Student's t-test: Difference of samples mean

(Independent random sample)

Assumptions:

The sample must meet these requirements specified by the t-distribution:

(i) The sample must be randomly selected. (Independently)

(ii) The population from which you are sampling must be normally distributed.

(iii) The variances of the populations σ_1^2 and σ_2^2 are equal.

Standard error of the difference in two sample means:

The standard error of the difference in the two sample means is

$$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

and hence test statistics for small samples becomes.

$$\boxed{t = \frac{\bar{x}-\bar{y}-D}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}$$

Calculation of S^2:-

$$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2}{n_1+n_2-2} \quad \text{---(1)}$$

Also, we know that for unbiased estimator of X and Y, we have

$$s_1^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n_1-1} \Rightarrow \sum_{i=1}^{n}(x_i-\bar{x})^2 = (n_1-1)\,s_1^2$$

and

$$s_2^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n_2-1} \Rightarrow \sum_{i=1}^{n}(y_i-\bar{y})^2 = (n_2-1)\,s_2^2$$

Thus,

$$\boxed{S^2 = \frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{n_1+n_2-2}}$$

For Biased estimator of X and Y

$$s_1^2 = \frac{1}{n_1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad \& \quad s_2^2 = \frac{1}{n_2}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$\boxed{S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}}$$

Degree of freedom in two samples:

We know that, for one small sample consists of n elements, the degree of freedom is $(n-1)$. Thus, considering the two independent samples X & Y having n_1 and n_2 elements, we have.

For sample X –:

S_1^2 is unbiased sample with (n_1-1) degree of freedom.

For sample Y –:

S_2^2 is unbiased sample variance with (n_2-1) degree of freedom.

- So, total degree of freedom is

$$(n_1-1)+(n_2-1) = (n_1+n_2-2).$$

t-statistics test for two samples mean –:

(1) Define the hypothesis –:

Null hypothesis –: $H_0 : \mu_1-\mu_2 = D$, where D is some specified task that you wish to test.

Alternative hypothesis –:

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

② Test statistics :

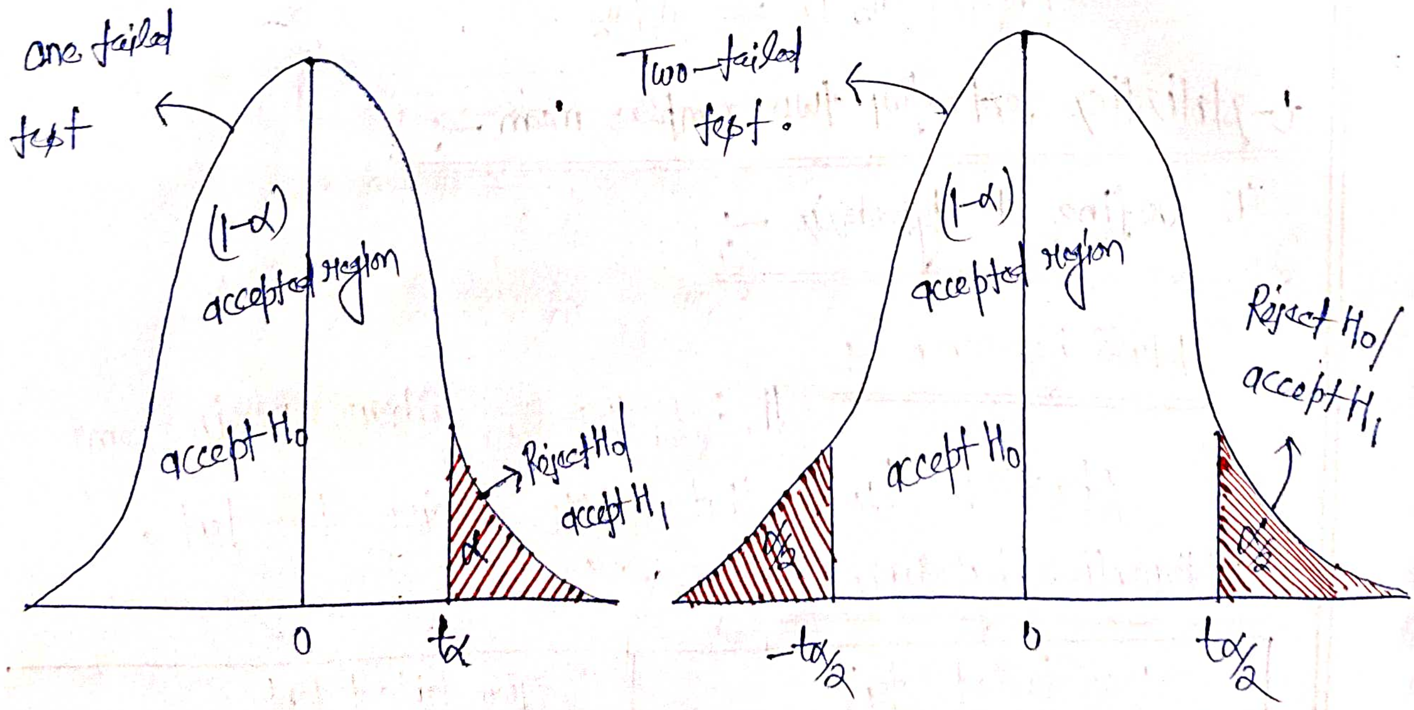
$$t = \frac{\bar{x} - \bar{y} - D}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where,

