Case 3: X has a general distribution, but we have a large sample size.

When population variance is known i.e.

σ^2 is known, the test statistics is

$$\frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

When population variance (σ^2) is unknown,

The test statistic is, $\frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} \sim N(0,1)$

Where, $\bar{x} =$ sample mean

$$S^2 = \text{sample variance} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}, n = \text{sample size}$$

$H_1: \mu > \mu_0$

The rejection region is $[\, Z_\alpha, +\infty[$

$H_1: \mu < \mu_0$

The rejection region is $]-\infty, -Z_\alpha]$

When $H_1: \mu \neq \mu_0$

The rejection region is $]-\infty, -Z_{\frac{\alpha}{2}}] \cup [\, Z_{\frac{\alpha}{2}}, +\infty[$

Example: From a long-term experience, a factory owner knows that a worker can produce a product in an average time of 89 min. However, on Monday morning, there is the impression **that takes longer.** To test whether this impression is correct a sample (n=50) is taken with sample mean $\bar{x} = 92.2$ and sample variance $s^2 = 144$. Verify whether this impression is correct at 5% level of significance.

Solution:

$$H_0: \mu = 89$$

$$H_1: \mu > 89$$

Test statistic is $\frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} \sim N(0,1)$

$$= \frac{92.2-89}{\frac{\sqrt{144}}{\sqrt{50}}}$$

$$= 1.88$$

The rejection region is $[\, Z_{\alpha}, +\infty[$

$$= [\, Z_{0.05}, +\infty[$$

$$= [\, 1.645, +\infty[$$ [From table page- 787]

Comment: Since test statistic's value (1.88) falls in the rejection region, so we reject H_0 at 5% level of significance.

That is the factory owner impression is correct.

Matched Pairs t test: Matched – Pairs t – test is used to test whether there is a significant mean difference between two sets of **paired** data.

For example: Suppose we are interested in evaluating the effectiveness of a company training program. To evaluate this training effectiveness, measure the performance of the employees, before and after completing the program and analyze the differences by using matched pair t test.

Step 1: $H_0 : \mu_D = 0$

$H_1 : \mu_D > 0$

Or $H_1 : \mu_D < 0$

Or $H_! : \mu_D \neq 0$

Here, $\mu_D = \mu_y - \mu_x$

Step 2: Test statistic = $\frac{\bar{D}}{\sqrt{\frac{S^2{}_D}{n}}} \sim t_{n-1}$

Here, D is the difference between paired value from two data sets.

$D = Y - X$

Where X is the variable in the first data set.

Y is the variable in the second data set.

$\overline{D}$ is the sample mean difference between paired observation.

${s^2}_D$ is the variance of the differences.

n is the number of paired observation.

Step 3: Rejection region

If $H_1 : \mu_D > 0$

Then rejection region is $[\, t_{\alpha,n-1}\, , +\infty[$

If $H_1 : \mu_D < 0$

Then rejection region is $]-\infty, -\, t_{\alpha,n-1}\,]$

If $H_! : \mu_D \neq 0$

Then rejection region is $\left]-\infty, -t_{\frac{\alpha}{2},n-1}\right]$ U $\left[t_{\frac{\alpha}{2},n-1}, +\infty\right[$

Step 4: Comment.

Example:

In 10 women the systolic blood pressure (mm Hg) is measured at the beginning of a clinical trial. Afterwards they have a fertility treatment with hormones. During this treatment they are again measured.

Id	before	during	Id	before	during
1	115	128	6	138	145
2	112	115	7	126	132
3	107	106	8	105	109
4	119	128	9	104	102
5	115	122	10	115	117

Does the fertility treatment with hormones have a significant effect on the systolic blood pressure. Test this at 5% level of significance.

Solution: $H_0 : \mu_D = 0$

$H_1 : \mu_D \neq 0$

Where $\mu_D = \mu_y - \mu_x$

Note: if we take $\mu_D = \mu_x - \mu_y$ conclusion will be same when $H_1 : \mu_D \neq 0$

μ_x indicate average blood pressure of population before fertility treatment with hormones.

