3.2.5 Paired sample t-test or Paired t-test or Matched pair of comparison test:

In the application of the two-sample t-test (two means, small/large samples), the samples are independent. But this test cannot be used when we deal with 'before and after' kind of data. Here the data is in pairs like a data related to 'before and after a training program', 'before and after taking a particular medicine', 'before and after listened a class' etc. In this case, we work with the differences of the paired data and these differences looked upon as a sample from the population of these differences. If this sample is small, we use the one sample t-test (one mean small sample test); otherwise, we use the corresponding large sample test (one mean large sample test). The one sample t-test used for this kind of paired data is called paired sample t-test.

(1) The blood pressures of 5 women before and after intake of a certain drug are given below.

Before	110	120	125	132	125
After	120	118	125	136	121

Test whether there is a significant change in blood pressure at the level of significance 0.01

(2) The following are the average weekly loss of working hours due to accidents in 10 industrial plants before and after a certain safety program was put into operation: 45 and 36, 73 and 60, 46 and 44, 124 and 119, 33 and 35, 57 and 51, 83 and 77, 34 and 29, 26 and 24, 17 and 11. Use the level of significance 0.05 test whether the safety program is effective. Also find the 90% confidence interval for the mean improvement in loss of working hours.

(3) In a study of the effectiveness of physical exercise in weight reduction program, a group of 11 persons engaged in a prescribed program of physical exercise for a month showed the following results.

Weight before	209	178	169	212	180	192	159	180	170	153	183
Weight after	196	171	160	207	177	190	128	196	164	152	179

Use the level of significance 0.01 test whether the prescribed program is effective

(4) Memory capacity of 10 students were tested before and after a training program. Test whether the training program was effective or not from the following scores.

Before Training	12	14	11	8	7	10	3	0	5	6
After Training	15	16	10	7	5	12	10	2	3	8

Units – 4: Estimation and Test of Hypothesis of Variances and Proportions

4.1 Estimation and Test of Hypothesis of Proportions

4.1.1 Proportion:

Let p, q be the success and failure probabilities of an event of a trial. Let the trial be conducted in any number of times. Then the collection of all successes and failures of the event is a population known as Binomial population. For this population, p is called **Proportion or True proportion**. If x is the number of successes in n trials, then $\frac{x}{n}$ is called the **Sample Proportion** and it is denoted by P ; that is, $P = \frac{x}{n}$

Example:

Let a coin be tossed 10 times and 'getting head H ' be the event. Suppose that the outcomes in these 10 tosses are $H, H, H, T, T, H, T, H, H, T$ respectively. Now the collection of all the successes and failures of the event is a population; that is, $Population = \{S, S, S, F, F, S, F, S, S, F\}$.

For this population, **Proportion** $p = \frac{1}{2}$

(i) If we collect 1st five outcomes $\{H, H, H, F, F\}$, then it is a sample of size $n = 5$

For this sample, sample proportion $P = \frac{x}{n} = \frac{3}{5}$

(ii) If we collect 1st two outcomes $\{H, H\}$, then it is a sample of size $n = 2$

For this sample, sample proportion $P = \frac{x}{n} = \frac{2}{2} = 1$

(iii) If we collect 1st four outcomes $\{H, H, H, F\}$, then it is a sample of size $n = 4$

For this sample, sample proportion $P = \frac{x}{n} = \frac{3}{4}$

(iv) If we collect last two outcomes $\{H, F\}$, then it is a sample of size $n = 2$

For this sample, sample proportion $P = \frac{x}{n} = \frac{1}{2}$

4.1.2 Maximum Error and Confidence Interval - One Proportion

Maximum Error of the Proportion:

(i) $E = z_{\frac{\alpha}{2}} \times \sqrt{\dfrac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$ is the maximum error of the Proportion with the probability $1-\alpha$

(ii) If we know the value of p , then the sample size is given by $n = p\,(1-p)\left[\dfrac{z_{\frac{\alpha}{2}}}{E}\right]^2$