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

③ Rejection region : Let α is the level of significance.

One tailed test	Two tailed test
$t > t_\alpha$ or $t < t_\alpha$	$t > t_{\alpha/2}$ or $t < -t_{\alpha/2}$



or

p-value.

One tailed test	Two-tailed test
p value $= P(T>t)$ or $P(T<-t)$	p-value $= P(T>t) + P(T<-t)$.

The critical value of t, t_α $t_{\alpha/2}$ are based on (n_1+n_2-2) degree of freedom.

④ Conclusion —:

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

4 steps rule for testing

$\Downarrow$

Define the Hypothesis (Assumptions).

$\Downarrow$

Test-statistics calculated (Evidence)

$\Downarrow$

Define rejection region (Enough / Not)

$\Downarrow$

Conclusion based on critical values or p-value.

Question: A random sample of 20 daily workers of state A was found to have average daily earning of Rs. 44 with sample variance 900. Another sample of 20 daily workers from state B was found to earn on an average Rs. 30 per day with sample variance 400. Test whether the workers in state A are earning more than those in state B.

Solution: Given that

$$n_1 = 20, \quad \bar{x} = 44 \quad s_1^2 = 900$$
$$n_2 = 20 \quad \bar{y} = 30 \quad s_2^2 = 400$$

Biased variance.

State	size(n)	mean	S.D.
A	20	44	30
B	20	30	20

① Define the hypothesis:

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

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

② t-statistics:

$$t = \frac{\bar{x} - \bar{y} - 0}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2}{n_1 + n_2 - 2} = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$\boxed{S^2 = 618.21}$$

Thus,

$$t = \frac{44 - 30}{\sqrt{618.21}\sqrt{\frac{1}{20} + \frac{1}{20}}} = 1.7389$$

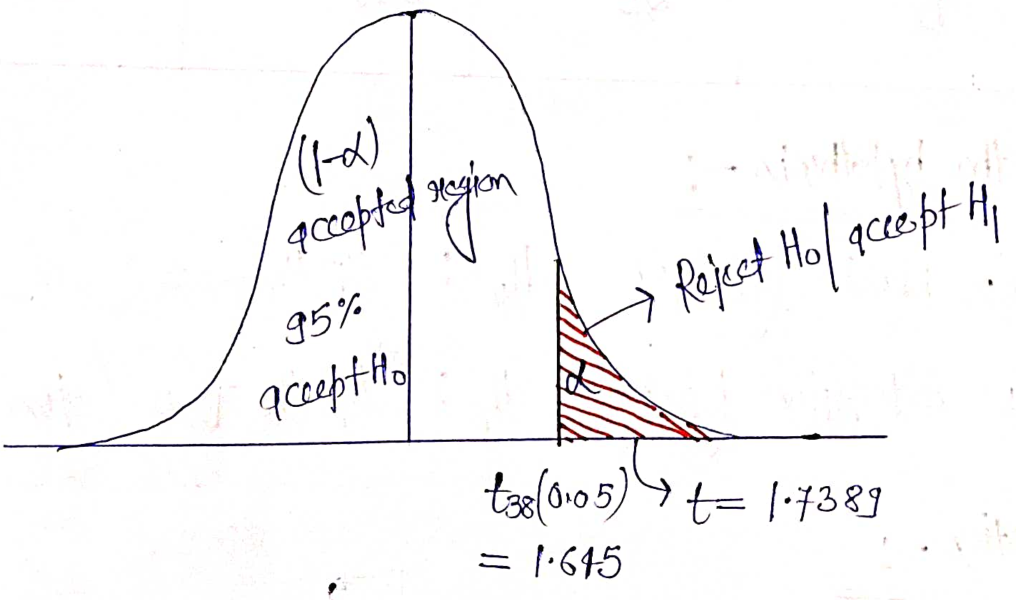
$$\boxed{t = 1.7389}$$

③ **Rejection region:-**

Degree of freedom $= n_1 + n_2 - 2 = 20 + 20 - 2 = 38$

Degree of freedom $= 38$

and $t_{38}(0.05) = 1.645$ (for one tailed)



