculated. The values of the statistic form its sampling distribution.

**Standard error:**

The standard deviation of the sampling distribution is called the standard error.

**Notation:**

Population mean = $\mu$ ; Population Standard Deviation =$\sigma$ ;

P - Population Proportion

Sample Mean = $\bar{x}$; Sample Standard Deviation = s;

p = Sample Proportion

**Null Hypothesis ($H_0$)**

The hypothesis tested for possible rejection under the assumption that it is true is usually called null hypothesis. The null hypothesis is a hypothesis which reflects no change or no difference. It is usually denoted by $H_0$.

**Alternative Hypothesis ($H_1$ )**

The Alternative hypothesis is the statement which reflects the situation anticipated to be correct if the null hypothesis is wrong. It is usually denoted by $H_1$ .

**For example:**

If $H_0$: $\mu_1 = \mu_2$ (There is no difference between the means) then the formulated alternative hypothesis is$H_1$: $\mu_1 \neq \mu_2$

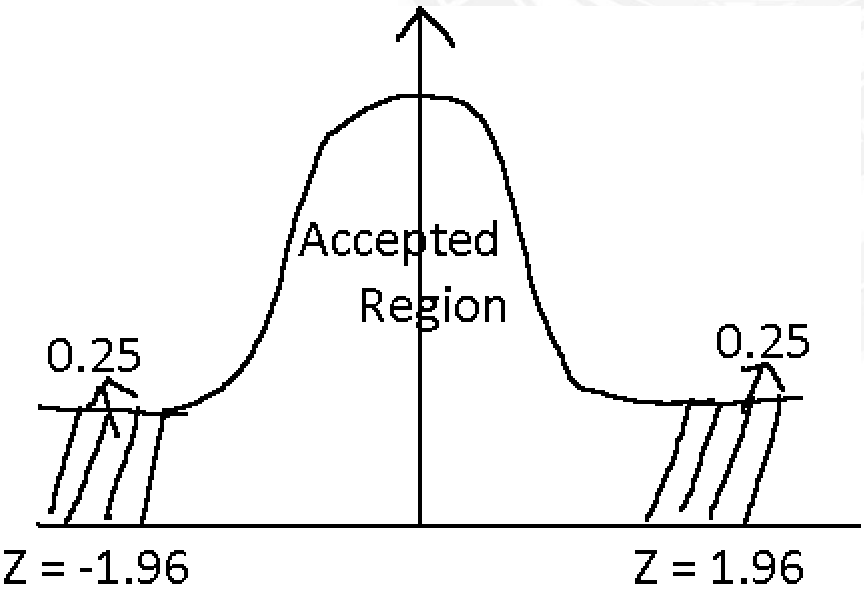
i.e., either $H_1$: $\mu_1 < \mu_2$ (or) $\mu_1 > \mu_2$

**Level of significance:**

It is the probability level below which the null hypothesis is rejected. Generally, 5% and 1% level of significance are used.

**Critical Region (or) Region of Rejection:**

The critical region of a test of statistical hypothesis is that region of the normal curve which corresponds to the rejection of null hypothesis




- **Critical values (or) Significant values:**
- The sample values of the statistic beyond which the null hypothesis will be rejected are called critical values or significant values
- **Level of significance**

| Types of test | 1% | 5% | 10% |
|---|---|---|---|
| Two tailed test | 2.58 | 1.96 | 1.645 |
| One tailed test | 2.33 | 1.645 | 1.28 |

- **Two tailed test and one-tailed tests:**
- When two tails of the sampling distribution of the normal curve are used, the relevant test is called two tailed test.
- The alternative hypothesis $H_1\colon\ \mu_1 \neq \mu_2$ is taken in two tailed test $H_0\colon\ \mu_1 = \mu_2$.
- When only one tail of the sampling distribution of the normal curve is used, the test is described as one tail test $H_1\colon \mu_1 < \mu_2$ (or) $\mu_1 > \mu_2$

**Type I and Type II Error:**

- Type I Error: Rejection of null hypothesis when it is correct
- Type II Error: Acceptance of null hypothesis when it is wrong