(iii) If we do not know the value of p , then the sample size is given by $n = \dfrac{1}{4}\left[\dfrac{z_{\frac{\alpha}{2}}}{E}\right]^2$

Confidence Interval – One Proportion:

With the probability $1-\alpha$ or $(1-\alpha)100\%$ confidence,

Upper and Lower confidence limits: $\dfrac{x}{n} \pm E$

Confidence Interval: $\left(\dfrac{x}{n} - E,\ \dfrac{x}{n} + E\right)$ where, $E = z_{\frac{\alpha}{2}} \times \sqrt{\dfrac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$

4.1.3 Test of Hypothesis – One Proportion:

(1) Null Hypothesis H_0 : $p = p_0$

(2) Alternative Hypothesis H_1 : Any one of the following $p < p_0$, $p > p_0$, $p \neq p_0$

(3) Level of Significance : α

(4) Test statistic : $Z = \dfrac{x - n\,p}{\sqrt{n\,p\,(1-p)}}$ or $Z = \dfrac{\frac{x}{n} - p}{\sqrt{\frac{p\,(1-p)}{n}}}$

(5) Criterion :

H_1	Reject H_0 if
$p < p_0$	$Z < -z_\alpha$
$p > p_0$	$Z > z_\alpha$
$p \neq p_0$	$\lvert Z \rvert > z_{\frac{\alpha}{2}}$

4.1.4 Test of Hypothesis – Confidence Interval - Two Proportions

Test of Hypothesis – Two Proportions:

(1) Null Hypothesis H_0 : $p_1 - p_2 = 0$

(2) Alternative Hypothesis H_1 : Any one of the following $p_1 - p_2 < 0$, $p_1 - p_2 > 0$, $p_1 - p_2 \neq 0$

(3) Level of Significance : α

(4) Test statistic : $Z = \dfrac{\dfrac{x_1}{n_1} - \dfrac{x_2}{n_2}}{\sqrt{\dfrac{\hat{p}\,(1-\hat{p})}{n_1} + \dfrac{\hat{p}\,(1-\hat{p})}{n_2}}}$ where, $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

(5) Criterion :

H_1	Reject H_0 if
$p_1 - p_2 < 0$	$Z < -z_\alpha$
$p_1 - p_2 > 0$	$Z > z_\alpha$
$p_1 - p_2 \neq 0$	$\lvert Z \rvert > z_{\frac{\alpha}{2}}$

Note: $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ is known as the **proportion by pooling**

Confidence Interval - Two Proportions:

With the probability $1-\alpha$ or $(1-\alpha)100\%$ confidence,

Upper and Lower confidence limits: $\left(\dfrac{x_1}{n_1} - \dfrac{x_2}{n_2}\right) \pm E$

Confidence Interval: $\left(\left(\dfrac{x_1}{n_1} - \dfrac{x_2}{n_2}\right) - E,\ \left(\dfrac{x_1}{n_1} - \dfrac{x_2}{n_2}\right) + E\right)$

Where, $E = z_{\frac{\alpha}{2}} \times \sqrt{\dfrac{\hat{p}\,(1-\hat{p})}{n_1} + \dfrac{\hat{p}\,(1-\hat{p})}{n_2}}$ and $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

4.1.5 Test of Hypothesis – Several Proportions:

Let $n_1, n_2, n_3, \cdots n_k$ be the sizes of k number of samples taken from k number of populations respectively. Let $x_1, x_2, x_3, \cdots x_k$ be the numbers of successes of these k samples respectively.
Let $n = n_1 + n_2 + n_3 + \cdots + n_k$ and $x = x_1 + x_2 + x_3 + \cdots + x_k$.
Now all these values can be tabulated as follows

	Sample 1	Sample 2	$\cdots$	Sample k	Total
No. of successes	x_1	x_2	$\cdots$	x_k	x
No. of failures	$n_1 - x_1$	$n_2 - x_2$	$\cdots$	$n_k - x_k$	$n - x$
Total	n_1	n_2	$\cdots$	n_k	n

