PROBABILITY AND STATISTICS (UCS401)

Lecture-32

(Testing of Hypothesis-Large sample)

Testing of Hypothesis (Unit –VII)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Dr. Rajanish Rai
Assistant Professor
School of Mathematics
Thapar Institute of Engineering and Technology, Patiala

Rajanish Rai

Large-Sample Testing of Hypothesis

Testing of Hypothesis about population parameters -:

In practical situation, statistical inference can involve either estimating a population parameter or making decisions about the value of the parameter.

The reasoning used in a statistical test of Hypothesis is similar to the process in a Court trail.

In trying a person for theft, the Court must decide between innocent and guilty. As the trails begins, the accused person is assumed to be innocent (initial assumption). The prosecution collects and presents all available evidence in an attempt to contradict the innocent hypothesis and hence obtain a conviction.

- If there is enough evidence against innocence, the Court will reject the innocence hypothesis, and declare the defendent Guilty.

- If the prosecution Does Not present enough evidence to prove the defendant guilty, the court will find him not guilty.

H_0 (Initial assumption)

$\Downarrow$

Give evidence Enough / Not Enough

$\Downarrow$

Decision (H_0) True / False.

Notice that this does not prove that the defendant is innocent, but merely that there was not enough evidence to conclude that the defendant was guilty.

We use this same type of reasoning to explain the basic concept of Hypothesis testing.

Large sample and small sample testing —:

We devide the Hypothesis testing into two classes:

- When sample sizes are large (when $n \geq 30$).
  - We always use Z — test of Hypothesis.
- When the sample sizes are small (when $n < 30$)
  we always use student's t — test of Hypothesis.

$$X \sim N(\mu, \sigma^2)$$

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad \text{or} \quad \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

A statistical test of Hypothesis -:

A statistical test of Hypothesis consists of four parts:

(i) The null and alternative hypothesis, denoted by H_0 and H_a ($\text{or } H_1$) (Initial assumption).

(ii) The test statistics and its p-value (evidence)

(iii) The rejection region (Enough or Not).

(iv) The Conclusion (Decision).

When you specify these four elements, you define a particular test.

(i) The null and alternative Hypothesis -:

Alternative hypothesis generally the hypothesis that the researcher wishes to support and the Null Hypothesis. ⇒ Contradiction of the alternative hypothesis.

Note that :- It is easier to show support for the alternative hypothesis by proving that the null hypothesis is false.

Hence, the statistical researcher always begins by assuming that the null Hypothesis H_0 is true.

Example-(1) You wish to show that the average hourly wages of Carpenters in the state of USA is different from \$ 14, which is the natinal average.

Null Hypothesis H_0 ; $\mu = 14$

Alternative Hypothesis H_1 $\mu \neq 14$ (Two tailed test for Hypothesis.)

Example (2) A miling process currently produces an average 3% defective. You are interested in showing that a sample adjustment on a machine will decrease p, the proportion of defective produced in the milling process.

