Testing Hypothesis

- A statistical hypothesis is an assertion or conjecture concerning one or more populations.

- Important factors of Hypothesis Testing
- Population: The complete collection of all the elements to be studied.

  Sample: A sub collection of elements drawn from a population.
- Hypothesis
  1) Null (H_0)
  2) Alternative $(H_a$ or $H_1)$
- Test (Statistics)

1) χ^2 test (Chi-square test)
2) t-student test or t-test
3) Fisher's z-test

Procedure for testing the Hypothesis:

1) To set the hypothesis.
2) To set the suitable significant level.
3) To set the test criteria.
4) Computation.
5) Decision.

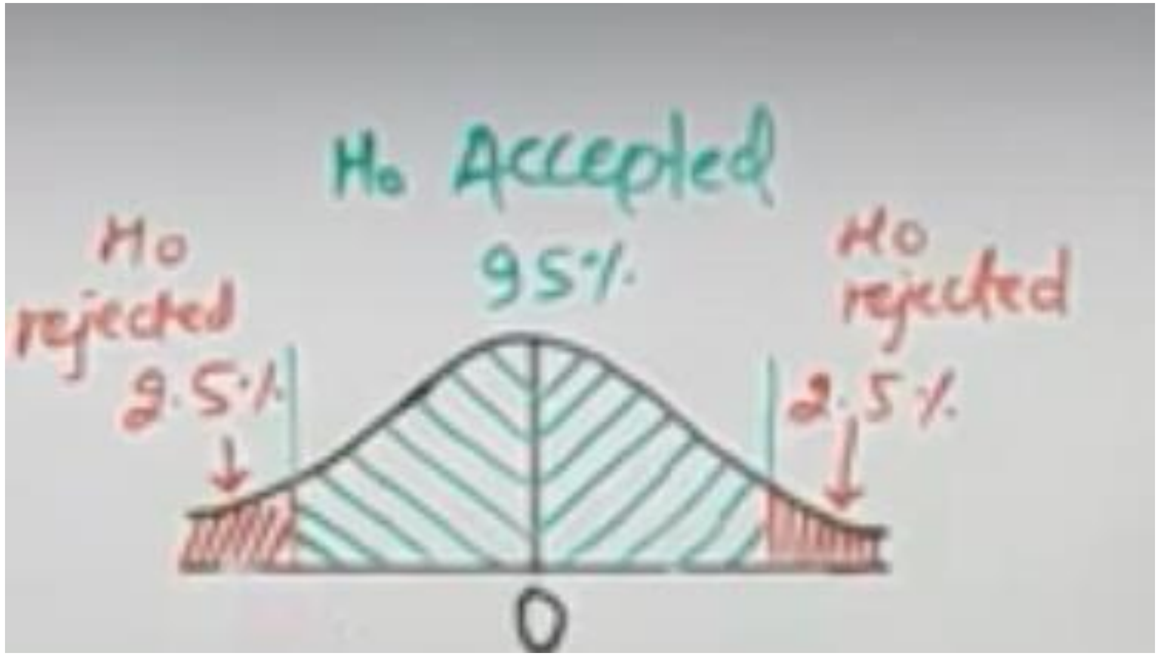
1) To set the hypothesis.

a) Null (H_0)

b) Alternative $(H_a$ or $H_1)$

2) To set the suitable significant level.

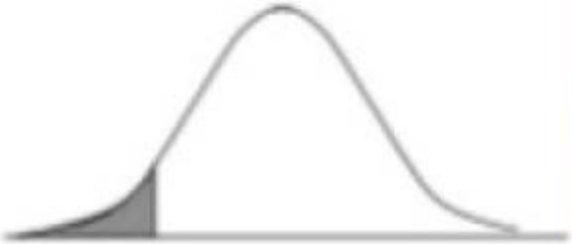
Test the validity of H_0 against H_1 at certain level of significance, i.e. 5% or 1%, etc.



