UNIT-IV TESTING OF HYPOTHESIS

Population:

A population in statistics means a set of object. The population is finite or infinite according to the number of elements of the set is finites or infinite.

Sampling:

A sample is a finite subset of the population. The number of elements in the sample is called size of the sample.

Large and small sample:

The number of elements in a sample is greater than or equal to 30 then the sample is called a large sample and if it is less than 30, then the sample is called a small sample.

Parameters:

Statistical constant like mean μ, variance σ^2 , etc., computed from a population are called parameters of the population.

Statistics:

Statistical constants like $\bar{x}$, variance S^2, etc., computed from a sample are called samlple staticts or statistics.

POPULATION (PARAMETER)	SAMPLE (STATISTICS)
Population size=N	Sample size=n
Population mean= μ	Sample mean= x
Population s.d.=σ	Sample s.d.=S
Population proportion= P	Sample proportion= p

<u>Tests of significance or Hypothesis Testing:</u>

Statistical Hypothesis:

In making statistical decision, we make assumption, which may be true or false are called Statistical Hypothesis.

Null Hypothesis(H_0):

For applying the test of significance, we first setup a hypothesis which is a statement about the population parameter. This statement is usually a hypothesis of no true difference between sample statistics and population parameter under consideration and so it is called null hypothesis and is denoted by H_0.

Alternative Hypothesis (H_1):

Suppose the null hypothesis is false, then something else must be true. This is called an alternative hypothesis and is denoted by H_1.

Eg. If H_0 is population mean μ=300, then H_1 is $\mu \neq 300$ $(ie. \mu < 300 \; or \; \mu > 300)$ or $H_1 \; is \; \mu > 300 \; or \; H_1 \; is \; \mu < 300$. So any of these may be taken as alternative hypothesis.

<u>Error in sampling:</u>

After applying a test of significance a decision is to be taken to accept or reject the null hypothesis H_0.

Type I error: The rejection of the null hypothesis H_0 when it is true is called type I error.

Type II error: The acceptance of the null hypothesis H_0 when it is false is called type II error.

Level of significance:

The probability of type I error is called level of significance of the test and it is denoted by α. We usually take either α=5% or α=1%.

One tailed and Two tailed test:

If θ_0 is a population parameter and θ is the corresponding sample statistics and if we setup the null hypothesis $H_0 : \theta = \theta_0$, then the alternative hypothesis which is complementary to H_0 can be anyone of the following:

(i) $H_1 : \theta \neq \theta_0$ $(\theta < \theta_0 \;\; or \; \theta > \theta_0)$ (ii) $H_1 : \theta < \theta_0$ (iii) $H_1 : \theta > \theta_0$

Alternative hypotheis, whereas H_1 given in (ii) is called a left-tailed test. And (iii) is called a right tailed test.

Level of significance:

The probability of Type I error is called the level of significance of the test and is denoted by α .

Critical region:

For a test statistic, the area under the probability curve, which is normal is divided into two region namely the region of acceptance of H_0 and the region of rejection of H_0. The region in which H_0 is rejected is called critical region. The region in which H_0 is accepted is called acceptance region.

Procedure of Testing of Hypothesis:

(i) State the null hypothesis H_0

(ii) Decide the alternative hypothesis H_1 (ie, one tailed or two tailed)

(iii) Choose the level of significance α (α=5% or α=1%).

(iv) Determine a suitable test statistic.

$$\text{Test statistic} = \frac{t - E(t)}{S.E \; of \; (t)}$$

(v) Compute the computed value of $|z|$ with the table value of z and decide the acceptane or the rejection of H_0.

If $|z|$<1.96, H_0 may be accepted at 5% level of significance. If $|z|$>1.96, H_0 may be rejection at 5% level of significance.

If $|z|$<2.58, H_0 may be accepted at 1% level of significance. If $|z|$>2.58, H_0 may be rejection at 1% level of significance.