μ_y indicate average blood pressure of population during fertility treatment with hormones.

Test statistic = $\frac{\bar{D}}{\sqrt{\frac{s^2{}_D}{n}}} \sim t_{n-1}$

From the data, we get

Id (i)	Di = Yi - Xi
1	13
2	3
3	-1
4	9
5	7
6	7
7	6
8	4
9	-2
10	2

$\therefore \bar{D} = \frac{13+3+\cdots\ldots+2}{10} = 4.8$

$s^2{}_D = \frac{\sum_{i=1}^{10}(Di-\bar{D})^2}{n-1}$

$= \frac{(D_1-\bar{D})^2+(D_2-\bar{D})^2+\cdots+(D_{10}-\bar{D})^2}{10-1}$

$= \frac{(13-4.8)^2+(3-4.8)^2+\cdots+(2-4.8)^2}{9}$

= 20.844

$\therefore$ Test statistic = $\frac{4.8}{\sqrt{\frac{20.844}{10}}}$

= 3.3247

Rejection region is: $\left]-\infty, -t_{\frac{\alpha}{2},n-1}\right] \cup \left[t_{\frac{\alpha}{2},n-1}, +\infty\right[$

$=]-\infty, -2.262] \cup [2.262, +\infty[$

Comment: Since the test statistics value falls in the rejection region, so we reject null hypothesis. i.e The fertility treatment with hormones have a significant effect on the systolic blood pressure.

Example:

In 10 women the systolic blood pressure (mm Hg) is measured at the beginning of a clinical trial. Afterwards they have a fertility treatment with hormones. During this treatment they are again measured.

Id	before	during	Id	before	during
1	115	128	6	138	145
2	112	115	7	126	132
3	107	106	8	105	109
4	119	128	9	104	102
5	115	122	10	115	117

Does the fertility treatment with hormones increase the systolic blood pressure? Test this at 5% level of significance.

Solution: $H_0 : \mu_D = 0$

$H_1 : \mu_D > 0$

Where $\mu_D = \mu_y - \mu_x$

μ_x indicate average blood pressure of population before fertility treatment with hormones.

μ_y indicate average blood pressure of population during fertility treatment with hormones.

Test statistic = $\frac{\bar{D}}{\sqrt{\frac{S^2{}_D}{n}}} \sim t_{n-1}$

From the data, we get

Id (i)	Di = Yi - Xi
1	13
2	3
3	-1
4	9
5	7
6	7
7	6
8	4
9	-2
10	2

$\therefore \bar{D} = \frac{13+3+\cdots\ldots+2}{10} = 4.8$

$s^2{}_D = \frac{\sum_{i=1}^{10}(Di-\bar{D})^2}{n-1}$

$= \frac{(D_1-\bar{D})^2+(D_2-\bar{D})^2+\cdots+(D_{10}-\bar{D})^2}{10-1}$

$= \frac{(13-4.8)^2+(3-4.8)^2+\cdots+(2-4.8)^2}{9}$

= 20.844

$\therefore$ Test statistic = $\frac{4.8}{\sqrt{\frac{20.844}{10}}}$

= 3.3247

Rejection region is: $[t_{\alpha,n-1}, +\infty[\;\; = [t_{0.05\,,10-1}, +\infty[$

$= [1.833, +\infty[$

Comment: Since the test statistics value falls in the rejection region, so we reject null hypothesis. i.e The fertility treatment with hormones increase the systolic blood pressure.

Example:

In 10 women the systolic blood pressure (mm Hg) is measured at the beginning of a clinical trial. Afterwards they have a fertility treatment with hormones. During this treatment they are again measured.

Id	before	during	Id	before	during
1	115	128	6	138	145
2	112	115	7	126	132
3	107	106	8	105	109
4	119	128	9	104	102
5	115	122	10	115	117

Does the fertility treatment with hormones decrease the systolic blood pressure? Test this at 5% level of significance.