One-Tail Test (left tail)

$$H_0 : \mu = \mu_0$$

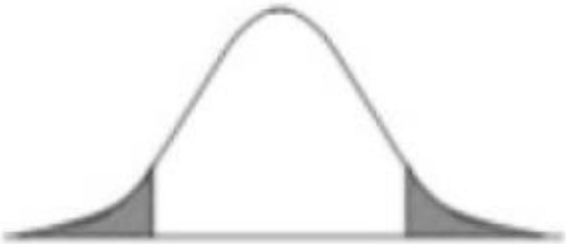
$$H_1 : \mu < \mu_0$$



Two-Tail Test

$$H_0 : \mu = \mu_0$$

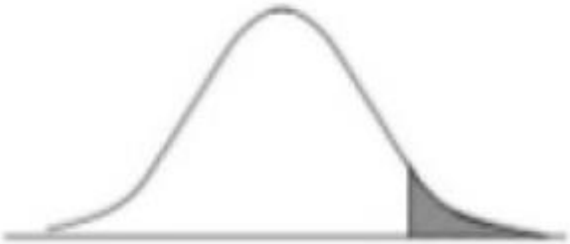
$$H_1 : \mu \neq \mu_0$$



One-Tail Test (right tail)

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

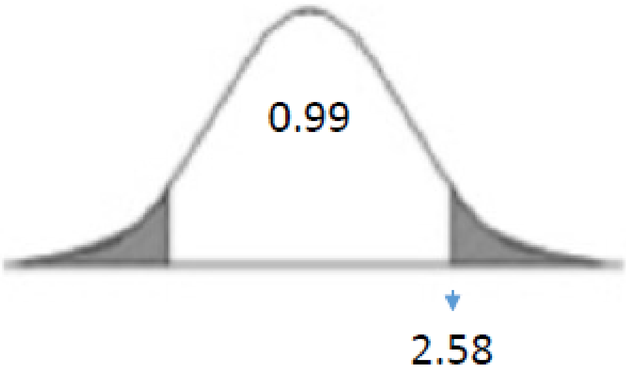


- For 1% level of significance

acceptance region is $\alpha = 1\%$

➢ <u>for two tailed</u>

$$P(z_1 < z < z_2) = 1 - \alpha = 1 - 0.01 = 0.99$$



since acceptance region is symmetric about mean

$$P(z_1 < z < z_2) = \frac{0.99}{2} = 0.495$$

The area under the normal curve with 0.495 is $z = 2.58$

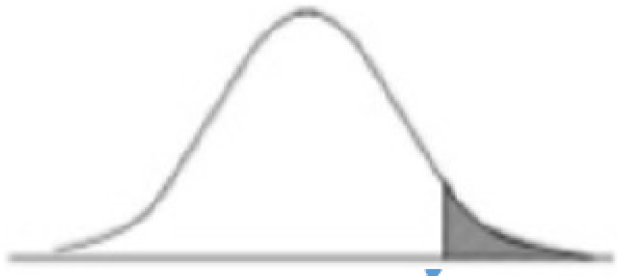
rejection region is =0.5-0.495=0.005

- For one tailed

right tailed:

$$P(Z > Z_\alpha) = \alpha = 0.01$$

$$P(O < Z < Z_1) = 0.5 - 0.01 = 0.49$$

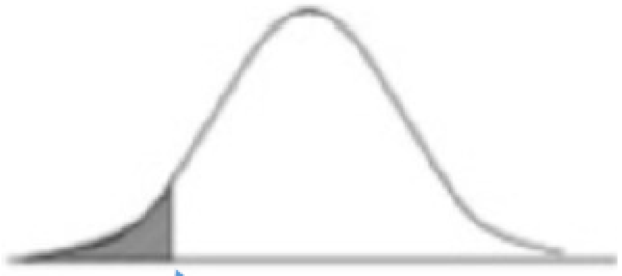


the area under the normal curve with 0.49 is $z_1 = 2.33$

left tailed:

$$P(Z < Z_\alpha) = \alpha = 0.01$$

$$P(O < Z < Z_2) = 0.5 - 0.01 = 0.49$$



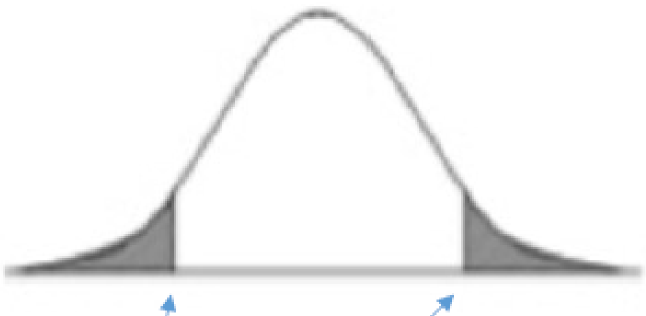
the area under the normal curve with 0.49 is $z_2 = 2.33$