For a single tail test(right tail or left tail) we compare the computed value of $|z|$ with 1.645(at 5% level) and 2.33(at 1% level) and accept or reject H_0 accordingly.

Test of significance of small sample:

When the size of the sample (n) is less than 30, then that sample is called a small sample. The following are some important tests for small sample,

(I) students t test

(II) F-test

(III) χ^2-test

I **<u>Student t test</u>**

(i). Test of significance of the difference between sample mean and population mean

(ii). Test of significance of the difference between means of two small samples

(i) **Test of significance of the difference between sample mean and population mean:**

The studemts 't' is defined by the statistic $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ where $\bar{x}$ =sample mean, μ=population mean, S=standard deviation of sample,

n= sample size.

<u>Note:</u>

If standard deviation of sample is not given directly then, the static is given by $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$, where

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

Confident Interval:

The confident interval for the population mean for small sample is $\bar{x} \mp t_\alpha \frac{s}{\sqrt{n}}$

$$\Rightarrow \left(\bar{x} - t_\alpha \frac{s}{\sqrt{n}}, \bar{x} + t_\alpha \frac{s}{\sqrt{n}} \right)$$

<u>Working Rule:</u>

(i) Let $H_0 : \mu = \bar{x}$ (there is no significant difference between sample mean and population mean)

$H_1 : \mu \neq \bar{x}$ (there is no significant difference between sample mean and population mean)(Two tailed test)

Find $t = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}$.

Let t_α be the table value of t with v=n-1 degrees of freedom at α % level of significance.

Conclusion:

If $|t| < t_\alpha$, H_0 is accepted at α % level of significance.

If $|t| > t_\alpha$, H_0 is rejected at α % level of significance.

Problem:

1. **The mean lifetime of a sample of 25 bulbs is found as 1550h, with standard deviation of 120h. The company manufacturing the bulbs claims that the average life of their bulbs is 1600h. Is the claim acceptable at 5% level of significance?**

Solution:

Given sample size n=25, mean $\bar{x}$=1550, S.D.(S)=120, population mean μ=1600

Let $H_0 : \mu = 1600$ (the claim is acceptable)

$H_1 : \mu \neq 1600$ ($\mu \neq \bar{x}$)(two tailed test)

Under H_0 , the test statistic is $t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}} = \dfrac{1550 - 1600}{120/\sqrt{25}} = -2.0833$

$\therefore |t| = 2.0833$

From the table, for v=24, $t_{0.05}$=2.064. Since $|t| > t_{0.05}$

$\therefore H_0$ is rejected

Conclusion: The claim is not acceptable.

2. **Test made on the breaking strength of 10 pieces of a metal gave the following results: 578,572,570,568,572,570,570,572,596, and 584kg. Test if the mean breaking strength of the wire can be assumed as 577kg.**

Solution:

let us first compute sample mean $\bar{x}$ and sample standard deviation S and then test if $\bar{x}$ differs significantly from the population mean μ=577.

x	$x - \bar{x}$	$(x - \bar{x})^2$
578	2.8	7.84
572	-3.2	10.24
570	-5.2	27.04
568	-7.2	51.84
572	-3.2	10.24
570	-5.2	27.04
570	-5.2	27.04
572	-3.2	10.24
596	20.8	432.64
584	8.8	77.44
5752	**0**	**681.6**

Where

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5752}{10} = 575.2,$$

$$S^2 = \frac{\sum_{i=1} (x_i - \bar{x})^2}{n-1} = \frac{681.6}{9} = 75.733$$

Let $H_0 : \mu = \bar{x}$,

$H_1 : \mu \neq \bar{x}$

Under H_0, the test statistic is $t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{572.2 - 577}{\sqrt{75.733}/\sqrt{10}} = -1.74$

$\therefore |t| = 1.74$

Tabulated value of t for v=9 degrees of freedom $t_{0.05}$ =2.262

Since $|t| < t_{0.05}$. $\therefore H_0$ is accepted

<u>Conclusion:</u>

$\therefore$ The mean breaking strength of the wire can be assumed as 577kg at 5% level of significance.