PROCEDURE FOR TESTING OF HYPOTHESIS

- State the null hypothesis $H_0$
- Decide the alternative hypothesis $H_1$ (i.e., one tailed or two tailed)
- Choose the level of significance $\alpha$ at 5% (or) 1%.
- Compute the test statistic $Z = \frac{t - E(t)}{S.E\ of\ (t)}$
- Compare the computed value of with the table value of $|Z|$ with the table value of Z and decide the acceptance or the rejection of $H_0$.
- If $|Z| < 1.96 H_0$ is accepted at 5% level of significance.
- If $|Z| > 1.96 H_0$ is rejected at 5% level of significance.
- If $|Z| < 2.58 H_0$ is accepted at 1% level of significance.
- If $|Z| > 2.58 H_0$ is rejected at 1% level of significance.
- For a single tail test (right tail or left tail) we compare the computed value of $|Z|$ with 1.645 (at 5% level of significance) and 2.33 (at 1% level of significance) and accept or reject $H_0$ accordingly.
- **Test of Hypothesis (Large Sample Tests)**

**1.1 Large sample tests (Test based in Normal distribution.)**

- **Type - I: (Test of significance of single mean)**
- Let $\{x_1,\ x_2,\ .\ .\ .\ ,\ x_n\}$ be a sample of size $(n \geq 30)$ taken from a population with mean $\mu$ and Standard Deviation $\sigma$. Let $\bar{x}$ be the sample mean. Assume that the population is Normal.

- To test whether the difference between Population mean $\mu$ and sample mean $\bar{x}$ is significant or not and this sample comes from the normal population whose mean is $\mu$ or not.
- $H_0$: $\mu$ = a specified value
- $H_1$: $\mu \neq$ a specified value
- we choose $\alpha$= 0.05(5%) (or) 0.01(1%) as the Level of significance

The test statistic is $Z = \frac{\bar{x}-\mu}{S.E(\bar{x})} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

**Note:**

If $\sigma$ is not known, for large n, S. E $\bar{x} = \frac{s}{\sqrt{n}}$ where s is the sample S.D

**PROBLEMS ON SINGLE MEAN**

1.. **A sample of 900 members is found to have a mean 3.5cm. Can it reasonably regarded as a simple sample from a large population whose mean is 3.38 and a standard deviation 2.4cm?**

**Solution:**

Set the null hypothesis $H_0$: μ= 3.38

Set the alternative hypothesis: $H_1$: μ ≠3.38

Level of significance $\alpha$= 0.05(5%)

The test statistic is $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Given $\bar{x} = 3.5,\ \mu = 3.38,\ n = 900,\ \sigma = 2.4$

$$\Rightarrow Z = \frac{3.5 - 3.38}{\frac{2.4}{\sqrt{900}}} = 1.5$$

**Critical value:** At 5% level, the tabulated value of $Z_\alpha$ is 1.96

**Conclusion:** Since |Z| = 1.5 < 1.96

Hence Null Hypothesis $H_0$ is accepted at 5% level of significance.

i.e., the sample comes from a population with mean 3.38cm

2. **A manufacturer claims that his synthetic fishing line has a mean breaking strength of 8kg and S.D 0.5kg. Can we accept his claim if a random sample of 50 lines yield a mean breaking of 7.8kg. Use 1% level of significance.**

**Solution:**

Set the null hypothesis $H_0$: μ= 8

Set the alternative hypothesis: $H_1$: μ ≠8

Level of significance $\alpha$= 0.01(1%)

The test statistic is $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Given $\bar{x} = 7.8,\ \mu = 8,\ n = 50,\ \sigma = 0.5$

$$\Rightarrow Z = \frac{7.8 - 8}{\frac{0.5}{\sqrt{50}}} = -2.828$$

**Critical value:**

At 1% level, the tabulated value of $Z_\alpha$ is 2.58

**Conclusion:** Since |Z| = 2.828 > 2.58

Hence Null Hypothesis $H_0$ is rejected at 1% level of significance.

i.e., the manufacturer's claim is not accepted.