- For 5% level of significance

  acceptance region is $\alpha = 5\%$

➢ for two tailed

$$P(z_1 < z < z_2) = 1 - \alpha = 1 - 0.05 = 0.95$$



since acceptance region is symmetric about mean

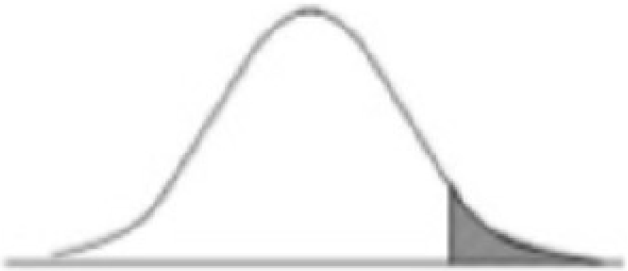
$$P(z_1 < z < z_2) = \frac{0.95}{2} = 0.475$$

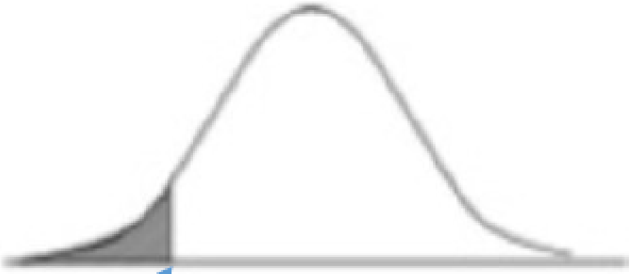
The area under the normal curve with 0.475 is $z = 1.96$

rejection region is =0.5-0.475=0.025

- Similarly,

for right tailed and left tailed





the value is $z = 1.645$

3) Test Statistic

compute the test statistic $z = \frac{t - E(t)}{S.E\ of\ t}$

$$z = \frac{Observed\ value - Expected\ value}{S.E.of\ t}$$

under the null hypothesis.

here t is sample statistic

5) Decision

compare the test statistic z with the critical value z_α at given level of significance (α).

if $|z| < z_\alpha$, we conclude that it is not significant,

we <u>accept</u> the null hypothesis.

if $|z| > z_\alpha$, then the difference is significant and hence

we <u>reject</u> the null hypothesis.

- Errors of sampling

(1) Type I error or α error

If the Null hypothesis H_0 is true but it is rejected by test procedure, then the error made is called Type I error.

(2) Type II error or β error

If the null hypothesis H_0 is false but it is accepted by test, the error committed is called Type II error.

	Accept H_0	Reject H_0
H_0 is true	Correct decision	Type I error
H_0 is false	Type II error	Correct decision

Test of hypothesis for large samples:

- Under **large sample test**, the following are the important tests to test the significance Z-TEST

  (1). Testing the significance of single mean

  (2). Testing the significance of difference of means

  (3). Testing the significance of single proportion

  (4). Testing the significance of difference of proportions

Testing the significance of single mean

- Aim: to test whether the difference between sample mean and population mean is significant or not.
- Procedure:

➢ Null hypothesis: let $\mu = \mu_0$

➢ Alternative hypothesis: may be $\mu \neq$ or $\mu > \mu_0$ or $\mu < \mu_0$

(depending on the given data)

➢ Level of significance:

choose either 1% or 5%

- Test statistic:

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \text{ (where } \sigma \text{ is known)}$$

- Conclusion:

compare the test statistic z with the critical value z_α at given level of significance (α).

if $|z| < z_\alpha$, we conclude that it is not significant, we accept the null hypothesis.

if $|z| > z_\alpha$, then the difference is significant and hence we reject the null hypothesis.

Example 1:

- A company manufacturing electric bulbs claims that the average life of its bulbs is 1600 hours. The average life and standard deviation of a random sample of 100 such bulbs were 1570 hours and 120 hours respectively. Should we accept the claim of the company?

Solution:

Given mean of population $\mu = 1600$, σ is unknown

from a sample of n$=$ 100,

mean $\bar{x} = 1570, s = 120$

- let the null hypothesis be : $\mu = 1600$
- And the alternative hypothesis be: $\mu \neq 1600$
- Level of significance be: 5% then $\mathrm{p} = 1.96$
- test of statistic: here σ is unknown

$$z = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{1570 - 1600}{\left(\frac{120}{\sqrt{100}}\right)} = -2.5$$

- Conclusion:

  Since $|z| = 2.5 > 1.96$

  therefore, the null hypothesis is rejected at 5% LOS.

  Hence, we conclude that the claim of the company should not be accepted at 5% level of significance(LOS).