3. A random sample of 10 boys had the following I.Q's: 70, 120, 110, 101, 88, 83, 95, 98,107,100. Do these data support the assumption of a population mean I.Q of 100 ? Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.

Solution:

Given $\mu = 100,\ n = 10$

Null Hypothesis:

$H_0 : \mu = 100$ i.e., The data are consist with the assumption of men IQ of 100 in the population

Alternate Hypothesis:

$H_1 : \mu \neq 100$ i.e., The data are consist with the assumption of men IQ of 100 in the population

Level of Significance : $\alpha = 5\% \Rightarrow \alpha = 0.05$

Test Statistic :

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

where $S^2 = \frac{1}{n-1} \sum (x - \bar{x})^2$

$$\bar{x} = \frac{\sum x}{n} = \frac{70+120+110+101+88+83+95+98+107+100}{10} = \frac{972}{10} = 97.2$$

$$S^2 = \frac{1}{10-1} \left[\begin{array}{l} (70-97.2)^2 + (120-97.2)^2 + (110-97.2)^2 + (101-97.2)^2 + (88-97.2)^2 \\ +(83-97.2)^2 + (95-97.2)^2 + (98-97.2)^2 + (107-97.2)^2 + (100-97.2)^2 \end{array} \right]$$

$$S^2 = \frac{1}{9}(1833.6) = 203.73 \Rightarrow S = 14.2734$$

$$t = \frac{97.2 - 100}{14.2734/\sqrt{10}} = \frac{2.8}{4.5136} = 0.6203$$

Table value : $t_{\alpha,n-1} = t_{5\%,10-1} = t_{0.05,9} = 2.262$ **(Two –tailed test)**

Conclusion :

Here $t > t_\alpha$

i.e., The table value >calculated value,

$\therefore$ we accept the null hypothesis and conclude that the data are consistent with the assumption of mean I.Q of 100 in the population.

To find the confidence limit:

$$\left(\bar{x} \mp t_\alpha \frac{S}{\sqrt{n}}\right) = \left(97.2 \mp 2.262 \times \frac{14.2734}{\sqrt{10}}\right) = \left(97.2 \mp (2.262)(4.514)\right) = (86.99, 107.41)$$

A reasonable range in which most of the mean I.Q. values of samples of 10 boys lies (86.99,107.41)

4. A random sample of 16 values from a normal population showed a mean of 41.5 inches and the sum of squares of deviations from this mean equal to 135 square inches. Show that the assumption of a mean of 43.5 inches for the population is not reasonable. Obtain 95 percent and 99 percent confidence limits for the same.

Solution:

Given $\bar{x} = 41.5,\ \mu = 43.5,\ n = 16$

Sum of squares of deviations from mean= $\sum\left(x - \bar{x}\right)^2 = 135$

The parameter of interest is μ .

Null Hypothesis H_0: μ =43.5 i.e., the assumption of a mean of 43.5 inches for the population is reasonable.

Alternative Hypothesis H_1: $\mu \neq 43.5$ i.e., the assumption of a mean of 43.5 inches for the population is not reasonable.

Level of significance: (i) α =5% = 0.05, degrees of freedom = 16–1=15

(ii)α =1% =0.01, degrees of freedom = 16–1=15

Test Statistic : $t = \dfrac{\bar{x} - \mu}{\dfrac{S}{\sqrt{n}}}$

where $S^2 = \dfrac{1}{n-1}\sum(x - \bar{x})^2 = \dfrac{1}{16-1}135 = 9 \Rightarrow S = 9$

$$t = \frac{41.5 - 43.5}{\dfrac{3}{\sqrt{16}}} = \frac{-8}{3} = -2.667 \Rightarrow |t| = 2.667$$

Conclusion:

(i) Since $|t|$=2.667 > 2.131 so we reject H_0 at 5% level of significance.

