Large sample – Test of Hypothesis for difference between two population mean.

Large and small sample testing –:

We devide the hypothesis testing into two classes:

- When the 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.

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 (or H_1).

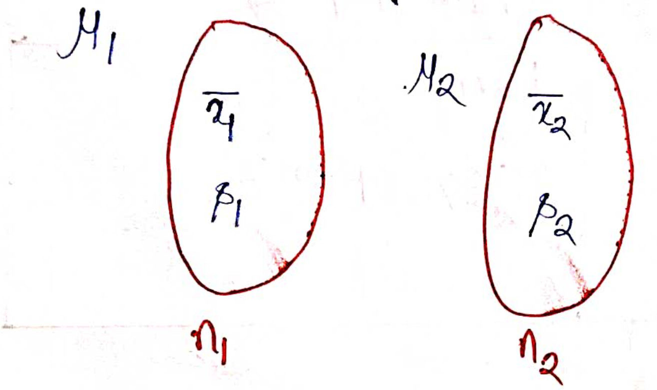
(ii) The test statistics and its p– value.

(iii) The rejection region.

(iv) The conclusion

many decision involve a comparison of two population means:

- For example, state government is interested to introduced the NEW POLICIES in schools by replacing the offline mode with the online computerized mode. To determine whether policies are significant or not.



Large-sample statistical test for $\mu_1 - \mu_2$ -:

(i) Define the Hypothesis -:

Null hypothesis -:

$H_0 : \mu_1 - \mu_2 = D_0$ where D_0 is some specified difference that you wish to test. [In most cases $D_0 = 0$]

Alternative hypothesis -:

One tailed test	Two-tailed test
$H_1 : \mu_1 - \mu_2 > D_0$ or $H_1 : \mu_1 - \mu_2 < D_0$	$H_1 : \mu_1 - \mu_2 \neq D_0$.

② Test statistics –:

$$X \sim N(\mu_1, \sigma_1^2) \qquad Y \sim N(\mu_2, \sigma_2^2)$$

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \qquad \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

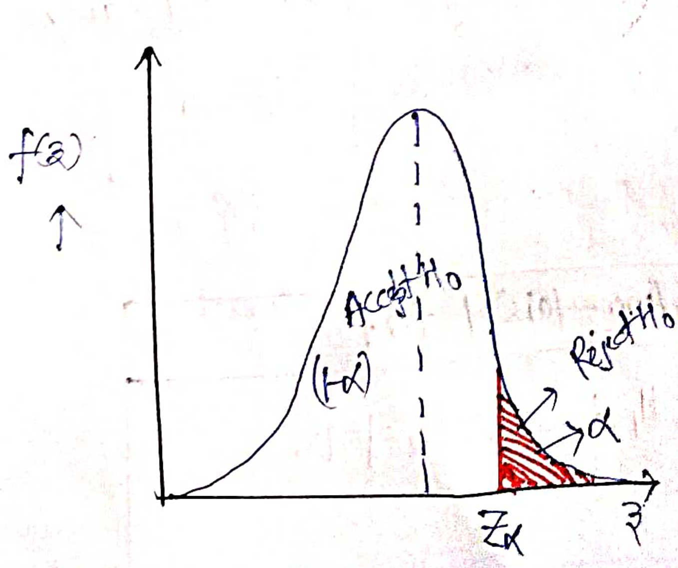
or p-value

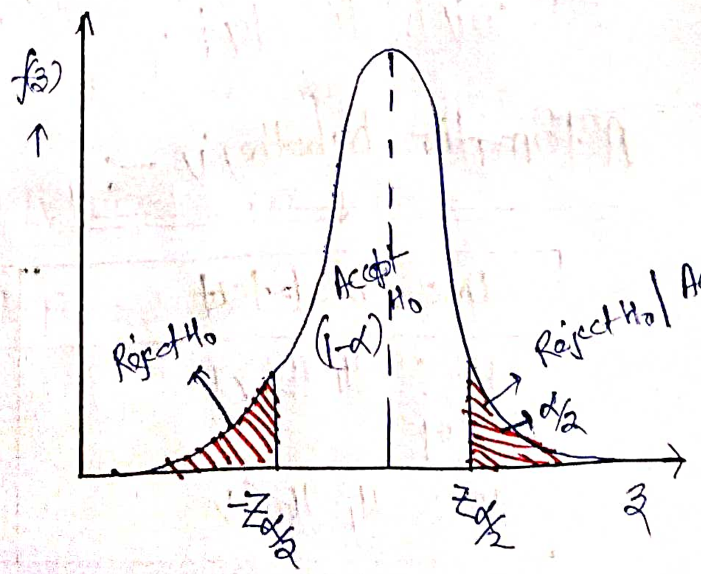
For one-tailed test	For two tailed test
$p\text{-value} = P(Z > z)$ or $P(Z < -z)$	$p\text{-value} = P(Z > z) + P(Z < -z)$

③ Rejection region –:

Here α is the level of significance

One tailed test	Two-tailed test
$Z > Z_\alpha$ or $Z < -Z_\alpha$	$Z > Z_{\alpha/2}$ or $Z < -Z_{\alpha/2}$





④ Conclusion -:

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

Question -: An investigation of relative merits of two kinds of FLASHLIGHT BATTERIES showed that a random sample of 100 batteries of brand A lasted on the average of 36.5 hours with a S.D. of 1.8 hours, while a random sample of 80 batteries of brand B lasted on the average of 36.8 hours with a S.D. of 1.5 hours. Use 5% level of significance ($Z_\alpha = 1.96$); test whether the observed difference between the average life times is significant.