**Type – II – Test of significant of difference between two means**

- Consider two samples of sizes $n_1$ and $n_2$ taken from two different populations with population means $\mu_1$ and $\mu_2$ and S. D $\sigma_1$and $\sigma_2$.
- Let $\overline{x_1}$ and $\overline{x_2}$ be the sample means and $s_1$ and $s_2$ be the sample S. D of the samples.
- The formulated Null and Alternative hypothesis is, $H_0\!:\ \mu_1 = \mu_2$

$$H_1\!:\ \mu_1 \neq \mu_2$$

- The test Statistic Z is defined by $Z = \frac{\overline{x_1}-\overline{x_2}}{S.E(\overline{x_1}-\overline{x_2})} = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}}$
- Where $\sigma_1$ and $\sigma_2$ S.D of populations.
- The test Statistic Z is defined by $Z = \frac{\overline{x_1}-\overline{x_2}}{S.E(\overline{x_1}-\overline{x_2})} = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{s_1{}^2}{n_1}+\frac{s_2{}^2}{n_2}}}$
- Where $s_1$ and $s_2$ S.D of populations.
- we choose $\alpha$= 0.05(5%) (or) 0.01(1%) as the Level of significance
- If $|Z_\alpha| < Z$ accept Null hypothesis.
- If $|Z_\alpha| > Z$ accept Alternative hypothesis.

1. **The mean of two large samples of 1000 and 200 members are 67.5 inches and 68 inches respectively. Can the samples be regard as drawn from the population of standard deviation of 2.5 inches? Test at 5% level of significance?**

**Solution:**

Set the null hypothesis $H_0\!:\ \mu_1 = \mu_2$

Set the alternative hypothesis $H_1\!:\ \mu_1 \neq \mu_2$

Level of significance $\alpha$= 0.05(5%)

The test Statistic Z is defined by $Z = \frac{\overline{x_1}-\overline{x_2}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

Given $\overline{x_1} = 67.5,\ \ \overline{x_2} = 68,\ n_1 = 1000,\ n_2 = 200,\ \sigma = 2.5$

$$\Rightarrow Z = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{200}}} = 5.164$$

**Critical value:**

At 5% level, the tabulated value of $Z_\alpha$ is 1.96

**Conclusion:** Since |Z| = 5.164 > 1.96

Hence Null Hypothesis $H_0$ is rejected at 5% level of significance.

i.e., The sample cannot be regards as drawn from the same population.

**2. Samples of students were drawn from two universities and from the weights is kilogram. The means and S.D's are calculated. Test the significance of the difference between the means of two samples**

| | Mean | S.D | Sample Size |
|---|---|---|---|
| University A | 55 | 10 | 400 |
| UniversityB | 57 | 15 | 100 |

**Solution:**

Set the null hypothesis $H_0$: $\mu_1 = \mu_2$

Set the alternative hypothesis $H_1$: $\mu_1 \neq \mu_2$

Level of significance $\alpha$= 0.05(5%)

The test Statistic Z is defined by $Z = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}}$

Given $\overline{x_1} = 55,\ \ \overline{x_2} = 57,\ n_1 = 400,\ n_2 = 100, s_1 = 10,\ s_2 = 15$

$$\Rightarrow Z = \frac{55 - 57}{\sqrt{\frac{10^2}{400} + \frac{15^2}{100}}} = 1.265$$

**Critical value:**

At 5% level, the tabulated value of $Z_\alpha$ is 1.96

**Conclusion:** Since |Z| = 1.265 < 1.96

Hence Null Hypothesis $H_0$ is accepted at 5% level of significance.

Hence the difference between the means isnot significant

3. **A sample of heights of 6400 Englishmen has a mean of 170cm and a S. D of 6.4cm, while a another sample of heights of 1600 Americans has a mean of 172cm and a S. D of 6.3cm. Do the data indicate that Americans are on the average taller than the Englishmen? Test the LOS at 1%?**

**Solution:**

Set the null hypothesis $H_0$: $\mu_1 = \mu_2$

Set the alternative hypothesis $H_1$: $\mu_1 < \mu_2$

Level of significance $\alpha$= 0.01(1%)

The test Statistic Z is defined by $Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}}$

Given $\overline{x_1} = 170$, $\overline{x_2} = 172$, $n_1 = 6400$, $n_2 = 1600, s_1 = 6.4$, $s_2 = 6.3$

$$\Rightarrow Z = \frac{170 - 172}{\sqrt{\frac{6.4^2}{6400} + \frac{6.3^2}{1600}}} = -11.325$$

**Critical value:** At 1% level, the tabulated value of $Z_\alpha$ is 2.33.

**Conclusion:** Since |Z| = 11.325 > 2.33

Hence Null Hypothesis $H_0$ is rejected at 1% level of significance.

i.e., The Americans are on the average, taller than the Englishmen