Null Hypothesis H_0 ; $p = 0.03$

Alternative Hypothesis H_1 $p < 0.03$

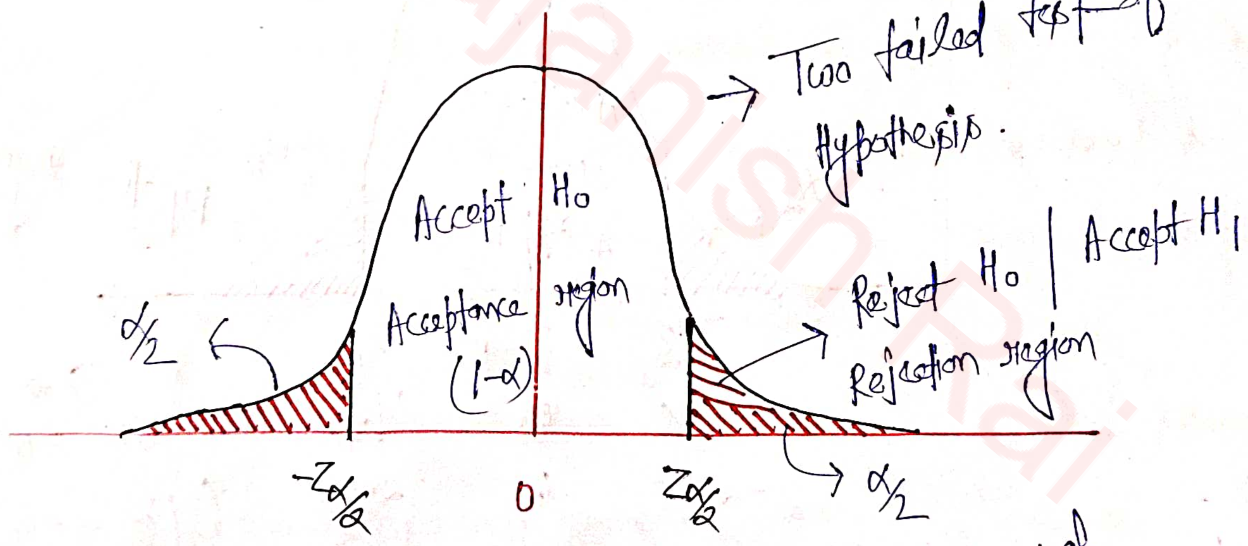
(one tailed test of Hypothesis.)

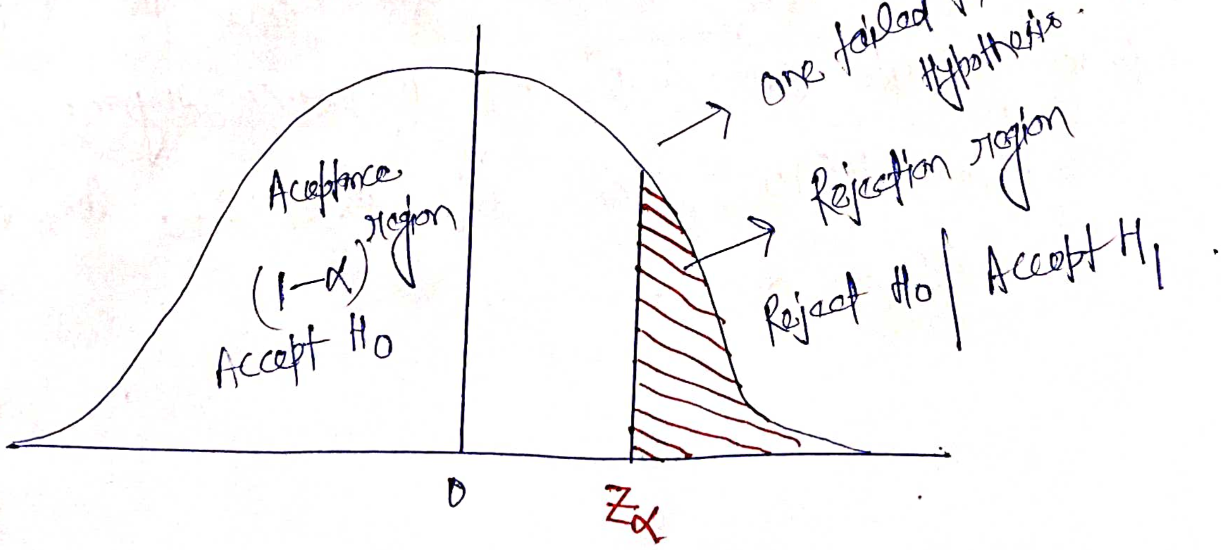
(2) **Test statistics :-** A single number (evidence) calculated from sample data.

p-value (probability of rejection) : a probability calculated using the test statistics.

Note :- Either or both of these measures act as decision makers for the researcher in deciding whether to reject or accept H_0.

