PROBABILITY AND STATISTICS

Lecture-1

Unit-III

Tests of Hypothesis for Large Sample

Dr. B. VIJAYA BASKER REDDY
Faculty of Mathematics
Department of Science & Humanities
SNIST

Syllabus

- Tests of Hypothesis
- Type-I and Type-II Errors
- Hypothesis testing concerning one mean and two means
- Test of Hypothesis concerning to one proportion and difference of proportions.

Topics of the Lecture:

- Hypothesis
- Types of Hypothesis
- Types of Errors in Hypothesis
- Level Of Significance(LOS)
- One Tailed Test and Two Tailed Test
- Procedure for Testing of Hypothesis
- Important tests of Hypothesis

Population:

Population is defined as a large collection of individuals or attributes or numerical data.

Note: Population may be finite or infinite.

Sample: A sample is defined as a finite subset of population.

Note1:If the size of a sample <30 then the sample is said to be a small sample.

Note2:If the size of sample $\geq$30,the sample is said to be large.

What is a Hypothesis?

A hypothesis is an assumption about the population parameter.

Where parameter is a Population mean or proportion etc.

Types of Hypothesis:

In general two types of Hypothesis are constructed namely,

1. Null Hypothesis
2. Alternative Hypothesis

Null Hypothesis (N.H):

N.H is the hypothesis which is tested for possible rejection under the assumption that it is true. (R.A.Fisher)

Null Hypothesis is denoted by H_0

Note: Null Hypothesis involving statements with equality (includes any value not stated by A.H)

1.H_0: $\mu = \mu_0$ *i.e.* population mean equals to specified constant μ_0

2.H_0: $\mu_1 - \mu_2 = \delta$

Alternative Hypothesis

The alternative hypothesis, denoted by H_1, is any Hypothesis other than null the hypothesis.

Note: The possible alternative hypothesis can be stated for mean in any one of the following forms.

1. $H_1: \mu > \mu_0$

2. $H_1: \mu < \mu_0$

3. $H_1: \mu \neq \mu_0$

- Example: A new teaching method is developed that is believed to be better than the current method.

Null Hypothesis:

The new method is not better than the old method.

Alternative Hypothesis:

The new teaching method is better.

Test of Hypothesis:

A procedure to decide whether to accept or reject the null hypothesis is called Test of Hypothesis or Test of Significance.

Note: In hypothesis testing, we usually proceed on the basis of N.H, at the end of the test, one of two decisions will be made:

1. Reject the null hypothesis, or
2. Fail to reject the null hypothesis.

Types of Errors in Hypothesis

- A **type-I error** occurs if the null hypothesis is rejected when it is true.

- A **type-II error** occurs if the null hypothesis is not rejected when it is false.

Types of Errors

	Actual Truth of H_0	
Decision	**Accept H_0**	**Reject H_0**
H_0 is True	Correct Decision	**Type I Error**
H_0 is False	**Type II Error**	Correct Decision

Level Of Significance(LOS)

The probability level below which we reject the null hypothesis is know as level of Significance and it is denoted by α.

Level of Significance with Example

α=5% means there are 5 chances in100 that Null Hypothesis is rejected when it is true or one is 95% confidence that the right decision is made.

Level of Significance with Example

α=1% means there are 1 case in100 that the Null Hypothesis is rejected when it is true i.e. we are about 99% confidence that we have made the right decision.

Critical Region

It is the region of rejection of null hypothesis.

The area of critical region is level of significance.

Critical values

The value of the test static S^{*}_{α} ,which separates the critical region(rejected region)and the acceptance region is called critical value or significant value.

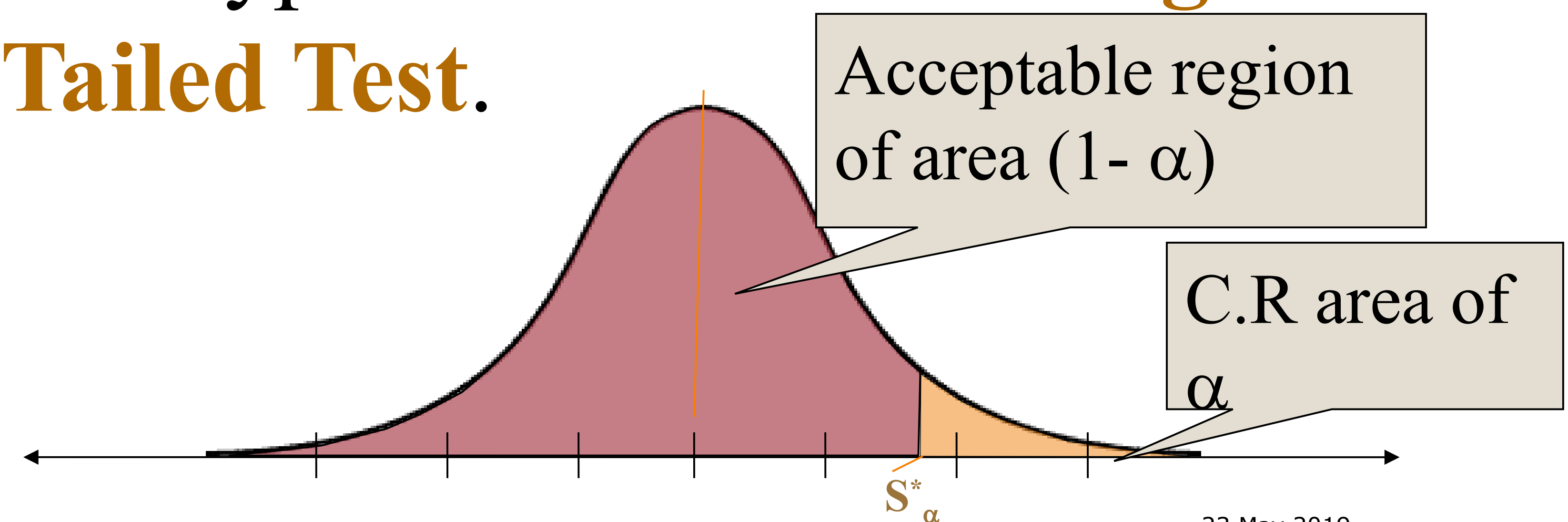
One Tailed Test

and

Two tailed Test

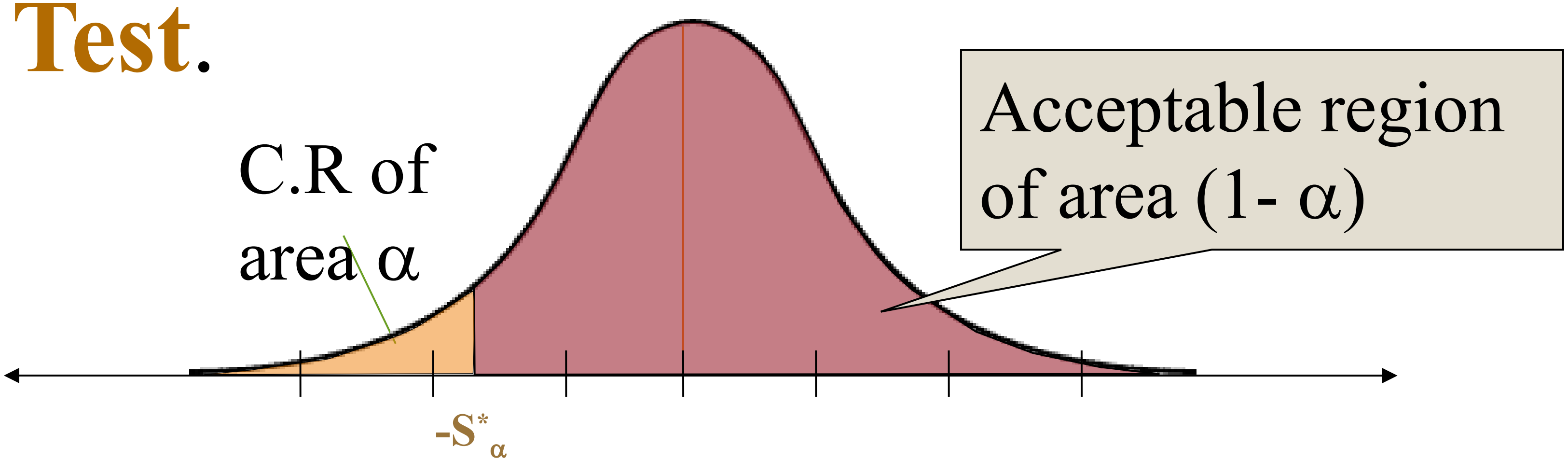
Right One Tailed Test (R.O.T.T)

If the Alternative Hypothesis contains the greater-than symbol (>) i.e H_1: $\mu > \mu_0$ or H_1: $\sigma_1^2 > \sigma_2^2$ *etc.*,, the hypothesis test is a **Right One Tailed Test**.



Left One Tailed Test (L.O.T.T)

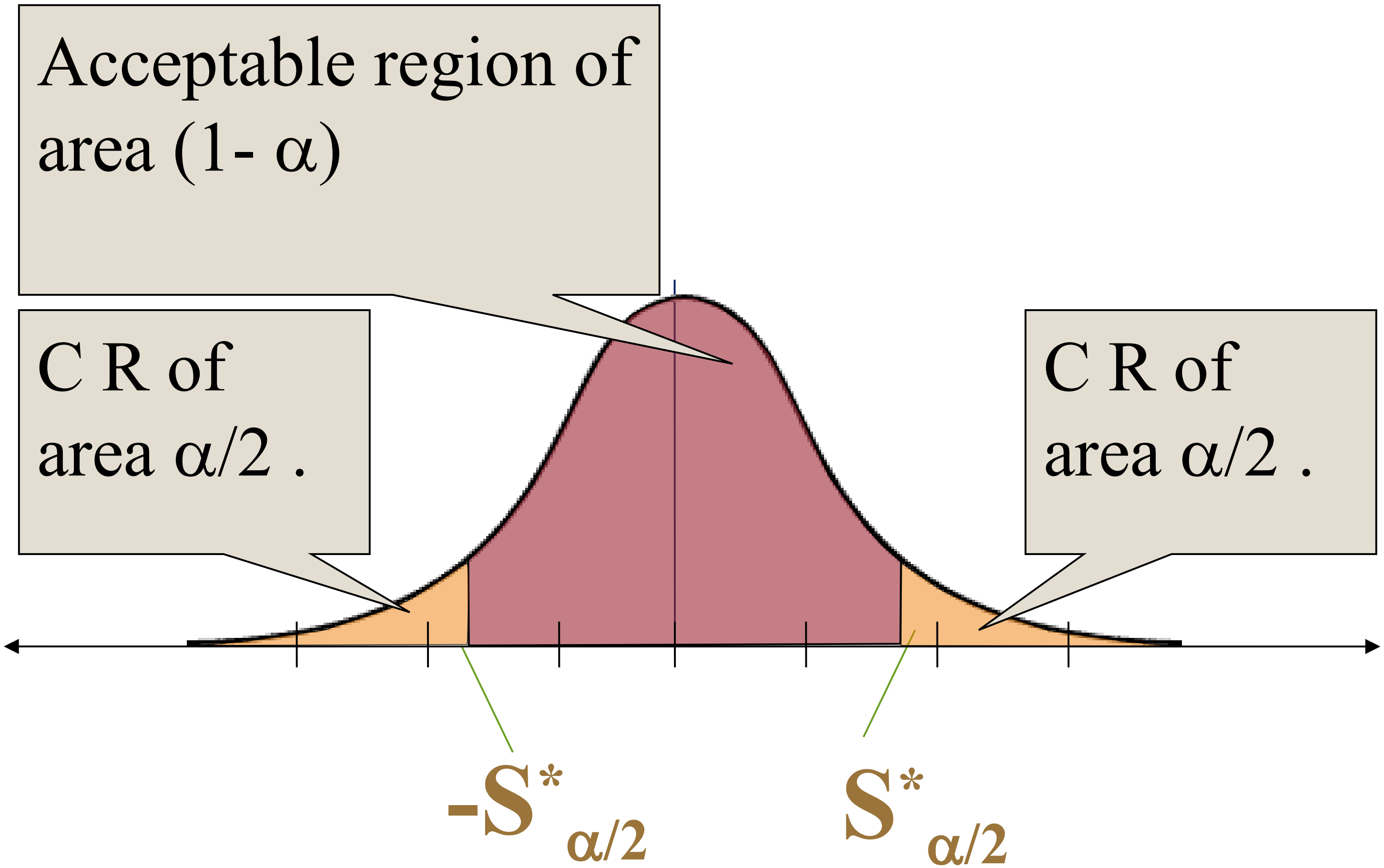
If the alternative hypothesis contains the greater-than symbol (<) i.e H_1: $\mu < \mu_0$ *or* H_1: $\sigma_1^2 < \sigma_2^2$ *etc.*, the hypothesis test is a **Left One Tailed Test**.