Solution: $H_0 : \mu_D = 0$

$H_1 : \mu_D < 0$

Where $\mu_D = \mu_y - \mu_x$

μ_x indicate average blood pressure of population before fertility treatment with hormones.

μ_y indicate average blood pressure of population during fertility treatment with hormones.

Test statistic = $\frac{\bar{D}}{\sqrt{\frac{S^2{}_D}{n}}} \sim t_{n-1}$

From the data, we get

Id (i)	Di = Yi - Xi
1	13
2	3
3	-1
4	9
5	7

6	7
7	6
8	4
9	-2
10	2

$\therefore \bar{D} = \frac{13+3+\cdots....+2}{10} = 4.8$

$s^2{}_D = \frac{\sum_{i=1}^{10}(Di-\bar{D})^2}{n-1}$

$= \frac{(D_1-\bar{D})^2+(D_2-\bar{D})^2+\cdots+(D_{10}-\bar{D})^2}{10-1}$

$= \frac{(13-4.8)^2+(3-4.8)^2+\cdots+(2-4.8)^2}{9}$

= 20.844

$\therefore$ Test statistic = $\frac{4.8}{\sqrt{\frac{20.844}{10}}}$

= 3.3247

Rejection region is: : $]-\infty,-t_{\alpha,n-1}]$ =: $]-\infty,-t_{0.05,10-1}]$

= $]-\infty,-1.833]$

Comment: Since the test statistics value does not fall in the rejection region, so we can not reject null hypothesis. i.e the fertility treatment with hormones does not decrease the systolic blood pressure.

■ **Independent sample t test:** also called the unpaired sample t test helps us to compare the means of two sets of data.

■ **Example:** In an experiment, we compare the result of treatment A and treatment B by seeing the survival time mouses.

Treatment A: 17 19 15 18 21 18

Treatment B: 18 15 13 16 13

■ **Step 1:** $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$

or $H_1 : \mu_1 > \mu_2$

or $H_1 : \mu_1 \neq \mu_2$

$\Downarrow$

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 < 0$

or $H_1 : \mu_1 - \mu_2 > 0$

or $H_1 : \mu_1 - \mu_2 \neq 0$

■ **Step 2:** Test statistic = $\frac{\bar{X}-\bar{Y}}{\sqrt{s^2{}_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$

Where $s^2{}_p = \frac{\sum_{i=1}^{n_1}(x_i-\bar{x})^2+\sum_{j=1}^{n_2}(y_j-\bar{y})^2}{n_1+n_2-2}$

Step 3: If $H_1 : \mu_1 - \mu_2 < 0$

The rejection region: $]-\infty, -t_\alpha]$

If $H_1 : \mu_1 - \mu_2 > 0$

The rejection region: $[t_\alpha, +\infty[$

If $H_1 : \mu_1 - \mu_2 \neq 0$

The rejection region: $\left]-\infty, -t_{\frac{\alpha}{2}}\right]$ U $\left[t_{\frac{\alpha}{2}}, +\infty\right[$

Step 4: Comment.

■ **Question:** In an experiment, we compare the result of treatment A and treatment B by seeing the survival time of mouses.

Treatment A: 17 19 15 18 21 18

Treatment B: 18 15 13 16 13

Investigate whether treatment A gives better result? Test this at 5% level of sig.