(3) **Rejection region :-**

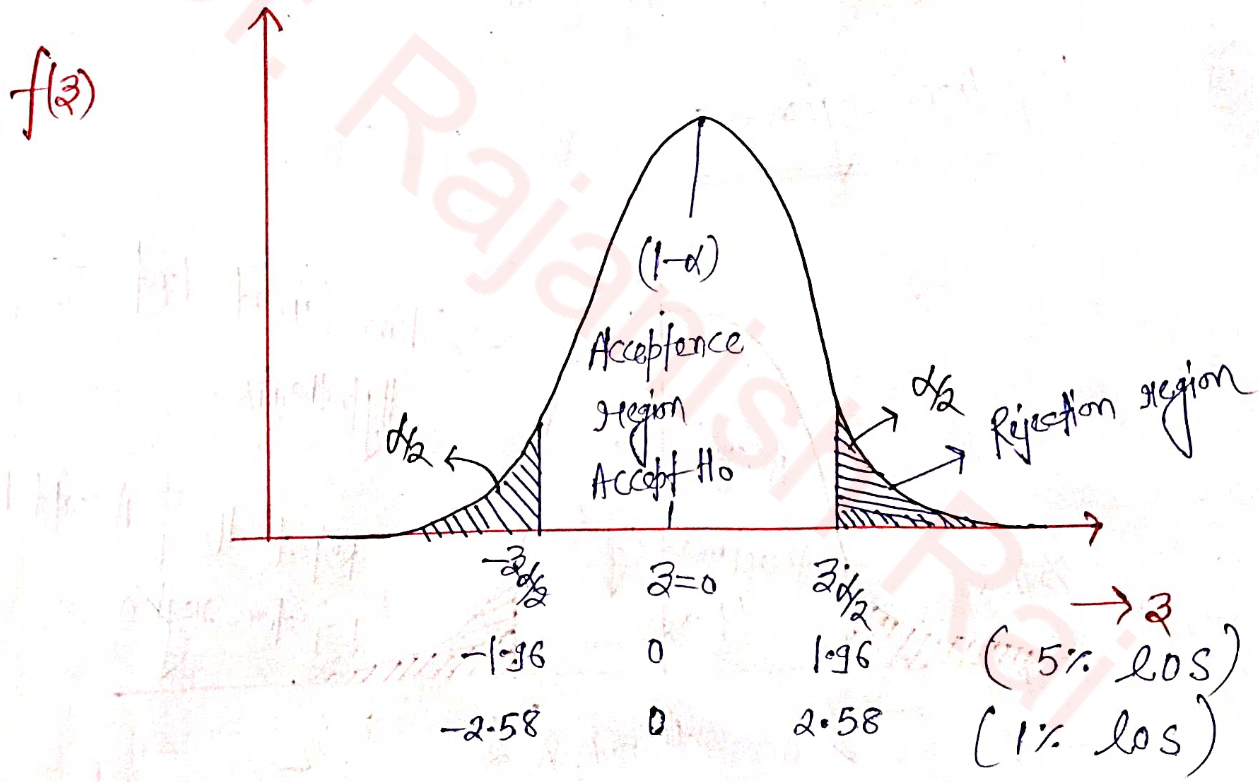




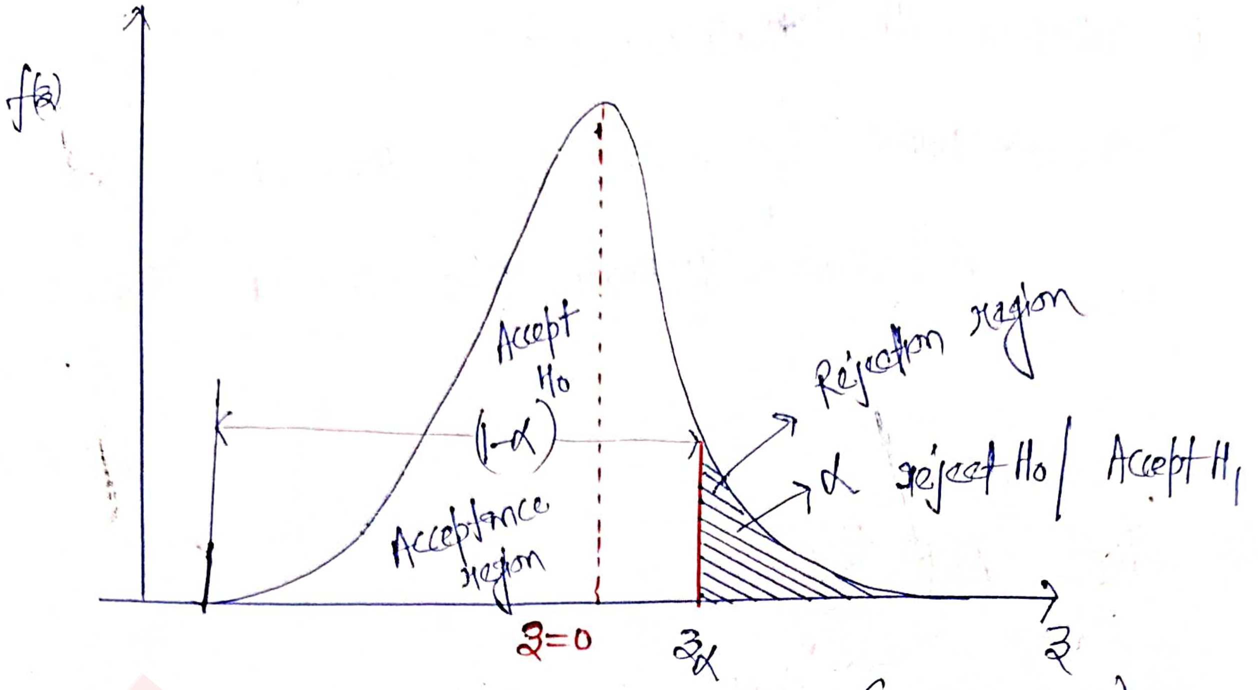
(4) Drawing the conclusion as:

- Reject H_0 and conclude that H_1 is true.
- Accept (do not reject) H_0 as true.

Note that:- For two tailed test, using $\alpha = 0.05$, the critical values cut off area of $\alpha/2 = 0.025$ in the right and left tails. These values are $z = \pm 1.96$ while for $\alpha = 0.01$, these values are $z = \pm 2.58$.



(ii) For one tailed test, using $\alpha = 0.05$, the critical value separate the rejection and acceptance regions is $z = 1.645$, while for $\alpha = 0.01$, the value is $Z_\alpha = 2.33$.



0	1.645	(5% los)
0	2.33	(1% los).

Question :- Consider a random sample of 100 USA Carpenters produces a sample mean \$15 with standard deviation 2. You wish to show that the average hourly wages of carpenters in the state of USA is different from \$14, which is the national average. Test the appropriite hypothesis using $\alpha = 0.05$ (95% Confidence interval):

Solution :-

Given that

$\mu = 14$

$\bar{x} = 15$

$s = 2$

$n = 100$ (>30) large sample.

(i) Define the Hypothesis :