Solution -: Given that

Sample	Size (n)	mean ($\bar{x}$)	S.D. (s)
A	100	36.5	1.8
B	80	36.8	1.5

(i) Define the Hypothesis -:

Null Hypothesis : $H_0 : \mu_1 - \mu_2 = 0$

Vs

Alternative hypothesis $H_1 : \mu_1 - \mu_2 \neq 0$ (Two tailed test)

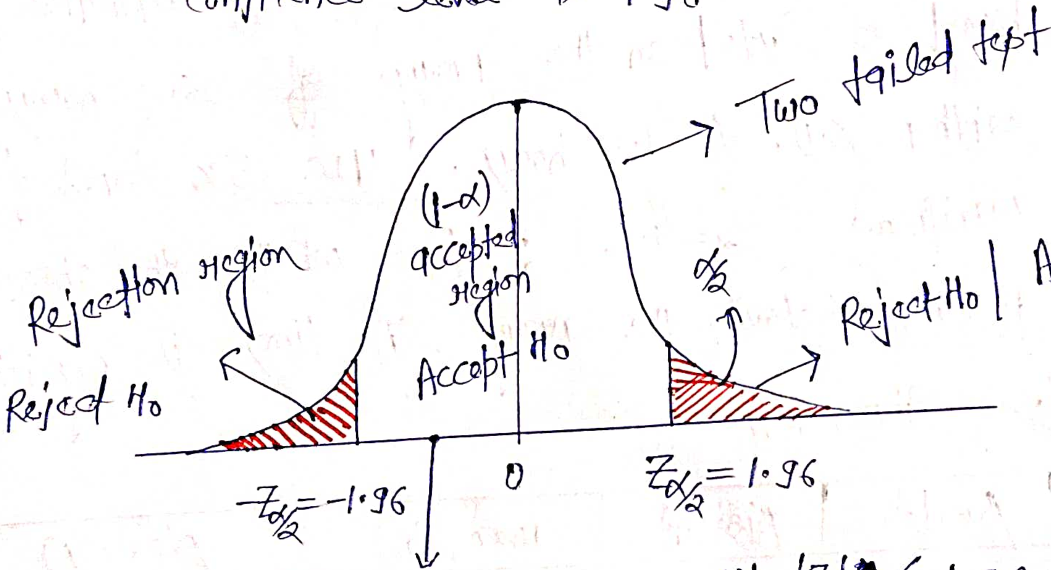
(2) Test - statistics -:

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{36.5 - 36.8}{\sqrt{\frac{(1.8)^2}{100} + \frac{(1.5)^2}{80}}}$$

$$\boxed{Z = -1.2195}$$

(3) Rejection region -:

In two tailed test, critical values at 95% Confidence level is 1.96



$z = -1.2195$ or $|Z| < 1.96$

(4) Conclusion -:

Since $-1.2195 > -1.96$, which fail to reject H_0. Hence we conclude that H_0 is accepted, i.e., there is no significant difference between the average life times of brand. A and B batteries.

Remark -: Revisit the previous example and suppose that an investigation will be concerned only if they detected a difference of more than 0.2 hours in the lives of two batteries. Based on your confidence interval 95%, Is the statistical significance in this case? Explain.

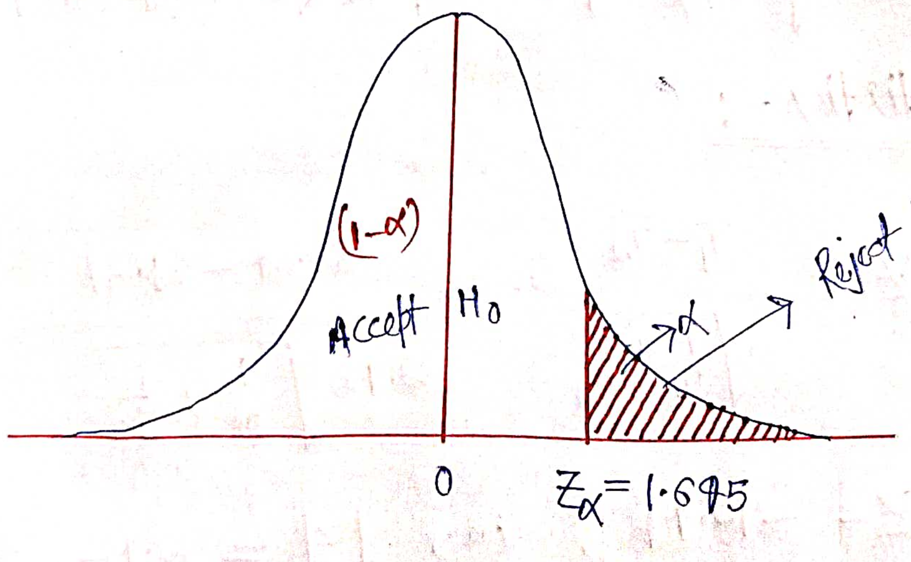
(i) $H_0 : \mu_1 - \mu_2 = 0.2$ (Null Hypothesis)