In the above table, each entry in (i, j)th cell is called Observed frequency and it is denoted by O_{ij} for $i = 1, 2$ and $j = 1, 2, 3, \cdots k$; that is,

$$O_{11} = x_1,\ O_{12} = x_2,\ \cdots\ O_{1k} = x_k,$$
$$O_{21} = n_1 - x_1,\ O_{22} = n_2 - x_2,\ \cdots\ O_{2k} = n_k - x_k$$

And the Expected frequency of each (i, j)th cell is given by

$$e_{ij} = \frac{\left(i^{\text{th}} \text{ row total}\right) \times \left(j^{\text{th}} \text{ column total}\right)}{n} \quad \text{for } i = 1, 2 \text{ and } j = 1, 2, 3, \cdots k; \text{ that is,}$$

$$e_{11} = \frac{x\,n_1}{n},\ e_{12} = \frac{x\,n_2}{n},\ \cdots\ e_{1k} = \frac{x\,n_k}{n},$$
$$e_{21} = \frac{(n-x)n_1}{n},\ e_{22} = \frac{(n-x)n_2}{n},\ \cdots\ e_{2k} = \frac{(n-x)n_k}{n}$$

(1) Null Hypothesis H_0 **:** $p_1 = p_2 = p_3 = \cdots = p_k$

(2) Alternative Hypothesis H_1 **:** Not all $p_1, p_2, p_3, \cdots p_k$ are equal

(3) Level of Significance : α

(4) Test statistic : $\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$ with $\nu = k - 1$

(5) Criterion : Reject H_0 if $\chi^2 > \chi^2_\alpha$

4.2 Estimation and Test of Hypothesis of Variances

4.2.1 Confidence Interval - One Variance

With the probability $1-\alpha$ or $(1-\alpha)100\%$ confidence,

Lower and Upper confidence limits are respectively given by $\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}$ and $\dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}$

Confidence Interval is given by $\left(\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}, \dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}\right)$

4.2.2 Test of Hypothesis – One Variance

(1) Null Hypothesis H_0 : $\sigma^2 = \sigma_0^2$

(2) Alternative Hypothesis H_1 : Any one of the following $\sigma^2 < \sigma_0^2$, $\sigma^2 > \sigma_0^2$, $\sigma^2 \neq \sigma_0^2$

(3) Level of Significance : α

(4) Test statistic : $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$, $\nu = n-1$

(5) Criterion :

H_1	Reject H_0 if
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{1-\alpha}$
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{\alpha}$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{1-\frac{\alpha}{2}}$ or $\chi^2 > \chi^2_{\frac{\alpha}{2}}$

4.2.3 Test of Hypothesis – Two Variances

(1) Null Hypothesis H_0 : $\sigma_1^2 = \sigma_2^2$

(2) Alternative Hypothesis H_1 : Any one of the following $\sigma_1^2 < \sigma_2^2$, $\sigma_1^2 > \sigma_2^2$, $\sigma_1^2 \neq \sigma_2^2$

(3) Level of Significance : α

(4) Test statistic & Criterion :

H_1	Test Statistic	Reject H_0 if
$\sigma_1^2 < \sigma_2^2$	$F = \dfrac{S_2^2}{S_1^2}$	$F > F_{\alpha}(n_2-1, n_1-1)$
$\sigma_1^2 > \sigma_2^2$	$F = \dfrac{S_1^2}{S_2^2}$	$F > F_{\alpha}(n_1-1, n_2-1)$
$\sigma_1^2 \neq \sigma_2^2$	If $\sigma_1^2 < \sigma_2^2$ then $F = \dfrac{S_2^2}{S_1^2}$ If $\sigma_1^2 > \sigma_2^2$ then $F = \dfrac{S_1^2}{S_2^2}$	$F > F_{\frac{\alpha}{2}}(n_2-1, n_1-1)$ $F > F_{\frac{\alpha}{2}}(n_1-1, n_2-1)$