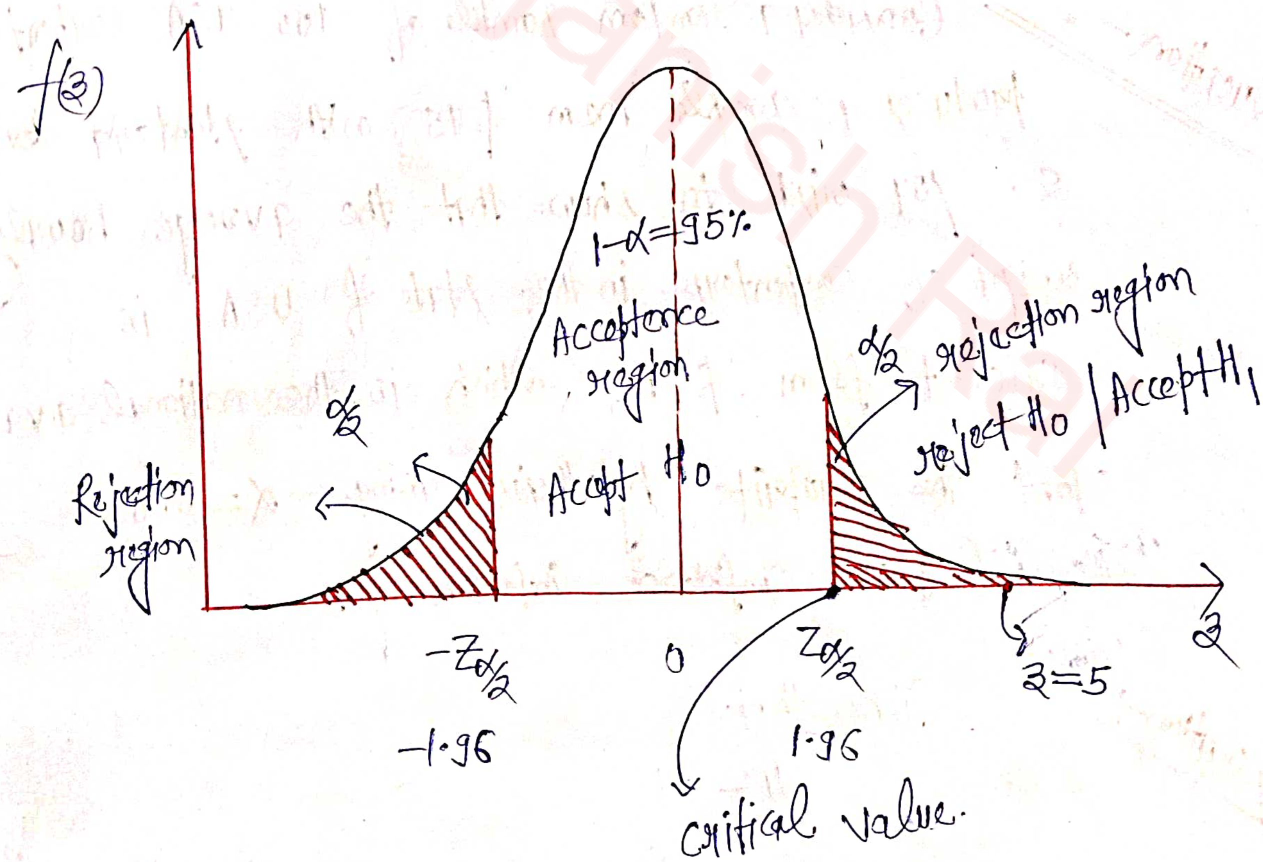
Null Hypothesis H_0 ; $\mu = 14$

Alternative Hypothesis H_1 : $\mu \neq 14$ (Two tailed test of Hypothesis)

(2) Test - Statistics

$$z = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{15-14}{2/\sqrt{100}} = 5$$

p value : $P(z>5) + P(z<-5) \approx 0$



(3) To decide whether to reject or accept H_0

If p-value is less than or equal to preassigned level of significance α, i.e. $(p \leq \alpha)$ then H_0 can be rejected.

or

Calculate the critical value.

Here for this two-tailed test $\alpha = 0.05$, the critical value is 1.96.

(4) **Conclusion-:**

- If the test statistics falls into the rejection region, the null Hypothesis is rejected.

- If the test statistics falls into acceptance region, the either null Hypothesis is accepted or the test is judged to be inconclusive.

"After comparing the test statistics value $z=5$ and critical value $Z_{\alpha/2} = 1.96$, we REJECT H_0 and conclude that H_1 is accepted, i.e., average hourly wages of carpenters in the state of USA is different."

#

Question :- The average weekly earnings for female social workers is \$ 670. Do men in the same positions have average weekly earnings that are higher than those for women ?. A random sample of $n = 40$ male social workers showed mean is \$ 725 and standard deviation \$ 102. Test the appropriate Hypothesis using $\alpha = 0.01$ (99% C.I.).

Solution :- Given that

$n = 40$ (large sample)

$\bar{x} = 725$

$s = 102$

You would like to show that the average weekly earnings for men are higher than \$ 670, the women average. If μ is the average weekly earnings for male social workers.