Two-Tailed Test

If the alternative hypothesis contains the not-equal-to symbol ($\neq$) i.e., H_1: $\mu \neq \mu_0$ etc., the hypothesis test is a **two-tailed test**.

Where the C.R lies on both sides of the right and left tails such that the area $\alpha/2$ lies on left tail and the area $\alpha/2$ lies on right tail .



Acceptable region of area (1- α)
C R of area α/2 .
C R of area α/2 .
-S*α/2
S*α/2

Procedure for testing of Hypothesis :

Testing of Hypothesis has following steps

(i)Formulate Null Hypothesis: H_0

(ii)Formulate Alternate Hypothesis: H_1

(iii)Choose Level of significance: α

(iv) Critical Region is decided according to alternate Hypothesis:

a) If A.H is Greater than type then test will be R.O.T.T.

b) If A.H is Less than type then test will be L.O.T.T.

c) If A.H is not equal type then test will be T.T.T.

(v) Compute the test static S^* using sample data i.e. s^*_{Cal}.

(vi)Decision :Accept or reject Null Hypothesis depending on the relation between S^* and S^*_α.

We will discuss the following topics in the next lecture

- **Test of Hypothesis concerning One Mean.**
- **Test of Hypothesis concerning Two Mean**

Thanks for watching this video

SNIST
Sreenidhi Institute of Science and Technology
Because, Life is all about taking the right decisions
AUTONOMOUS

PROBABILITY AND STATISTICS

Lecture-2

Unit-III

Tests of Hypothesis for Large Sample

Dr. B. VIJAYA BASKER REDDY
Faculty of Mathematics
Department of Science & Humanities
SNIST

Topics of the previous Lecture:

- Hypothesis
- Types of Hypothesis
- Types of Errors in Hypothesis
- Level Of Significance(LOS)
- One Tailed Test and Two Tailed Test
- Procedure for Testing of Hypothesis
- Important tests of Hypothesis

Topics of the Lecture:

- Types of Hypothesis concerning one mean for large sample

- Types of Hypothesis concerning two means for large sample

Test of Hypothesis concerning single mean.

Assumptions:

1. σ known
2. Normal Population or large sample
3. simple random sample from the population.

Note: For large sample $n \geq 30$, even if σ is unknown, σ can be replaced by sample variance S.

Procedure:

Step 1: $H_0 : \mu = \mu_0$

Step 2: $H_1 : \mu \neq \mu_0$ or

$H_1 : \mu < \mu_0$ or

$H_1 : \mu > \mu_0$

Step 3: level of Significance: α

Step 4:Critica Region (C.R): Find the

Critical value Z_α of Z at LOS α

from the normal table.

If H_1 is $\mu \neq \mu_0$ then use Two-Tailed Test.

If H_1 is $\mu < \mu_0$ then use Left-one Tailed Test.

If H_1 is $\mu > \mu_0$ then use Right-one

Step5: Test statistic

$$Z_{\mathrm{Cal}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Step6: H_0 is rejected if Z_{Cal} value falls in Critical region.

Critical value based on z distribution

Level of significance (α%)	**10%**	**5%**	**1%**
Level of significance(α)	**0.1**	**0.05**	**0.01**
two-tailed test (α/2)	**1.645**	**1.96**	**2.58**
one-tailed test (α)	**1.28**	**1.645**	**2.33**

Example 1.

The length of life X of certain computer is approximately normally distributed with mean 800 hours and standard deviation 40 hours. If a random sample of 30 computers has an average life of 788 hours.

Test the null hypothesis that $\mu=800$ hours against alternative hypothesis that $\mu\neq800$ hours at (a)1% (b) 5% level of significance.

Solution:

Given data

Population mean: μ

Population Standard deviation : $\sigma = 40$

Sample size: $n=30$

Sample mean: $\bar{x}=788$

Specified constant: $\mu_0=800$ hours

Case (a):

Step 1. Null hypothesis: $\mu=800$

Step 2.Alternative hypothesis : $\mu\neq800$

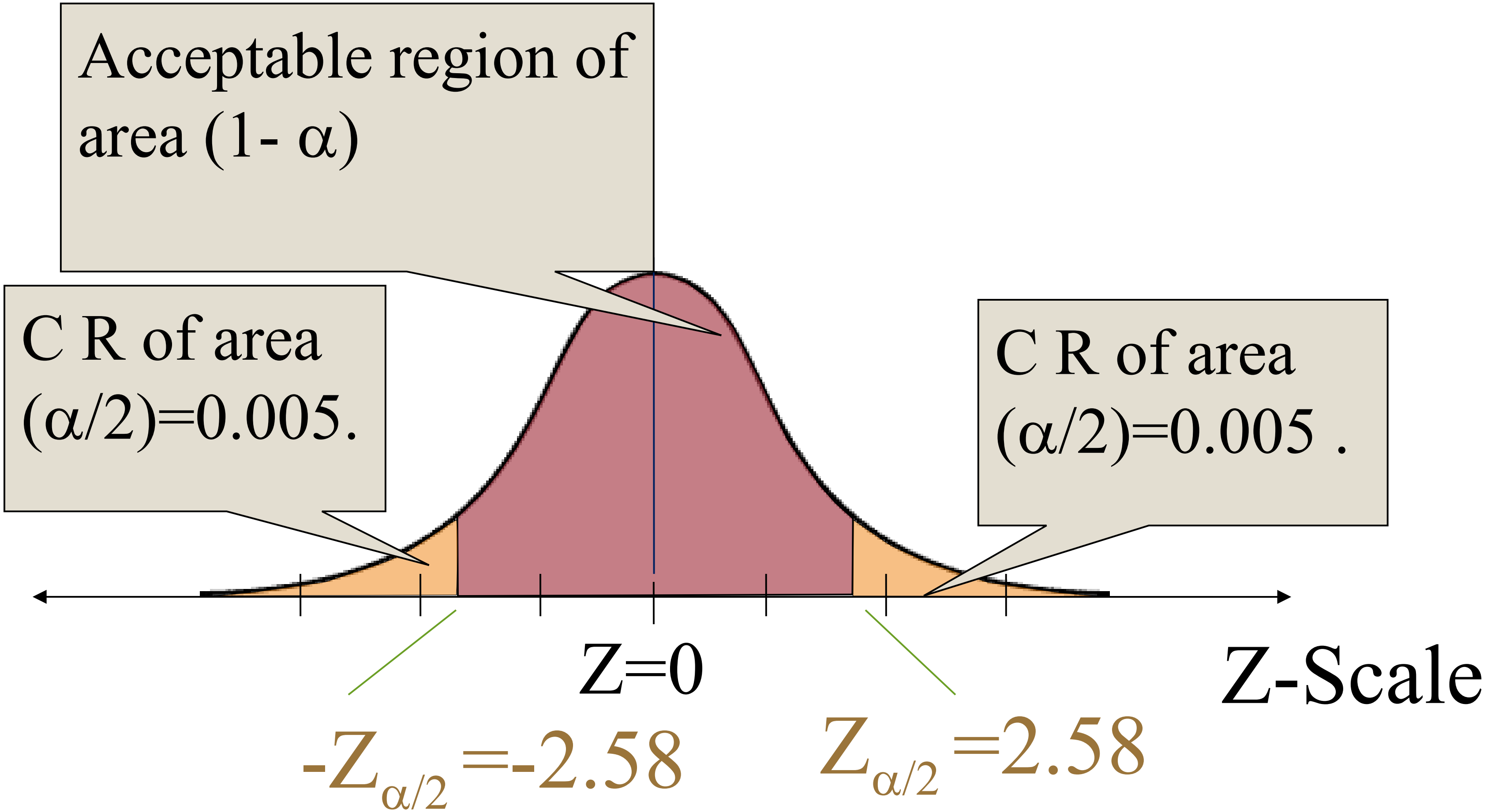
Step 3.Levl of Significance :

$\alpha=1\% =0.01$

Step 4. Critical region: Since Alternate Hypothesis is not equal type (≠),the test is two tailed .

$$Z_{\alpha/2}=Z_{0.01/2}=Z_{0.005}=2.58$$

$$-Z_{\alpha/2}=-Z_{0.01/2}=-Z_{0.005}=-2.58$$



Acceptable region of area (1- α)
C R of area (α/2)=0.005.
C R of area (α/2)=0.005 .
Z=0
Z-Scale
-Zα/2=-2.58
Zα/2=2.58

Step 5.Test Static

$$Z_{\text{Cal}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{788 - 800}{\frac{40}{\sqrt{30}}}$$

$$= -1.643$$

Step 6. Conclusion: Accept Null Hypothesis(H_0)

Since Z_{Cal} falls in acceptable region i.e

$-Z_{\alpha/2}=-2.58 < Z_{Cal}=-1.64 < Z_{\alpha/2}=2.58$

Case (b):