④ **Conclusion:-** Since $1.7389 > 1.645$, so H_0 is rejected at 5% level of significance. Hence, we conclude that with 95% confident that the workers in state (A) are earning more than those in state (B).

\#

Question :- The average number of articles produced by two machines per day is 200 and 250 with standard deviations 20 and 25, respectively on the basis of record of 25 days production. Can you regard both the machine equally efficient at 1% level of significance.

Solution :- We have

Machines	size (n)	mean	S.D. (biased)
A	25	200	20
B	25	250	25

(i) Define the hypothesis :-

(a) Null-hypothesis $H_0 : \mu_1 - \mu_2 = 0$

(b) Alternative hypothesis $H_1 : \mu_1 - \mu_2 \neq 0$ (two tailed test).

(ii) t-statistics :-

$$t = \frac{\bar{x} - \bar{y} - D}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Where,

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2}{n_1 + n_2 - 2} = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$S^2 = 533.85$$

Thus,

$$t = \frac{200-250}{\sqrt{533.25\left(\frac{1}{25}+\frac{1}{25}\right)}} = -7.65$$

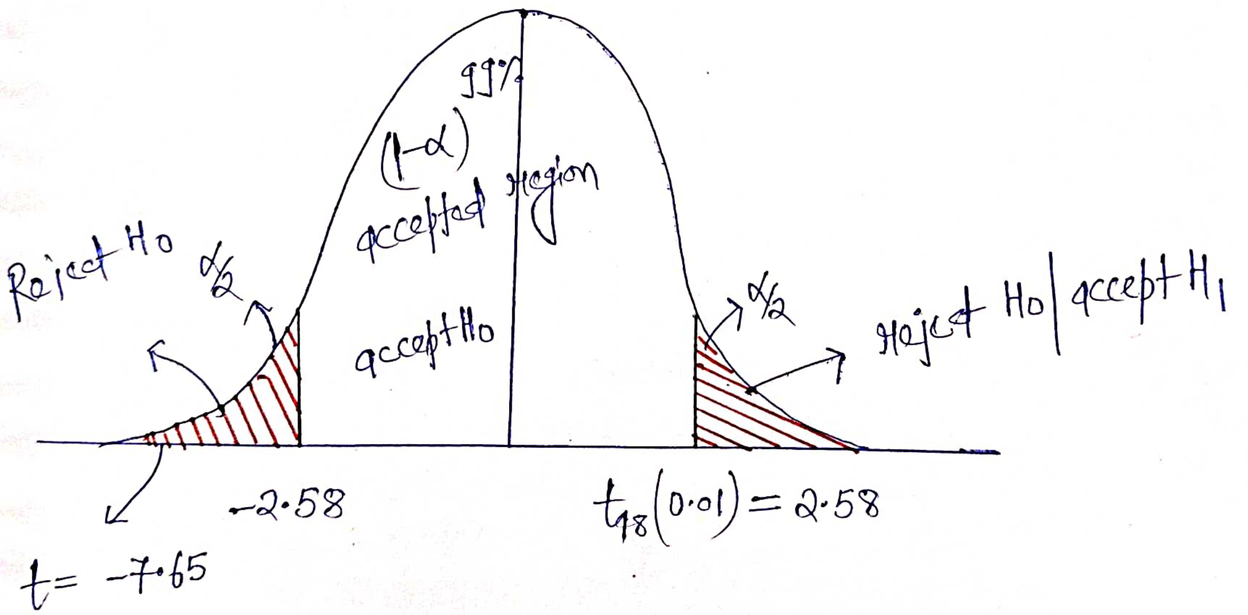
$$\boxed{t = -7.65}$$

③ Rejection region –:

Degree of freedom $= n_1 + n_2 - 2 = 25 + 25 - 2$

Degree of freedom $= 48$

and $t_{48}(0.01) = 2.58$ (for two tailed).



④ Conclusion –: Since $-7.65 < -2.58$, so H_0 is rejected at 1% level of significance. He we conclude that with 99% confident that the both the machines are not equaly efficiant.