(i) Define Hypothesis :-

(a) Null hypothesis H_0 ; $\mu = 670$.

(b) Alternative hypothesis H_1 ; $\mu > 670$

(one tailed test of Hypothesis).

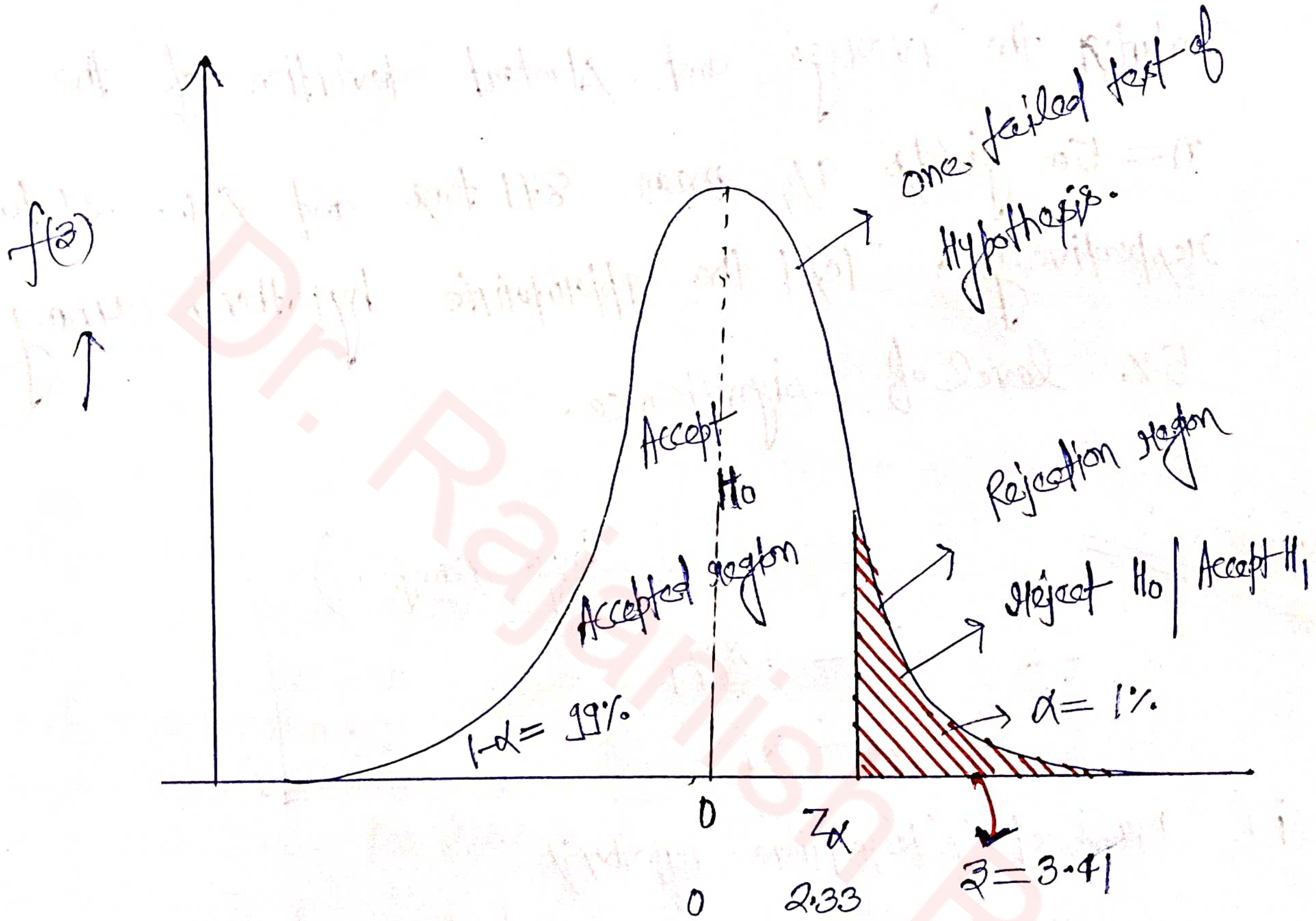
(2) Test - statistics :

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{725 - 670}{102/\sqrt{40}} = 3.41$$

$$\boxed{z = 3.41}$$

③ Rejection region :-

For this one-tailed test, $\alpha = 0.01$, the critical value Z_α is 2.33.