Step 1. Null hypothesis: $\mu=800$

Step 2. Alternative hypothesis : $\mu\neq800$

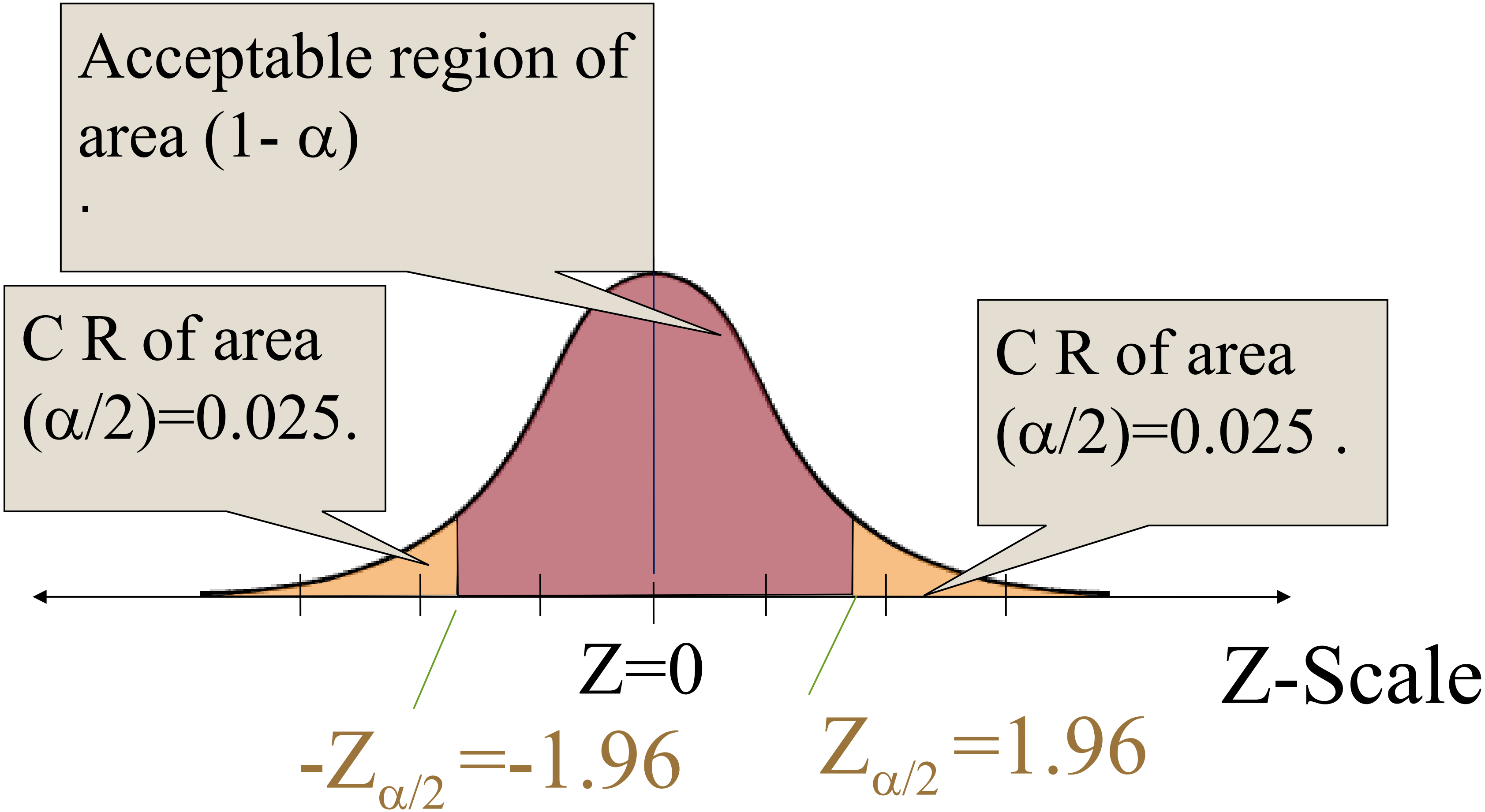
Step 3. Levl of Significance :

$\alpha=5\% =0.05$

Step 4. Critical region: Since Alternate Hypothesis is not equal type (≠),the tail is two tailed .

$$Z_{\alpha/2}=Z_{0.05/2}=Z_{0.025}=1.96$$

$$-Z_{\alpha/2}=-Z_{0.05/2}=-Z_{0.025}=-1.96$$


Acceptable region of area (1- α) .
C R of area (α/2)=0.025.
C R of area (α/2)=0.025 .
Z=0
Z-Scale
$-Z_{\alpha/2}=-1.96$
$Z_{\alpha/2}=1.96$

Step 5.Test Static

$$Z_{\mathrm{Cal}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{788 - 800}{\frac{40}{\sqrt{30}}}$$

$$= -1.643$$

Step 6. Conclusion: Accept Null Hypothesis(H_0)

Since Z_{Cal} falls in acceptable region i.e $-Z_{\alpha/2} < Z_{Cal} < Z_{\alpha/2}$.

Example 2.

Mice with an average lifespan of 32 months will live upto 40 months when fed by a certain nutritious food. If 64 mice fed on this diet have an average lifespan of 38 months and standard deviation of 5.8 months, is there any reason to believe that average lifespan is less than 40 months.

Solution:

Given data

Let μ= average lifespan of mice fed with nutritious food

Sample Standard deviation :$S = 5.8$

Sample size: $n=64$

Sample mean: $\bar{x} =38$

Specified constant:$\mu_0 =40$

Step 1. Null hypothesis: $\mu=40$

Step 2. Alternative hypothesis : $\mu<40$

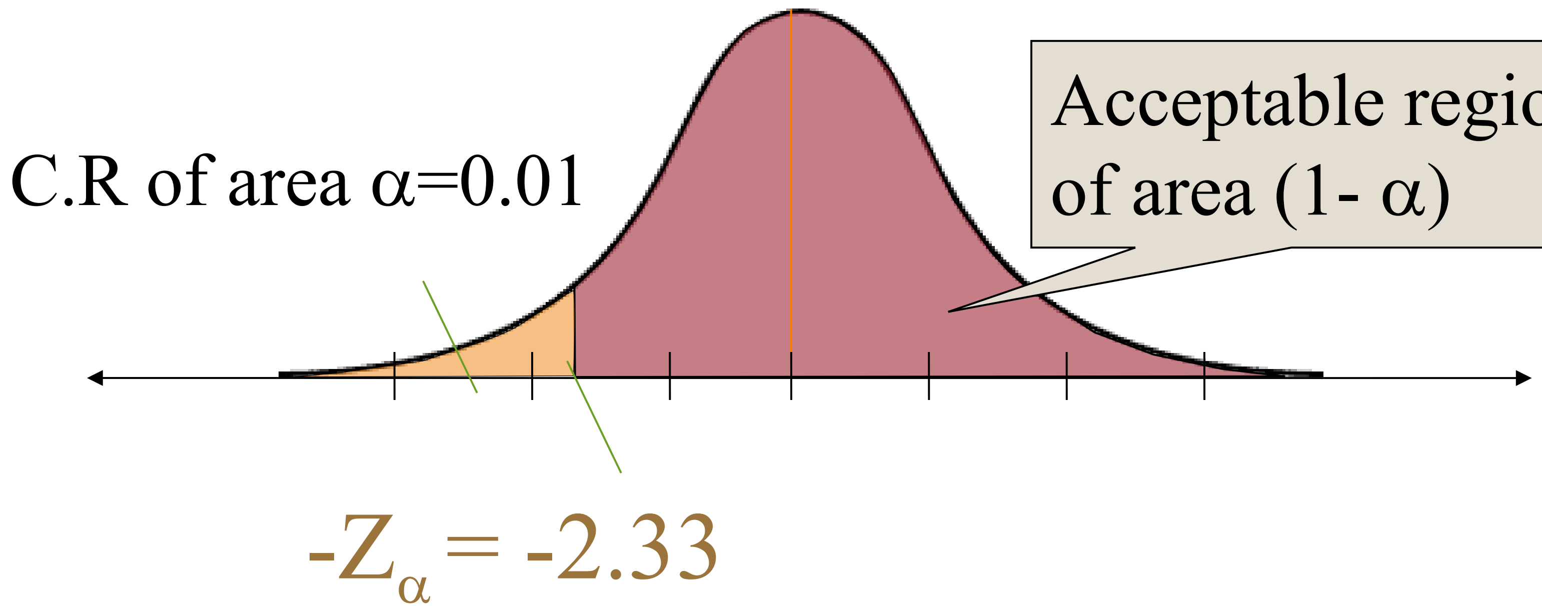
Step 3. Levl of Significance :

$\alpha=1\% =0.01$

Step 4. Critical region: Since Alternate Hypothesis is less than type (<),the test is left one tailed.

$$Z_{\alpha}=Z_{0.01}=2.33$$

Left One Tailed Test (L.O.T.T)



Step 5.Test Static

$$Z_{\mathrm{Cal}} = \frac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$$

$$= \frac{38 - 40}{\dfrac{5.8}{\sqrt{64}}}$$

$$= -2.76$$

Step 6. Decision: Reject Null Hypothesis(H_0)

Since Z_{Cal} falls in critical region i.e

$Z_{Cal} < -Z_{\alpha}$

Hence there is reason to believe that the average lifespan of mice with nutritious food is less than 40 months.

Example 3.

A machine runs on an average of 125 hours /year. A random sample of 49 machines has an annual average use of 126.9 hours with standard deviation 8.4 hours. Does this suggest to believe that machines are used on the average more than 125 hours annually at 0.05 LOS?

Solution:

Given data

Let μ= average number of hours a machine runs in an year.

Sample Standard deviation :S = 8.4

Sample size: n=49

Sample mean: $\bar{x}$ =126.9

Specified constant: μ_0 =125

Step 1. Null hypothesis: $\mu=125$

Step 2. Alternative hypothesis : $\mu>125$

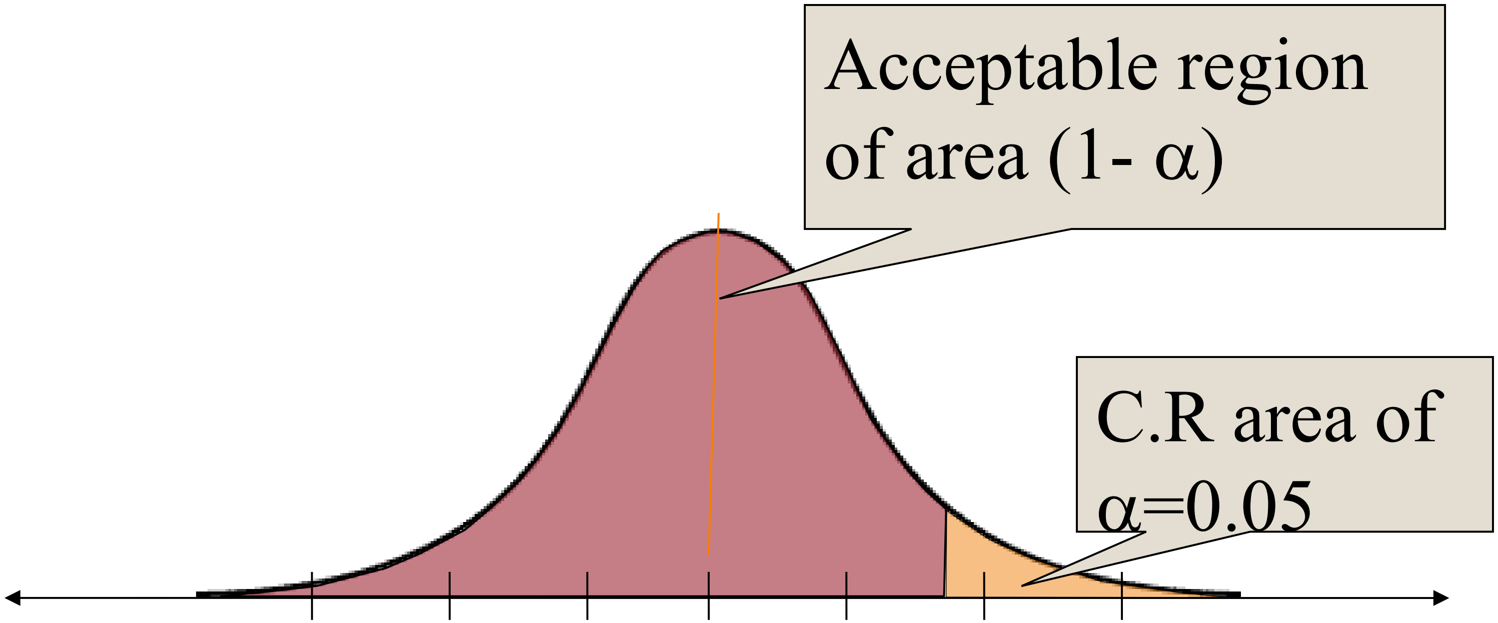
Step 3. Levl of Significance :

$\alpha=5\% =0.05$

Step 4. Critical region: Since Alternate Hypothesis is greater than type (>),the test is right one tailed.

$$Z_{\alpha}=Z_{0.05}=1.645$$

Right One Tailed Test (R.O.T.T)



$$Z_\alpha = 1.645$$

Step 5.Test Static

$$Z_{\text{Cal}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{126.9 - 125}{\frac{8.4}{\sqrt{49}}}$$

$$= 1.58$$

Step 6. Decision: Accept Null Hypothesis(H_0)

Since Z_{Cal} falls in acceptable region

i.e $Z_{Cal} < Z_{\alpha}$

Hence can not believe that machine works more than 125 hours in an year.