Solution: $H_0 : \mu_1 = \mu_2 \quad \Rightarrow \quad H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 > \mu_2 \quad \Rightarrow H_1 : \mu_1 - \mu_2 > 0$

Test statistic = $\dfrac{\bar{X}-\bar{Y}}{\sqrt{s^2{}_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Here $s^2{}_p = \dfrac{\sum_{i=1}^{6}(x_i-\bar{x})^2+\sum_{j=1}^{5}\left(y_j-\bar{y}\right)^2}{n_1+n_2-2}$

And $\bar{x} = \dfrac{17+\cdots+18}{6} = 18$

$\bar{y} = \dfrac{18+\cdots+13}{5} = 15$

$s^2{}_p = \dfrac{(17-18)^2+(19-18)^2+\cdots+(18-18)^2+(18-15)^2+(15-15)^2\ldots+(13-15)^2}{6+5-2}$

= 4.22

∴ Test statistic = $\dfrac{18-15}{\sqrt{4.22\left(\frac{1}{6}+\frac{1}{5}\right)}}$

= 2.41

Rejection region: $[\, t_\alpha\, , +\infty[$

= $[1.833, +\infty[$

Comment: Since the calculate value falls in the rejection region, we reject H_0 . That is, treatment A gives better result.

■ **Question:** In an experiment, we compare the result of treatment A and treatment B by seeing the survival time of mouses.

Treatment A: 17 19 15 18 21 18

Treatment B: 18 15 13 16 13

Investigate whether treatment B gives better result? Test this at 5% level of sig.

Solution: $H_0 : \mu_1 = \mu_2 \;\; \Rightarrow \;\; H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 < \mu_2 \;\; \Rightarrow H_1 : \mu_1 - \mu_2 < 0$

Test statistic = $\dfrac{\bar{X}-\bar{Y}}{\sqrt{s^2{}_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Here $s^2{}_p = \dfrac{\sum_{i=1}^{6}(x_i-\bar{x})^2+\sum_{j=1}^{5}\left(y_j-\bar{y}\right)^2}{n_1+n_2-2}$

And $\bar{x} = \dfrac{17+\cdots+18}{6} = 18$

$\bar{y} = \dfrac{18+\cdots+13}{5} = 15$

$s^2{}_p = \dfrac{(17-18)^2+(19-18)^2+\cdots+(18-18)^2+(18-15)^2+(15-15)^2 \ldots+(13-15)^2}{6+5-2}$

= 4.22

∴ Test statistic = $\dfrac{18-15}{\sqrt{4.22\left(\frac{1}{6}+\frac{1}{5}\right)}}$

= 2.41

Rejection region: $]-\infty, -t_\alpha]$

$=]-\infty, -1.833]$

Comment: Since the calculate value does not fall in the rejection region, so we can not reject H_0 . That is, treatment B does not give better result.

Question: In an experiment, we compare the result of treatment A and treatment B by seeing the survival time of mouses.

Treatment A: 17 19 15 18 21 18

Treatment B: 18 15 13 16 13

Investigate whether treatment A and treatment B gives different result? Test this at 5% level of sig.

Solution: $H_0 : \mu_1 = \mu_2 \;\; \Rightarrow \;\; H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 \neq \mu_2 \;\; \Rightarrow H_1 : \mu_1 - \mu_2 \neq 0$

Test statistic = $\frac{\bar{X}-\bar{Y}}{\sqrt{s^2{}_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Here $s^2{}_p = \frac{\sum_{i=1}^{6}(x_i-\bar{x})^2+\sum_{j=1}^{5}\left(y_j-\bar{y}\right)^2}{n_1+n_2-2}$

And $\bar{x} = \frac{17+\cdots+18}{6} = 18$

$\bar{y} = \frac{18+\cdots+13}{5} = 15$

$s^2{}_p = \frac{(17-18)^2+(19-18)^2+\cdots+(18-18)^2+(18-15)^2+(15-15)^2 \ldots+(13-15)^2}{6+5-2}$

= 4.22

$\therefore$ Test statistic = $\frac{18-15}{\sqrt{4.22\left(\frac{1}{6}+\frac{1}{5}\right)}}$

= 2.41

Rejection region: $]-\infty, -t_{\frac{\alpha}{2}}]$ U $\left[t_{\frac{\alpha}{2}}, +\infty\right[$

$=]-\infty, -2.262]$ U $[2.262, +\infty[$

Comment: Since the calculate value falls in the rejection region, so we reject H_0 . That is, treatment A and treatment B give different result.