④ Conclusion :-

Compare the test statistics value $z = 3.41$ and critical value $Z_\alpha = 2.33$, so we reject H_0 and conclude that H_1 is accepted, i.e., the average weekly earnings for male social workers are higher than the average for female social workers.

Question :- The daily yield for a local chemical plant has averaged 880 tons for the last several years. The quality control manager would like to know whether this average has changed in recent months. She randomly select 50 days from the computer database and computes the average and standard deviation of the $n = 50$ yields as mean 871 tons and S.D. 21 tons, respectively. Test the appropriate hypothesis using 5% level of significance.

Solution :-

$$n = 50 \text{ (large sample)}$$

$$\bar{x} = 871$$

$$s = 21$$

(1) Null and alternative hypothesis

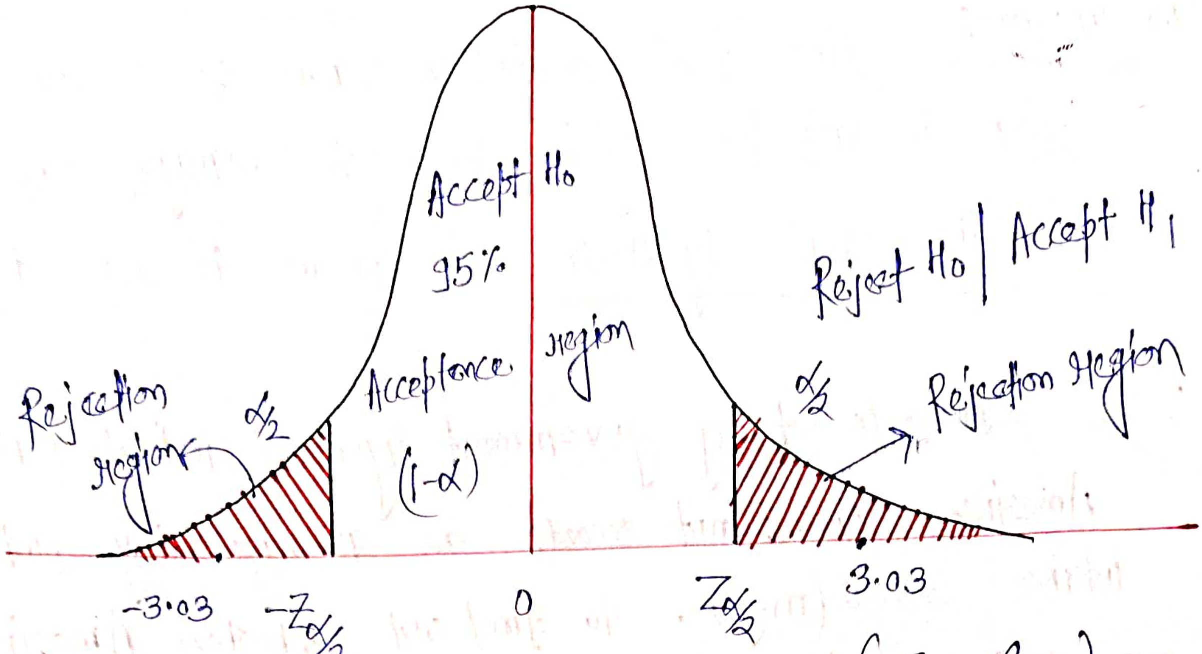
$$H_0 : \mu = 880$$

$$H_1 : \mu \neq 880 \quad \text{(Two tailed test of Hypothesis)}$$

(2) Test statistics :-

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

$$\boxed{z = -3.03}$$



-1.96	0	1.96	(5% los)
-2.58	0	2.58	(1% los)

$$p \text{ value} = P(Z > 3.03) + P(Z < -3.03)$$

$$= 0.0024$$

③ Rejection region -:

For this two tailed test, $\alpha = 0.05$, so the critical value of $\alpha/2$ in the right and left tails are $Z_{\alpha/2} = \pm 1.96$

while for $\alpha = 0.01 \Rightarrow Z_{\alpha/2} = \pm 2.58$

Based on p-value -: If p value $\leq \alpha = 0.05$, H_0 can be rejected. In this example, you can reject H_0 at either the 1% or 5% level of significance.