<u>Test of Hypothesis concerning two means.</u>

Let $\bar{x}_1$ be the mean of random sample of size n_1 drawn from normal population with mean μ_1 and variance σ_1^2.

Let $\bar{x}_2$ be the mean of random sample of size n_2 drawn from another normal population with mean μ_2 and variance σ_2^2.

To test the hypothesis for difference of means

Consider the null hypothesis $\mu_1 - \mu_2 = \delta$ = given constant.

So when $\delta = 0$ there is no difference between means.

If $\delta \neq 0$ the means of two populations are different.

In above cases, the test static

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Note: When two variance σ_1^2 and σ_2^2 are unknown, they can replaced by sample variances S_1^2 and S_2^2 provided both samples are large in this case the test static

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{S_1^{\ 2}}{n_1} + \frac{S_2^{\ 2}}{n_2}}}$$

Example 1

In a random sample of 100 tube lights produced by company A , the mean lifetime of tube light is 1190 hours with with standard deviation of 90 hours. Also in a random sample of 75 tube lights from company B the mean lifetime is 1230 hours with standard deviation of 120 hours.

Is there a difference between the mean lifetimes of the two brands of tube lights at a significance level of 0.05.

Given data for company A

Mean life time of tube lights of company A is $\bar{x}_A$=1190,

Standard deviation of tube lights of A is s_A=90.

Sample size of the tube lights of A =n_A=100.

Given data for company B

the mean life time of tube lights of company B is $\bar{x}_B = 1230$,

standard deviation of tube lights of B is $s_B = 120$

Sample size of the tube lights of B $= n_B = 75$.

Step1 Null hypothesis :H_0 : $\mu_A - \mu_B = \delta = 0$

i.e No difference

Step2 Alternate hypothesis:H_1:

$\mu_A - \mu_B \neq 0$

i.e there is difference

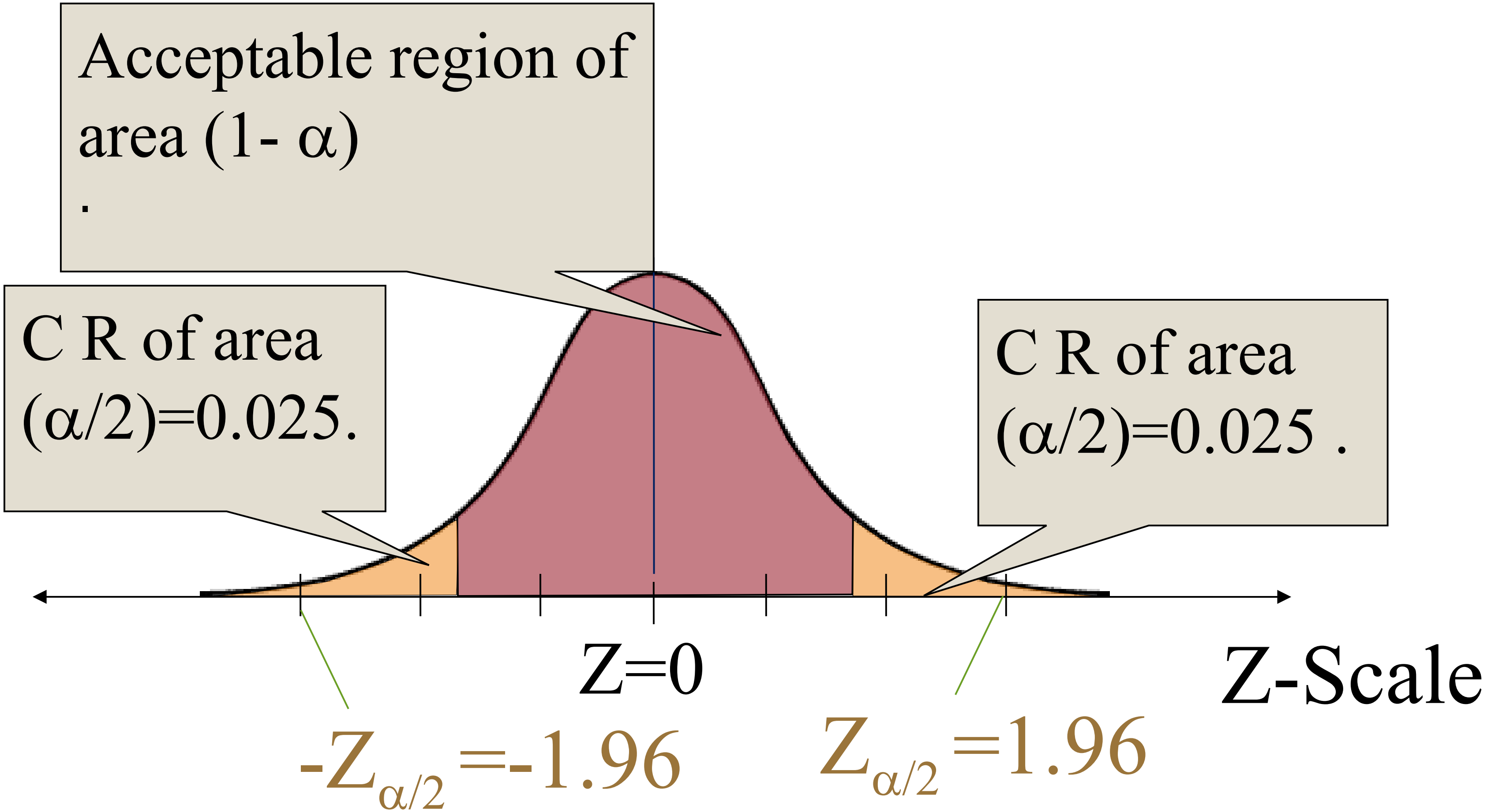
Step3: L.O.S: $\alpha = 0.05$

Step4: Critical region: Two tailed test

Step 4. Critical region: Since Alternate Hypothesis is not equal type (≠),the test is two tailed.

$$Z_{\alpha/2}=Z_{0.05/2}=Z_{0.025}=1.96$$

$$-Z_{\alpha/2}=-Z_{0.05/2}=-Z_{0.025}=-1.96$$


Acceptable region of area (1- α)
.
C R of area (α/2)=0.025.
C R of area (α/2)=0.025 .
Z=0
Z-Scale
$-Z_{\alpha/2}=-1.96$
$Z_{\alpha/2}=1.96$

Step5: Computation

$$\mu_A - \mu_B = \delta = 0$$

$$z_{cal} = \frac{(\bar{x}_A - \bar{x}_B) - \delta}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

$$= \frac{(1190 - 1230) - 0}{\sqrt{\frac{90^2}{100} + \frac{120^2}{75}}} = -2.421$$

Step 6. Conclusion: Reject Null Hypothesis(H_0)

Since Z_{Cal} falls in rejectible region i.e

$Z_{Cal} < Z_{\alpha/2}$

Hence there is difference between the mean life times of tube lights produced by company A and B

Example: 2

To test effects a new pesticide on rice production, a farm land was divided into 60 units of equal areas, all portions having identical qualities as to soil, exposure to sunlight etc. The new pesticide applied to 30 units while old pesticide to the remaining 30.

Is there reason to believe that the new pesticide is better than the old pesticide .

If the mean number of kgs of rice harvested /unit using the new pesticide(N.P) is 496.31 with the standard deviation of 17.18 kgs while the old pesticide (O.P) is 485.41 kgs and 14.73 kgs. Test the level of significance at α=0.05.

Given data for new pesticide N

Mean number of kgs of rice harvested per unit using new pesticide $\bar{x}_N$=496.31,

Standard deviation number of kgs of rice harvested per unit using new pesticide is S_N=17.18.

Number of units applied by old pesticide =n_N=30.

Given data for old pesticide O

Mean number of kgs of rice harvested per unit using old pesticide $\bar{x}_o$=485.41

Standard deviation number of kgs of rice harvested per unit using old pesticide is S_o=14.73

Number of units applied by old pesticide =n_o=30

Step1 Null hypothesis :H_0 : $\mu_N - \mu_O = \delta = 0$

i.e No difference

Step 2: Alternate hypothesis:H_1:

$\mu_N - \mu_O > 0$

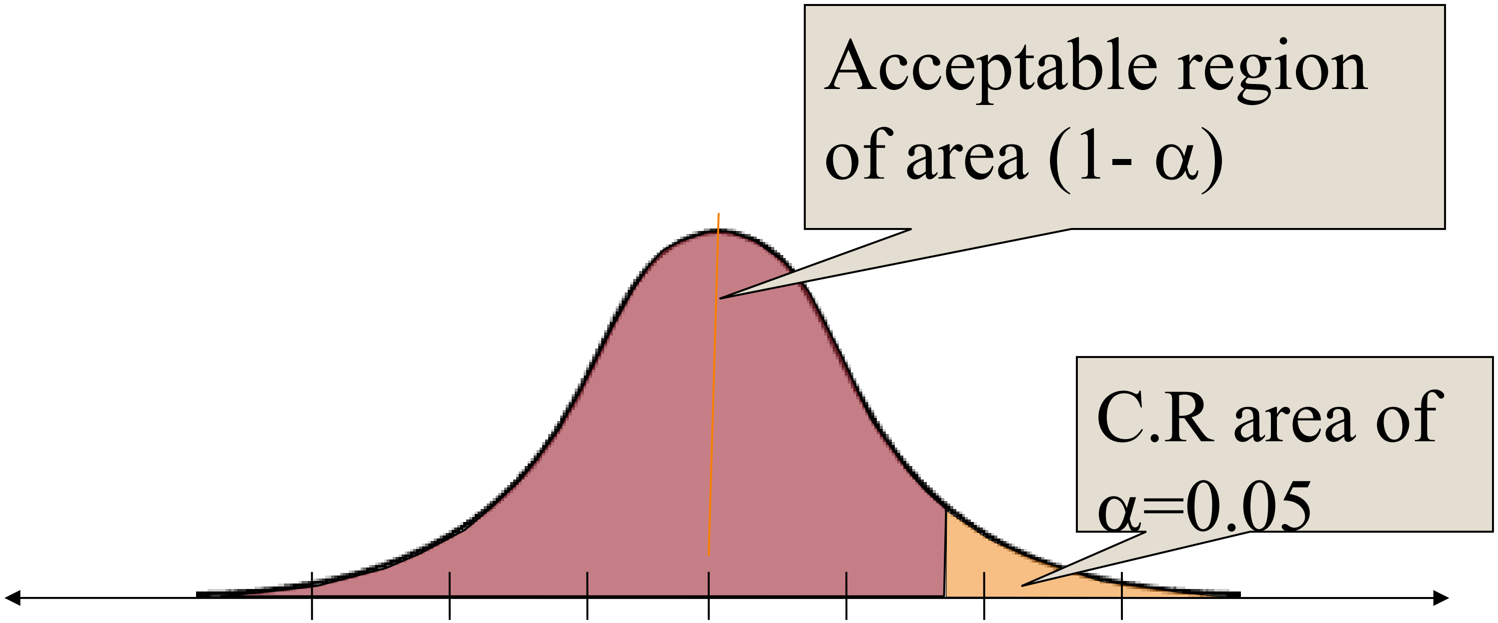
i.e New pesticide is batter than old pesticide

Step3: L.O.S: $\alpha = 0.05$

Step 4. Critical region: Since Alternate Hypothesis is greater than type (>),the tail is right one tailed.

$$Z_\alpha=Z_{0.05}=1.645$$

Right One Tailed Test (R.O.T.T)



$$Z_{\alpha} = 1.645$$

Step5: Computation

$$\mu_N - \mu_O = \delta = 0$$

$$z_{cal} = \frac{(\bar{x}_N - \bar{x}_O) - \delta}{\sqrt{\frac{s_N^2}{n_N} + \frac{s_O^2}{n_O}}}$$

$$= \frac{(496.31 - 485.41) - 0}{\sqrt{\frac{17.18^2}{30} + \frac{14.73^2}{30}}} = 2.63814$$

Step 6. Decision: Reject Null Hypothesis(H_0)

Since $Z_{Cal} = 2.638 > Z_{\alpha} = Z0.05 = 1.645$

Hence New pesticide is superior to old pesticide

Practice Problems

Example: An oceanographer wants to check whether the average depth of the ocean in a certain region is 57.4 fathoms , as previously recorded What can he conclude at the level of significance $\alpha = 0.05$, if soundings taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a standard deviation of 5.2 fathoms ?