$H_1 : \mu_1 - \mu_2 > \underset{\Downarrow}{0.2}$ (Alternative hypothesis)
D

One tailed test

For 95% C.I. $\Rightarrow Z_\alpha = 1.645$

(ii) $$z = \frac{\bar{x}_1 - \bar{x}_2 - D}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ (do it).



Question :- If 60 B. Tech Computer science student are found to have mean height of 63.60 inches and 50 B. Tech mechanical students a mean height of 69.51 inches. Would you conclude that mechanical students are taller than Computer science students? Assume that the standard deviation of height of under-graduate students to be 2.48 inches. Use 5% level of significance ($Z_\alpha = 1.645$, one tailed test).

Solution :- Given that

$\bar{x}_1 = 63.60$ $\quad n_1 = 60 \rightarrow \mu_1$

$\bar{x}_2 = 69.51$ $\quad n_2 = 50 \rightarrow \mu_2$

large sample ($n \geq 30$).

$\sigma = 2.48$

(i) Define the Hypothesis :-

Null hypothesis H_0 : $\mu_1 - \mu_2 = 0$

Alternative hypothesis H_1 : $\mu_1 - \mu_2 < 0$ (One tailed test).

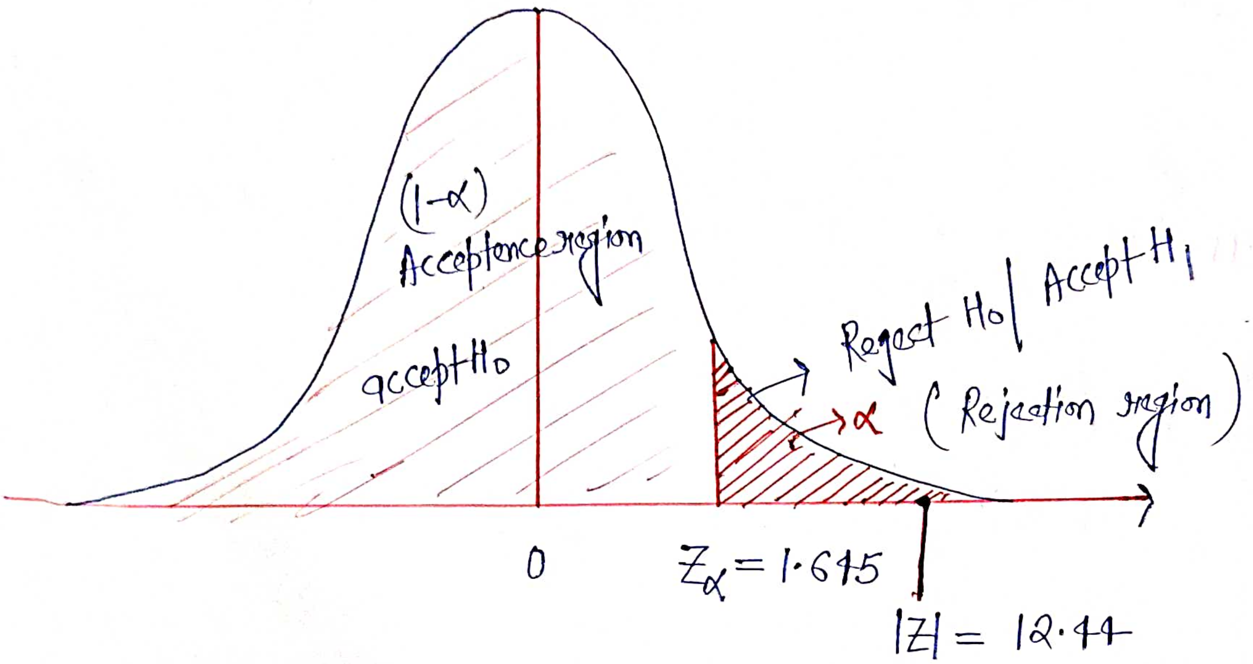
(ii) Test statistics :-

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{63.60 - 69.51}{2.48\sqrt{\frac{1}{60} + \frac{1}{50}}}$$

$Z = -12.44$, i.e., $|Z| = 12.44$.

⑥ **Rejection region -:**

For one tail, 95% confidence interval, the critical value $Z_\alpha = 1.645$.



⑦ **Conclusion -:** Since $|Z| > Z_\alpha = 1.645$, so H_0 is rejected, and we conclude that H_1 is accepted. That is yes, we can conclude that there is highly significant difference between their mean height, i.e., machenial students are taller than computer science students.

#