(7) Conclusion :-

Since test-statistics value $z = -3.23$ falls in rejection region, i.e., the manager can reject the null hypothesis and conclude it has changed.

Question :- Standards set by government agencies indicate that Americans should not exceeds an average daily sodium intake 3300 (m.g.). To find out whether Americans are exceeding this limits, a sample of 100 Americans is selected and the mean and standard deviation of daily sodium intake are found to be 3400 mg and 1100 mg, respectively. Use 5% level of significance to conduct a test of Hypothesis.

Solution :- Given that

$\mu = 3300$

$\bar{x} = 3400$

$s = 1100$

$n = 100$ (large sample - Z test).

(i) Define Hypothesis :-

(a) Null Hypothesis

$H_0 : \mu = 3300$

(b) Alternative Hypothesis

$H_1 : \mu > 3300$ (one tailed test).

② Test statistics :-

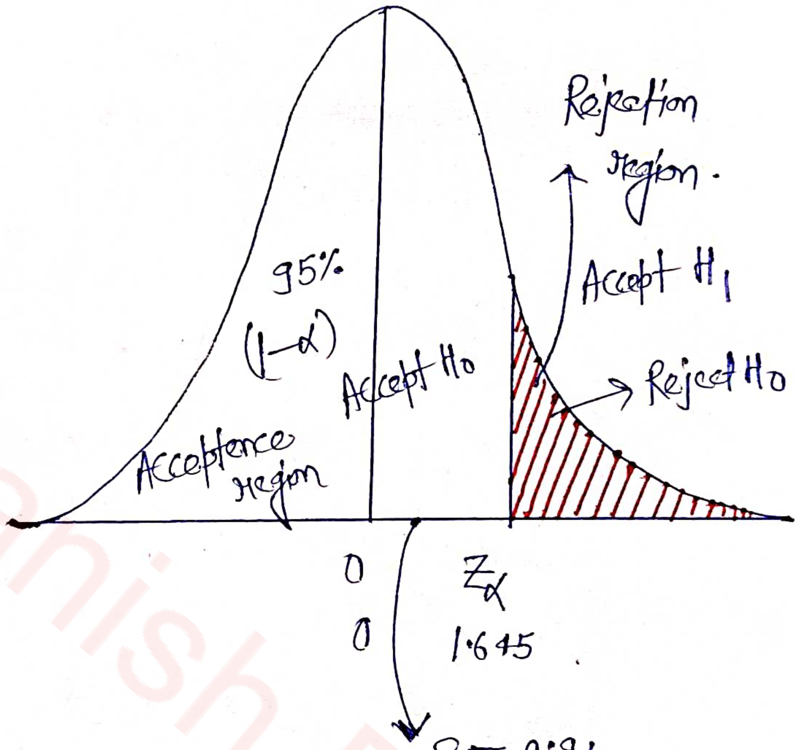
$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{3100 - 3300}{1100/\sqrt{100}} = 0.91$$

p value :

$$P(z > 0.91) = 0.1814$$

$\alpha = 0.05$

$Z_\alpha = 1.645$



③ Critical region :-

For one failed at 5% level of significance The critical value of z is 1.645.

④ Conclusion :-

Since $0.91 < 1.645 = Z_\alpha$

or

$\because$ p value $= 0.1814 \not< 0.05 = \alpha$

$\Rightarrow$ Ho is not rejected and result are not statistically significant. There is not enough evidence to indicate that average daily sodium intake exceeds 3300 mg.