Solution: Given $n = 40$, $\bar{x} = 59.1$ and $\sigma = 5.2$

1. **Null Hypothesis** $H_0 : \mu = 57.4$

2. **Alternative hypothesis** $H_1 : \mu \neq 57.4$

3. **Level of significance :** $\alpha = 0.05$

4. **The test statistic is** $Z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{59.1 - 57.4}{5.2/\sqrt{40}} = 2.067$

Tabulated value of Z at 5% level of significance is 1.96

Hence calculated $Z >$ tabulated Z.

$\therefore$ The null hypothesis H_0 is rejected.

Example: In a large lot of electric bulbs, the mean life and standard deviation of the bulbs are 360 hours and 90 hours respectively. A sample of 625 bulbs is chosen. It is found that the mean life and standard deviation of the bulbs in the sample are 355 hours and 90 hours respectively. Can we conclude that the sample is drawn from the given population? Test at 5% level of significance. If we assume that the population is normal and its mean is unknown, find the 98% confidence limits of the mean.

Null hypothesis H_0 :

The sample has been drawn from the population with mean $\mu = 360$ hours and $S.D = \sigma = 90$ hours.

Alternate hypothesis H_1 : $\mu \neq 360$ hours.

We shall use the two tailed test. Define the test statistic as (see Table 20.1)

$$Z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} \sim N(0, 1).$$

From the given data, we have $\bar{x} = 355$ hours, $\mu = 360$ hours, $n = 625$ and $\sigma = 90$ hours. We get

$$Z = \frac{(355 - 360)25}{90} \approx -1.389.$$

Since, $|Z| = 1.389 < 1.96$, we accept the null hypothesis, that the sample is drawn from the population (at 5% level of significance), (see Table 20.2).

The 98% confidence interval for μ is given by (see 20.19)

$$\left(\bar{x} - 2.33\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + 2.33\left(\frac{\sigma}{\sqrt{n}}\right)\right) \text{ or } \left(355 - \frac{2.33(90)}{25}, 355 + \frac{2.33(90)}{25}\right)$$

or (346.612, 363.388).

The 98% confidence limits are 346.612 and 363·388.

Example: A coin was tossed 960 times and returned heads 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance.

<u>**Solution**</u> **:** Here $n = 960$, $p =$ Probability of getting head $= 1/2$

$$\therefore\ q = 1 - p = \frac{1}{2}; \quad \mu = np = 960\left(\frac{1}{2}\right) = 480$$

$$\sigma = \sqrt{npq} = \sqrt{(np)\,q} = \sqrt{480 \times \frac{1}{2}} = \sqrt{240} = 15.49$$

$x =$ number of successes $= 183$

1. **Null Hypothesis** H_0 : The coin is unbiased
2. **Alternative Hypothesis** H_1 : The coin is biased
3. **Level of significance** : $\alpha = 0.05$

4. The test statistic is $Z = \dfrac{x-\mu}{\sigma} = \dfrac{183-480}{15.49} = \dfrac{-297}{15.49} = -19.17$

$\therefore\ |Z| = 19.17$

As $|Z| > 1.96$, the null hypothesis H_0 has to be rejected at 5% level of significance and we conclude that the coin is biased.

Example: The sizes and means of two independent random samples are 400, 225; 3.5 and 3.0 respectively. Can we conclude that the samples are drawn from the same popultion with standard deviation 1.5?

Solution We have $n_1 = 400$, $\bar{x}_1 = 3.5$, $n_2 = 225$, $\bar{x}_2 = 3.0$, $\sigma = 1.5$. Define,

Null hypothesis $H_0 : \mu_1 = \mu_2$ and $\sigma = 1.5$ (samples are drawn from the same population).

Alternate hypothesis H_1: $\mu_1 \neq \mu_2$, (two tailed test).

We have
$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{3.5 - 3.0}{(1.5)\sqrt{\frac{1}{400} + \frac{1}{225}}} = 4.0.$$

Since, $|Z| = 4 > 3$, H_0 is rejected. We conclude that in all probability, the samples are not drawn from the same population.

Example: The mean height of 80 boys, who participated in the athletic competetion in a college was 167 cm with a standard deviation of 9 cm. The mean height of the remaining 160 boys who did not participate in the atheletic competetion was 163 cm with a standard deviation of 10 cm. Test the hypothesis at 5% level of significance, whether the students who participated in atheletics are taller than the other students.

Solution We have $n_1 = 80$, $\bar{x}_1 = 167$ cm, $s_1 = 9$ cm; $n_2 = 160$, $\bar{x}_2 = 163$ cm and $s_2 = 10$ cm. Define

Null hypotheis $H_0 : \mu_1 = \mu_2$ (no significant difference between means of participating and non-participating students).

Alternate hypothesis $H_1 : \mu_1 > \mu_2$ (right tailed test : students participating in athletics are taller). The test statistic is given by

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{167 - 163}{\sqrt{\frac{81}{80} + \frac{100}{160}}} = 3.126.$$

We find $Z = 3.126 > 1.645$. Hence, we conclude that at 5% level of significance for right tailed test, the difference of means is significant and H_0 is rejected. Therefore, in all probability, the students who participated in athletics are taller than the non-participating students.

Example: Test at 0.05 L.O.S. a manufacturer's claim that the mean tensile strength (mts) of a thread A exceeds the mts of thread B by at least 12 kgs. if 50 pieces of each type of thread are tested under similar conditions yielding the following data:

	sample size	mts (kgs)	s.d. (kgs)
Type A	50	86.7	6.28
Type B	50	77.8	5.61

Hint: $H_0 : \mu_A - \mu_B \geq 12$, $H_1 : \mu_A - \mu_B < 12$, reject H_0 if $Z < Z_\alpha = -1.64$

$$Z = \frac{(86.7-77.8)-12}{\sqrt{\frac{(6.28)^2}{50} + \frac{(5.61)^2}{50}}} = -2.60, \text{ reject } H_0,$$

Accept $H_1 : \mu_A - \mu_B < 12$.

Ans: Claim not tenable.

We will discuss the following topics in the next lecture

- Test of Hypothesis concerning to one proportion.

- Test of Hypothesis concerning to difference of proportions.

Thanks for watching this video

Sreenidhi Institute of Science and Technology

AUTONOMOUS

PROBABILITY AND STATISTICS

Lecture-3 **Unit-III**

Tests of Hypothesis for Large Sample

Dr. B. VIJAYA BASKER REDDY

Faculty of Mathematics
Department of Science & Humanities
SNIST

Topics of the previous Lecture:

- Test of Hypothesis concerning one mean for large sample.

- Test of Hypothesis concerning two means for large sample.

We will discuss the following topics in this lecture

- Test of Hypothesis concerning one proportion.

- Test of Hypothesis concerning difference of proportions.

Proportion:

Proportion is the fraction of favourable cases from the total number of cases. Proportion is defined for both population and sample.

Moreover, the population proportion is estimated by the sample proportion.

Population Proportion (large sample).

Population Proportion is a fraction of population(or percentage of population) that has certain characteristics.

Population proportion is obtained by the ratio of favourable cases from the population. It is denoted by p.

The formula for population proportion is

$$p = \frac{\text{Number of favourable cases}}{\text{Total cases in Population}}$$

Sample proportion:

A sample proportion is obtained by the ratio of favourable cases from the sample.

Test of Hypothesis concerning to one proportion.

Let $\hat{p}$ be the sample proportion in a large random sample of size n, drawn population with proportion p.

Working rule Test of Hypothesis concerning to one proportion.

Step 1: $H_0 : p = p_0$

Step 2: $H_1 : p \neq p_0$ or

$H_1 : p < p_0$ or

$H_1 : p > p_0$

Step 3: level of Significance: α

Step 4: Critical Region (C.R): Find the Critical value Z_α of Z at LOS α from the normal table.

If H_1 is $p \neq p_0$ then use Two-Tailed Test.

If H_1 is $p < p_0$ then use Left-one Tailed Test.

If H_1 is $p > p_0$ then use Right-one Tailed Test.

Step5: Test statistic

$$Z_{Cal} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Step6: Decision: H_0 is rejected if Z_{Cal} value falls in Critical region.

Critical value based on z distribution

Level of significance (α%)	**10%**	**5%**	**1%**
Level of significance(α)	**0.1**	**0.05**	**0.01**
two-tailed test (α/2)	**1.645**	**1.96**	**2.58**
one-tailed test (α)	**1.28**	**1.645**	**2.33**

Example 1.

If in a random sample of 600 cars making a right turn at a certain traffic junction 157 drove into the wrong lane, test whether actually 30% of all drivers make this mistake or not at this given junction. Use (a) 0.05 (b) 0.01 L.O.S.

Solution:

Given data

Population proportion: p

Specified population proportion constant: p_0=30%=0.3

Sample size: n=600

Number of drivers drove in to the wrong line in sample: x =150 .

Sample proportion $= \hat{p} = \frac{x}{n} = \frac{157}{600}$

Case (a):

Step 1. Null hypothesis: $p=0.3$

Step 2.Alternative hypothesis :$p\neq 0.3$

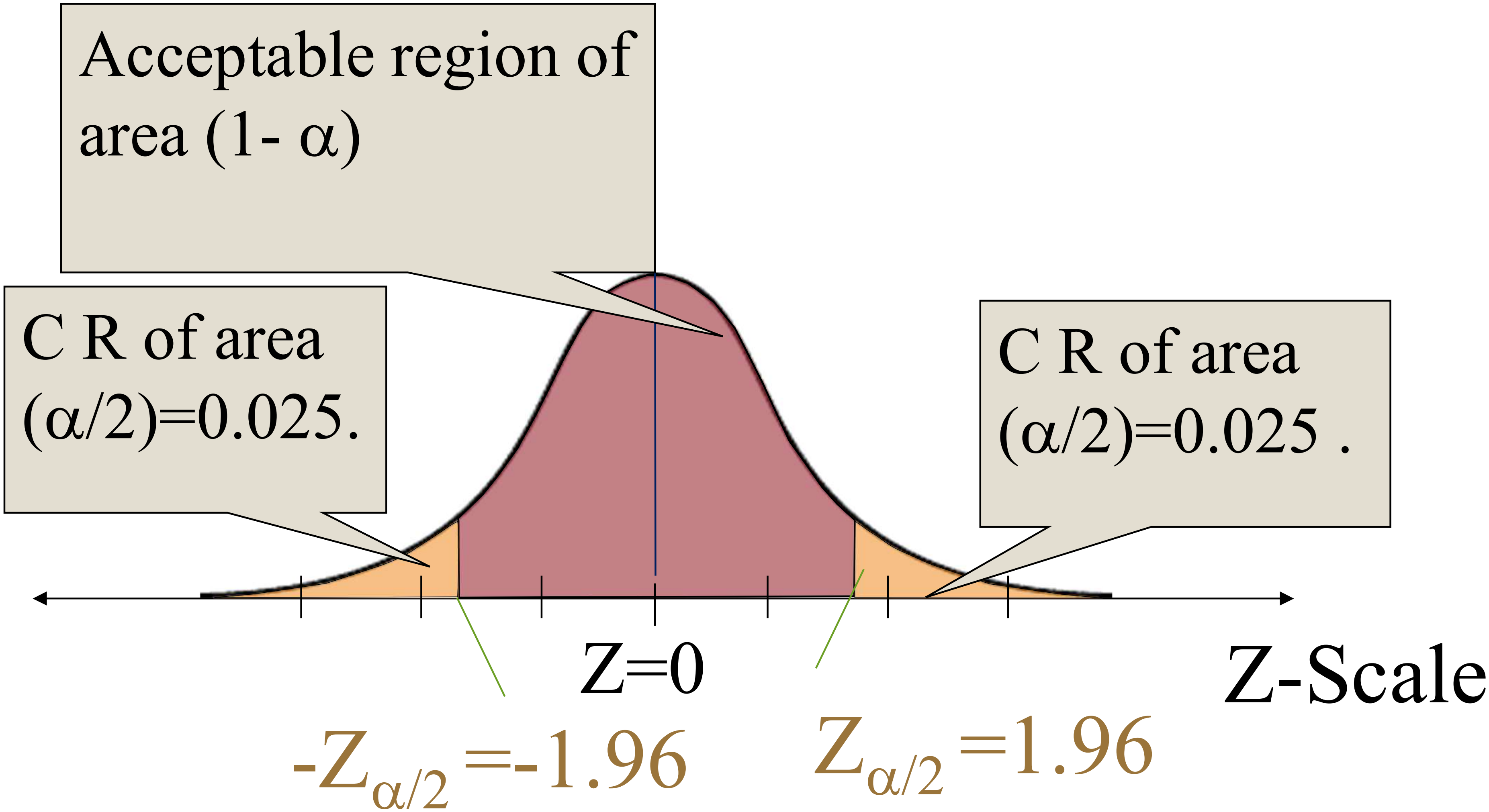
Step 3.Levl of Significance :

$\alpha=5\% =0.05$

Step 4. Critical region: Since Alternate Hypothesis is not equal type ($\neq$),the test is two tailed .

$$Z_{\alpha/2}=Z_{0.05/2}=Z_{0.025}=1.96$$

$$-Z_{\alpha/2}=-Z_{0.05/2}=-Z_{0.025}=-1.96$$


Acceptable region of area (1- α)
C R of area (α/2)=0.025.
C R of area (α/2)=0.025 .
Z=0
Z-Scale
-Zα/2 =-1.96
Zα/2 =1.96

Step 5.Test Static

$$Z_{Cal} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$= \frac{\frac{157}{600} - \frac{30}{100}}{\sqrt{\frac{30}{100} \times \frac{70}{100} \times \frac{1}{600}}}$$

$$= -2.04$$ (Approx)

Step 6. Conclusion: Reject Null Hypothesis(H_0)

Since Z_{Cal} falls in critical region region i.e $Z_{Cal} \notin (-Z_{\alpha/2}, Z_{\alpha/2})$

Case (b):

Step 1. Null hypothesis: $p=0.3$

Step 2.Alternative hypothesis : $p\neq 0.3$

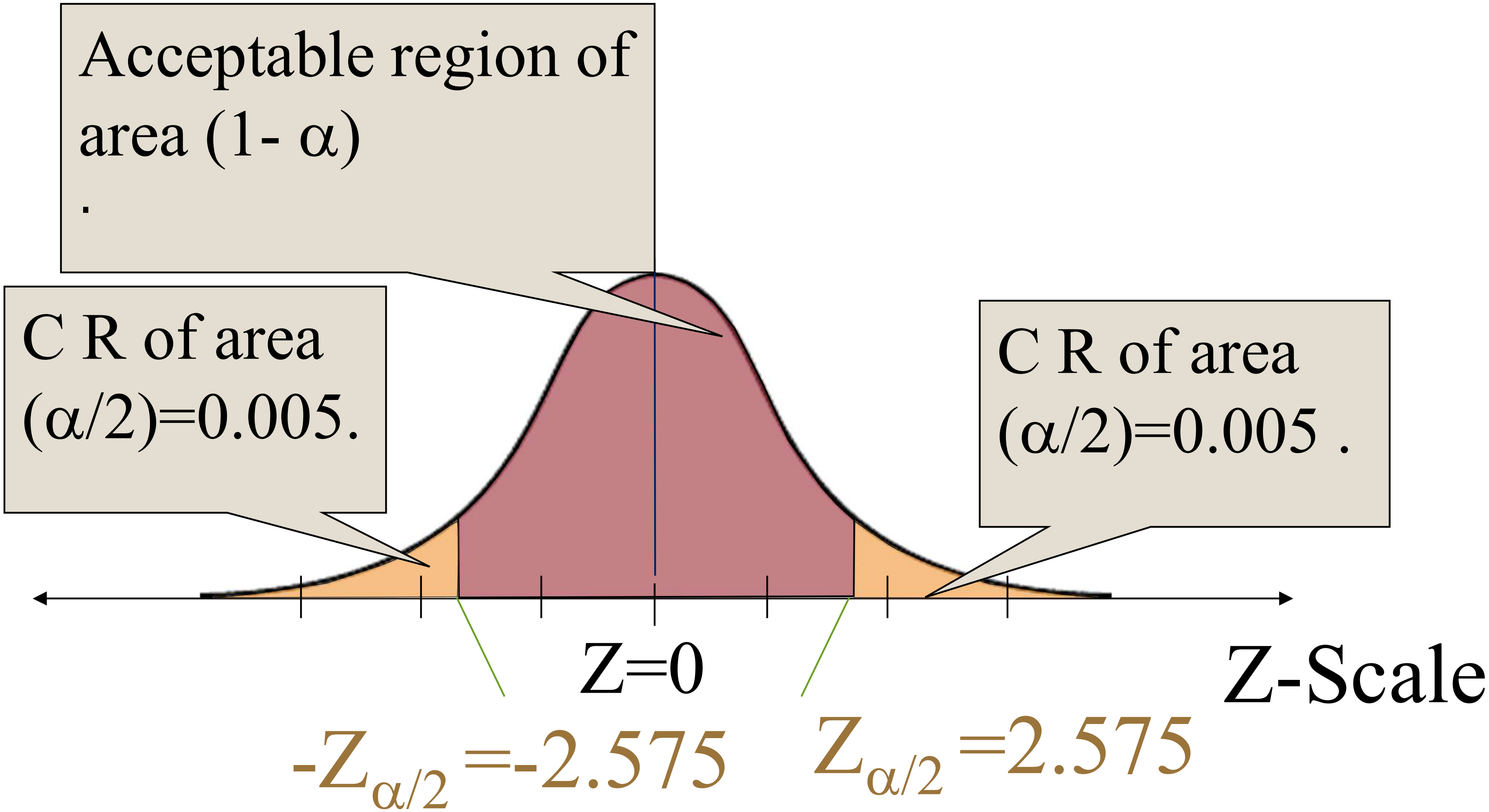
Step 3.Levl of Significance :

$\alpha=1\% =0.01$

Step 4. Critical region: Since Alternate Hypothesis is not equal type ($\neq$),the tail is two tailed .

$$Z_{\alpha/2}=Z_{0.01/2}=Z_{0.005}=2.575$$

$$-Z_{\alpha/2}=-Z_{0.01/2}=-Z_{0.005}=-2.575$$



Acceptable region of area (1- α) .
C R of area (α/2)=0.005.
C R of area (α/2)=0.005 .
Z=0
Z-Scale
-Zα/2 =-2.575
Zα/2 =2.575

Step 5.Test Static

$$Z_{Cal} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$= \frac{\frac{157}{600} - \frac{30}{100}}{\sqrt{\frac{30}{100} \times \frac{70}{100} \times \frac{1}{600}}}$$

$$= -2.04$$

Step 6. Conclusion:

The null hypothesis is accepted

Since $Z_{cal} \in (-Z_{\alpha/2}, Z_{\alpha/2})$.

Example 2.

Test the claim of a manufacturer that 95% of his ‘stabilizers’ confirm to ISI specifications, if out of a random sample of 200 stabilizers produced by this manufacturer 18 were faulty.
Use 0.01.

Solution:

Given data

Population proportion: p

Specified population proportion constant: p_0=95%.

Sample size: n=200

Number of stabilizers with ISI specifications=200-18=182

Sample proportion = $\hat{p} = 182/200$

Step 1. Null hypothesis :H_0: $p=0.95$

Step 2.Alternative hypothesis:H_1 : $p<0.95$

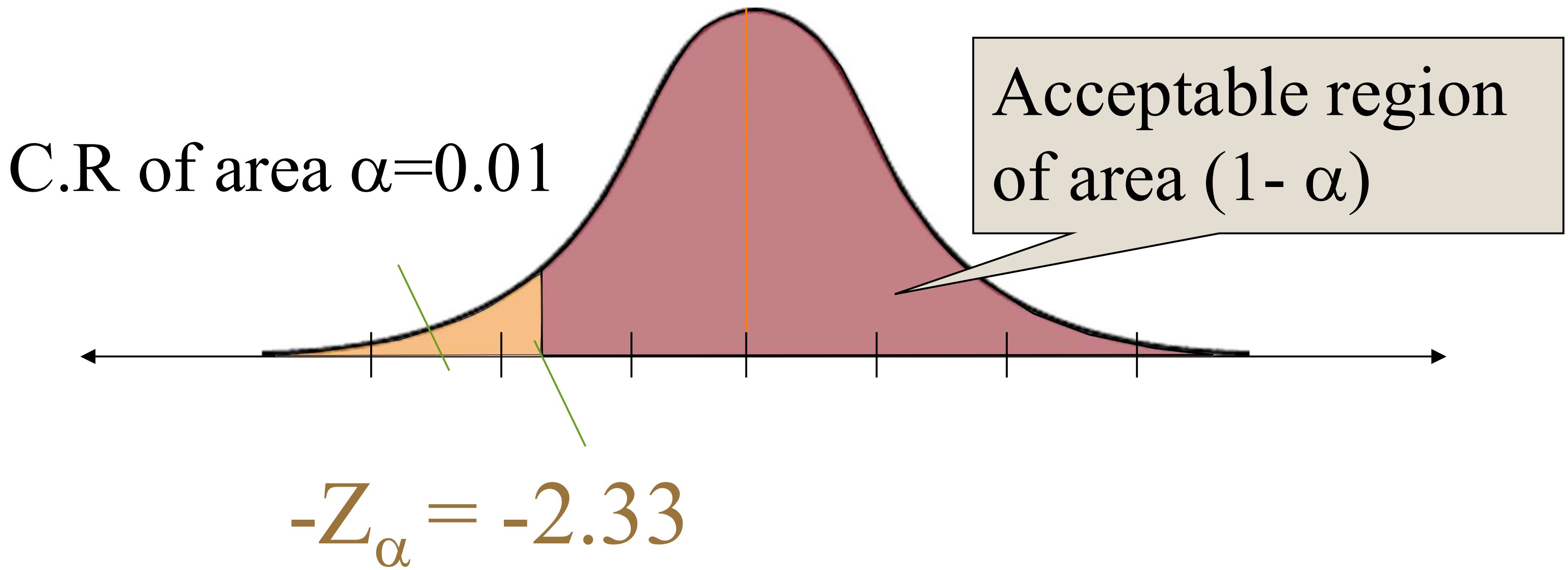
Step 3.Levl of Significance : $\alpha=1\% =0.01$

Step 4. Critical region: Since Alternate Hypothesis is less than type (<),the test is left one tailed.

$A(Z_{\alpha})=0.5-0.1=0.49$

$Z_{\alpha}=Z_{0.005}=-2.33$

Left One Tailed Test (L.O.T.T)



Step 5.Test Static

$$Z_{Cal} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{182}{200} - \frac{95}{100}}{\sqrt{\frac{95}{100} \times \frac{5}{100} \times \frac{1}{200}}}$$

$$= -2.597$$

Step 6. Decision: Reject Null Hypothesis(H_0)

Since Z_{Cal} falls in critical region i.e $Z_{Cal} < -Z_{\alpha}$

-

Example 3.

In a sample of 1,000 people in Telangana state, 540 are rice eaters and the rest are wheat eaters.
Can we assume that both rice and wheat are equally popular in this State at 1% level of significance ?

Solution:

Given data

p= Population proportion of rice eaters

=50/100=0.5

q=1-p=0.5

Sample size: n=1000

Sample proportion of rice : $\hat{p} = \frac{x}{n} = \frac{540}{1000}$

Step 1. Null hypothesis: $p=0.5$

Step 2. Alternative hypothesis : $p \neq 0.5$

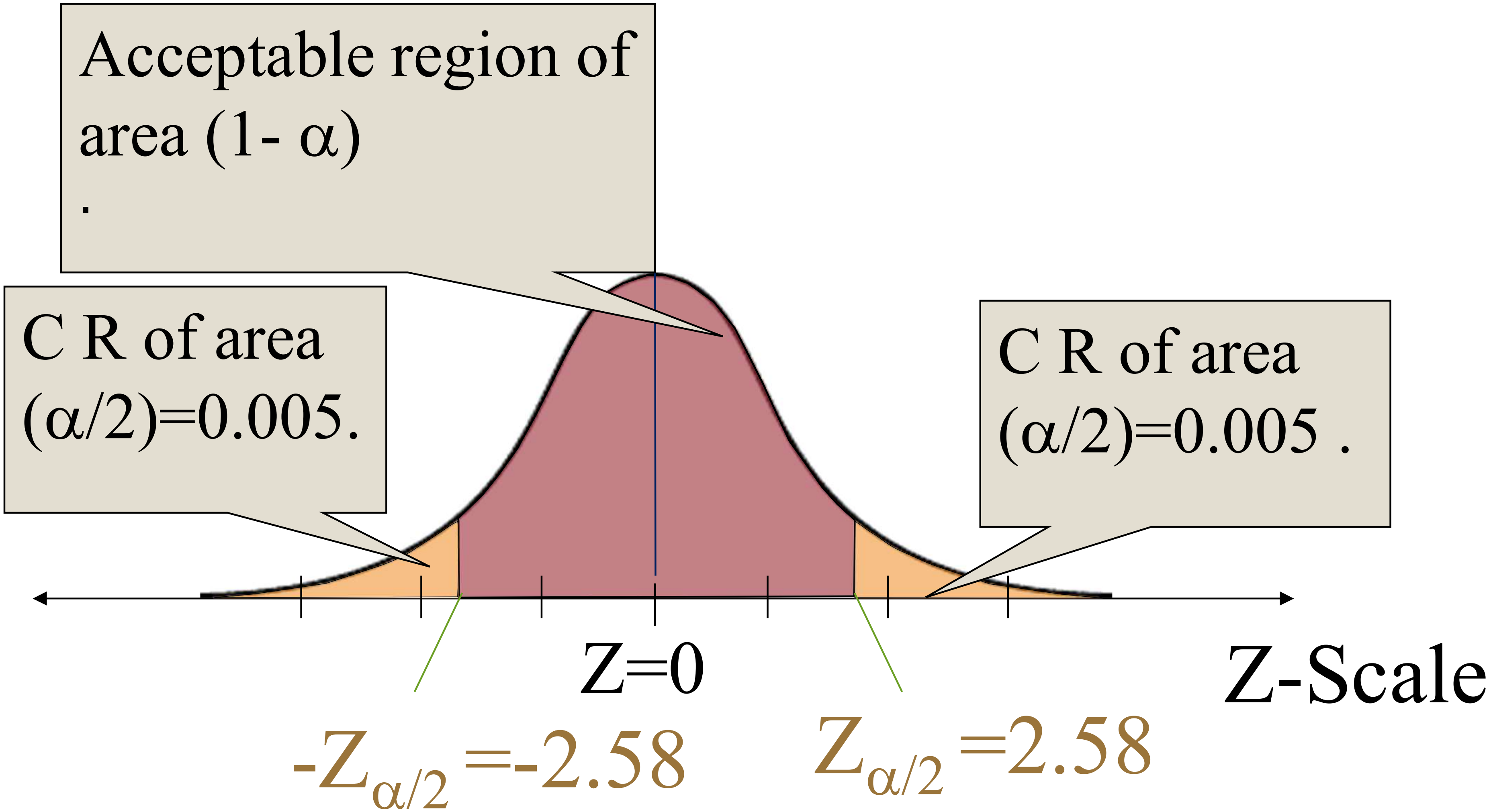
Step 3. Levl of Significance :

$\alpha = 1\% = 0.01$

Step 4. Critical region: Since Alternate Hypothesis is not equal type (≠),the tail is two tailed .

$$Z_{\alpha/2}=Z_{0.01/2}=Z_{0.005}=2.58$$

$$-Z_{\alpha/2}=-Z_{0.01/2}=-Z_{0.005}=-2.58$$


Acceptable region of area (1- α) .
C R of area (α/2)=0.005.
C R of area (α/2)=0.005 .
Z=0
Z-Scale
-Zα/2 =-2.58
Zα/2 =2.58

Step 5.Test Static

$$Z_{Cal} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$= \frac{\frac{50}{100} - \frac{540}{1000}}{\sqrt{\frac{50}{100} \times \frac{50}{100} \times \frac{1}{1000}}} = -2.53$$

Step 6. Conclusion:

The null hypothesis is accepted

Since $Z_{cal} \in (-Z_{\alpha/2}, Z_{\alpha/2})$.

Test of Hypothesis concerning two Propotions.

Suppose there are two different populations A and B and let each item of these two populations are mutually exclusive.

Let x_1 and x_2 be number of items having attribute c(success) in a random sample of size n_1 and n_2 drawn from population A and B.

	With attribute c (Success)	Without attribute c (Failure)	Total
Sample from population A	x_1	n_1-x_1	n_1
Sample from population B	x_2	n_2-x_2	n_2

$\hat{p}_1 = \frac{x_1}{n_1}$ and $\hat{p}_2 = \frac{x_2}{n_2}$ be the sample proportions of success in A and B respectively.

p_1 and p_2 population proportions of population A and B.

Working rule:

Step1 Null hypothesis : H_0 : $p_1 - p_2 = \delta = 0$

i.e No difference.

Step2 Alternate hypothesis: H_1:

$p_1 - p_2 \neq 0$ or $p_1 < p_2$ or $p_1 > p_2$

i.e there is difference.

Step3: L.O.S: α

Step4: Critical region:

Step5: Computation

$$Z_{Cal} = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right) - (p_1 - p_2)}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

Step 6: Conclusion

Example 1:

Out of two vending machines at a super market ,the first machine fails to work 13 times in 250 trials and second fails to work 7 times in 250 trials. Test 0.05 LOS whether the difference between corresponding population proportion is significant?

Given data for first machine

Sample size of first machine $=n_1=250$.

Sample proportion of first machine

$$= \hat{p}_1 = \frac{237}{250}$$

Given data for second machine

Sample size of second machine $=n_2=250$.

Sample proportion of second machine $=\quad \hat{p}_2 = \dfrac{243}{250}$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{480}{500}$$

Step1 Null hypothesis :H_0 :p_1=p_2=δ=0

i.e No difference.

Step2 Alternate hypothesis:H_1:

p_1-$p_2 \neq 0$

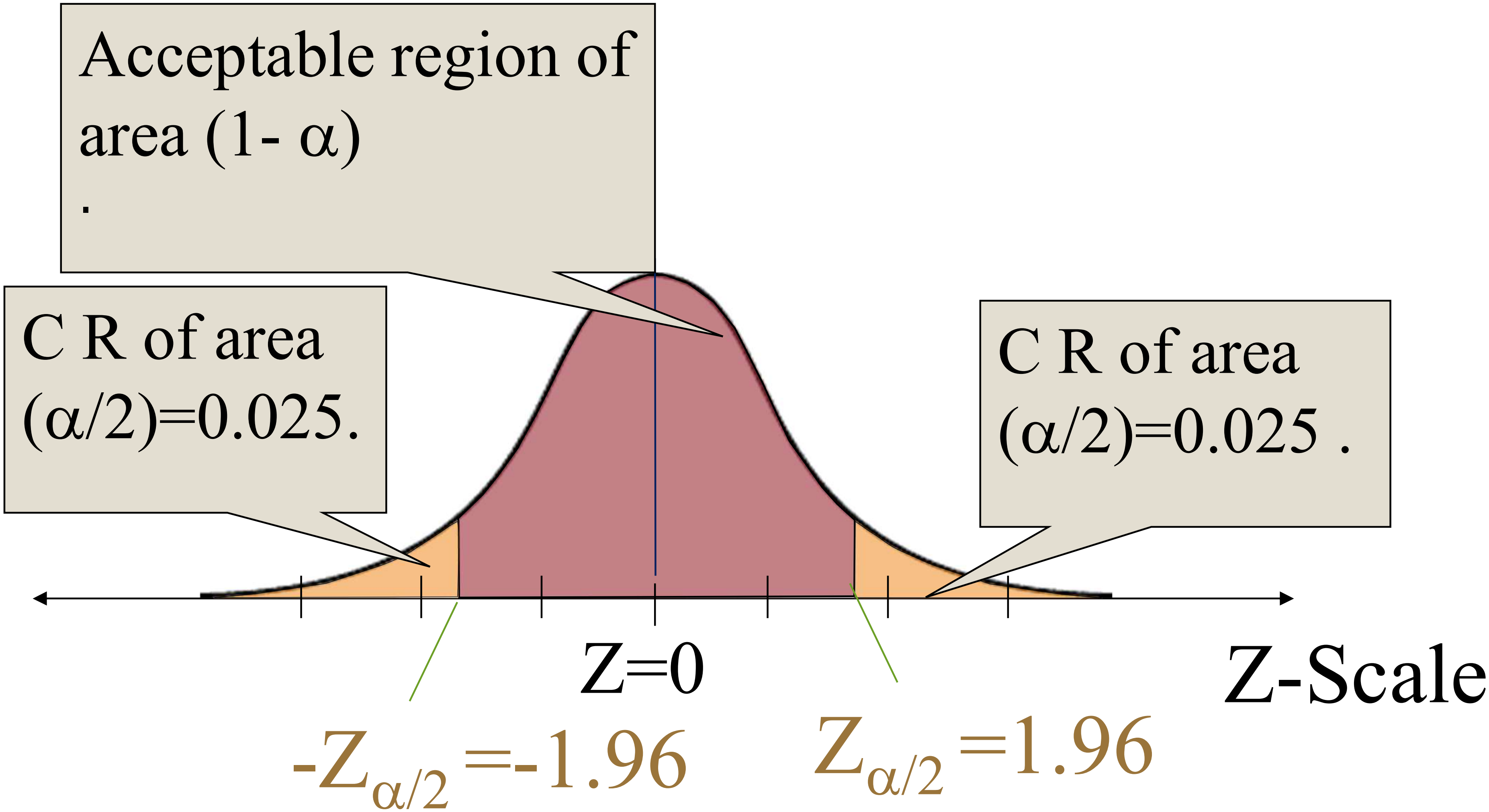
i.e there is difference.

Step3: L.O.S: α=0.05.

Step 4. Critical region: Since Alternate Hypothesis is not equal type (≠),the test is two tailed.

$$Z_{\alpha/2}=Z_{0.05/2}=Z_{0.025}=1.96$$

$$-Z_{\alpha/2}=-Z_{0.05/2}=-Z_{0.025}=-1.96$$


Acceptable region of area (1- α) .
C R of area (α/2)=0.025.
C R of area (α/2)=0.025 .
Z=0
Z-Scale
$-Z_{\alpha/2}=-1.96$
$Z_{\alpha/2}=1.96$

Step5: Computation

$$p_1 - p_2 = \delta = 0$$

$$Z_{Cal} = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right) - (p_1 - p_2)}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -1.369$$

Step 6. Conclusion: Null Hypothesis(H_0) is accepted

Hence there is no difference between population proportion of the first machine and second machine

Example: 2

If 57 of 150 patients suffering with certain disease are cured by Allopathy medicine and 33 out of 100 patients of same disease are cured by homeopathy . Is there any reason to believe that allopathic is greater than homeopathy? Use LOS 0.05.

Solution:Given data n_H=100, n_A=150 .

Sample proportion of homeopathy Medicine $\hat{p}_H = \frac{33}{100}$

Sample proportion of Allopathic Medicine $\hat{p}_A = \frac{57}{150}$

$$\hat{p} = \frac{x_A + x_H}{n_A + n_H} = \frac{90}{250}$$

Step1 Null hypothesis :H_0: $p_A - p_H = \delta = 0$

i.e No difference

Step 2: Alternate hypothesis:H_1:

$p_A > p_H$

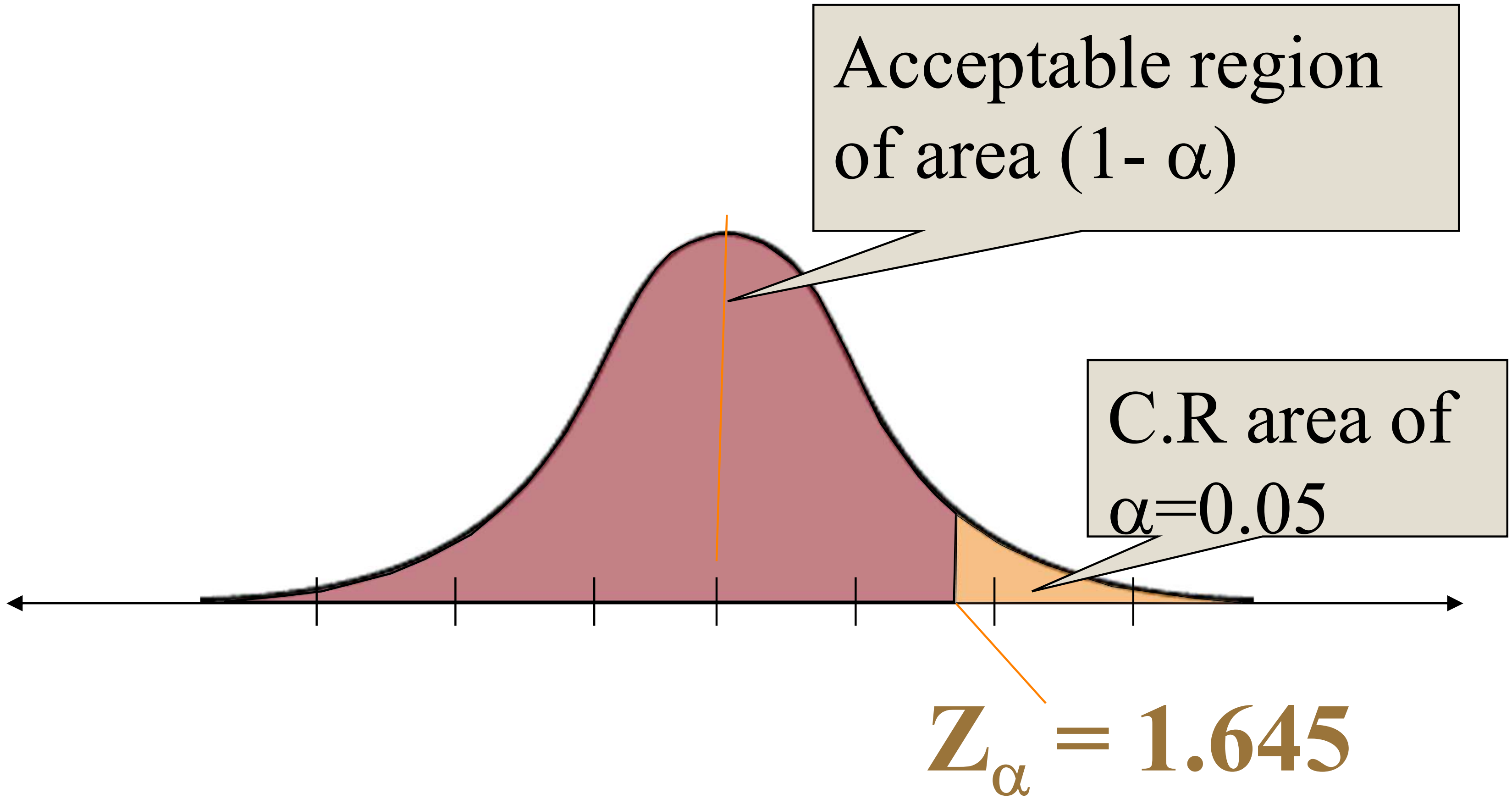
Step3: L.O.S: $\alpha = 0.05$

Step4. Critical region: Since Alternate Hypothesis is greater than type (>),the tail is right one tailed.

$$A(Z_{\alpha})=A(Z_{0.05})=0.5-0.05=0.45$$

$$Z_{\alpha}=Z_{0.05}=1.645$$

Right One Tailed Test (R.O.T.T)



Step5: Computation:

$$p_A - p_H = 0$$

$$Z_{Cal} = \frac{(\hat{p}_A - \hat{p}_H) - (p_A - p_H)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_H}\right)}} = 0.8068$$

Step 6. Decision: The null hypothesis is accepted

i.e no significant difference between Allopathic and Homeopathic medicine.

Example: 3

A question in true /false quiz is considered to be smart if it discriminates between intelligent person(IP) & average person (AP). suppose, 205 of 250 intelligent persons (IP's) 137 of 250 average person (AP's) answer a quiz question correctly.

Test the 0.01 level of significance, whether for a given question, the proportion of correct answers can be expected to be at least 15% lower among IP's than among AP's.

Given data

$n_I=250$ $n_A=250$

$\hat{p}_I = \frac{205}{250}$ and $\hat{p}_A = \frac{137}{250}$

$$\hat{p} = \frac{x_A + x_I}{n_A + n_I} = \frac{342}{500}$$

Step1 Null hypothesis :H_0:p_I-p_A=0.15.

Step 2: Alternate hypothesis:H_1:

p_I-p_A>0.15

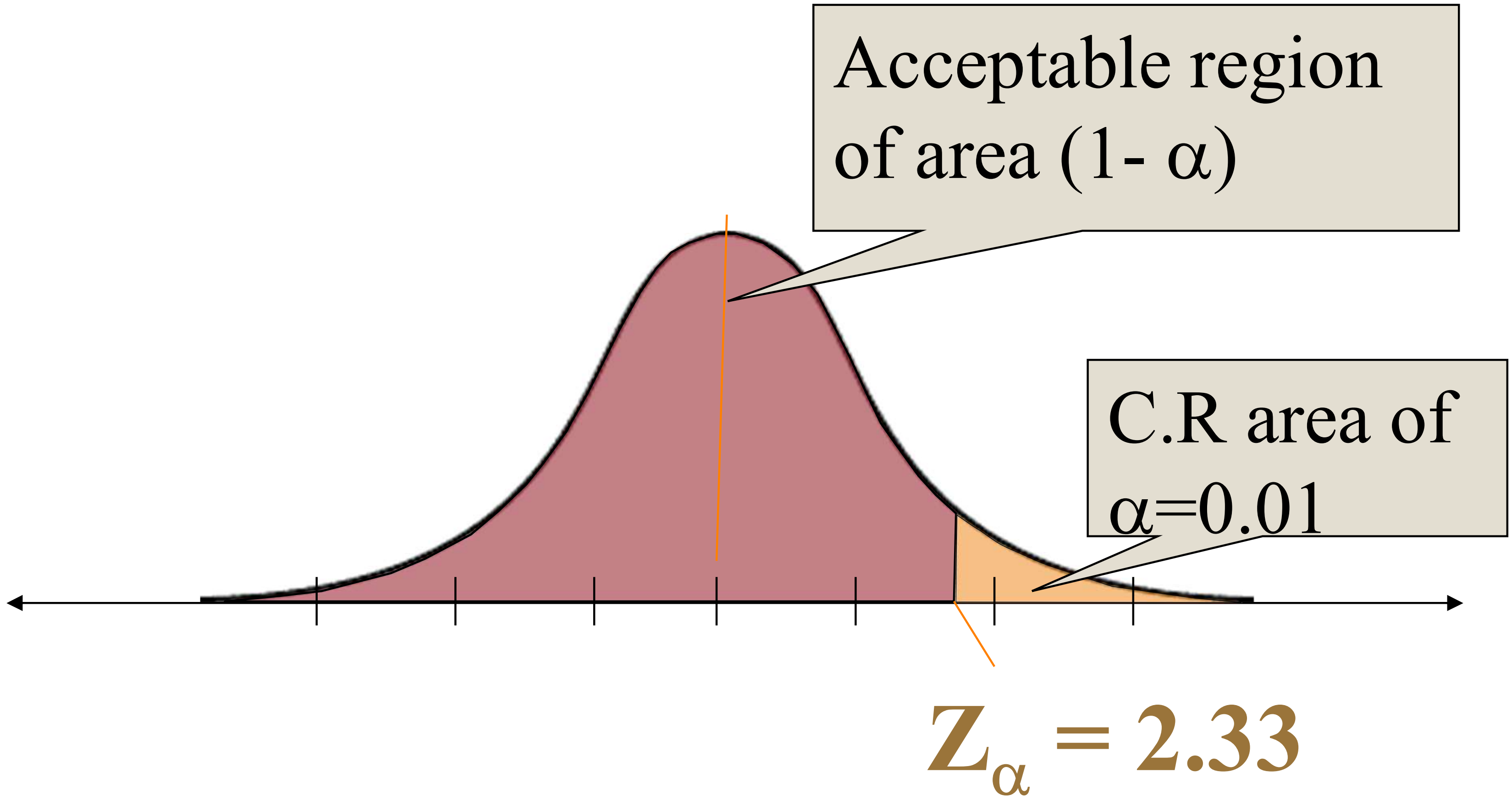
Step3: L.O.S: α=0.01

Step 4. Critical region: Since Alternate Hypothesis is greater than type (>),the tail is right one tailed.

$$A(Z_\alpha)=A(Z_{0.01})=0.5-0.01=0.49$$

$$Z_\alpha=Z_{0.01}=2.33$$

Right One Tailed Test (R.O.T.T)



Step5: Computation:

$p_I - p_A = 0.15$

$$Z_{Cal} = \frac{(\hat{p}_I - \hat{p}_A) - (p_I - p_A)}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{n_I} + \frac{1}{n_A}\right)}}$$

$$= 3.068$$

Step 6. Decision: The null hypothesis is rejected i.e The proportion of correct answer among IP's is 15% higher than the AP's.

Thanks for watching this video