So we conclude that the assumption of mean of 43.5 inches for the population is not reasonable.

(ii) Since $|t|$=2.667 < 2.947 so we accept H_0 at 1% level of significance.

So we conclude that the assumption is reasonable.

95% confidence limits:

$$\left(\bar{x} \mp t_{\alpha} \frac{S}{\sqrt{n}}\right) = \left(41.5 \mp \left(2.947 \times \frac{3}{4}\right)\right) = (41.5 \mp 1.5983) = (39.9, 43.09)$$

$\therefore 39.902 < \mu < 43.098$

99% confidence limits:

$$\left(\bar{x} \mp t_{\alpha} \frac{S}{\sqrt{n}}\right) = \left(41.5 \mp \left(2.947 \times \frac{3}{4}\right)\right) = (41.5 \mp 2.2101) = (39.29, 43.71)$$

$\therefore 39.29 < \mu < 43.71$

5. Ten oil tins are taken at random from an automatic filling machine. The mean weight of the tins is 15.8 kg and standard deviation is 0.5 kg. Does the sample mean differ significantly from the intended weight of 16 kg?

Solution:

Given $\bar{x} = 15.8,\ \mu = 16,\ s = 0.50,\ n = 10$

Null Hypothesis H_0: $\mu = 16$ the sample mean weight is not different from the intended weight.

Alternative Hypothesis H_1: $\mu \neq 16$ i.e., the sample mean weight is not different from the intended weight.

Level of significance: α =5% = 0.05, degrees of freedom = 10-1=9

Test Statistic : $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$

$$t = \frac{15.8 - 16}{\dfrac{0.50}{\sqrt{10}}} = \frac{-0.2}{0.1581} = -1.27 \Rightarrow |t| = 1.27$$

Critical value : The critical value of t at 5% level of significance with degrees of freedom 9 is 2.26

Conclusion:

Here calculated value < table value.

so we accept H_0 at 5% level of significance.

Hence the sample mean weight is not different from the intended weight.

(ii) **Test of significance of the difference between means of two small samples:**

To test the significance of the difference between the means $\bar{x}_1$ and $\bar{x}_2$ of sample of size n_1 and n_2.

Under H_0, the test statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$,

$$where\ S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}\ \ or\ S^2 = \frac{\sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}\ (if\ s_1, s_2\ is\ not\ given\ directly)$$

Degrees of freedom(df) v = $n_1 + n_2$ -2

<u>Note:</u>

If $n_1 = n_2$ =n and if the pairs of values x_1 and x_2 are associated in some way (or correlated).

Then we use the statistic is $t = \dfrac{\bar{d}}{S/\sqrt{n-1}}$, *where* $\bar{d} = \dfrac{\sum d}{n}$ *and* $S^2 = \dfrac{\sum\left(d-\bar{d}\right)^2}{n}$

Degrees of freedom v = n-1

Confident Interval:

The confident interval for difference between two population means for small sample is $\left(\bar{x}_1 - \bar{x}_2\right) \mp t_\alpha \; S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$

<u>Problem:</u>

1. **Samples of two types of electric bulbs were tested for length of life and the following data were obtainded.**

Sample	Size	Mean	S.D
I	8	1234h	36h
II	7	1036h	40h

Is the difference in the means sufficient to warrant that type I bulbs are superior type II bulbs?

Solution:

Here $\bar{x}_1$ =1234, $\bar{x}_2$ =1036, n_1 =8, n_2 =7, s_1 =36, $s_2 = 40$

Let $H_0 : \bar{x}_1 = \bar{x}_2$,

$H_1 : \bar{x}_1 > \bar{x}_2$ (ie. Type I bulbs are superior to type II bulbs) (one tail test)

Under H_0 , the test statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$,

$$where\ S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = 40.7317$$

$$\therefore t = \frac{1234-1036}{40.7317\sqrt{\dfrac{1}{8}+\dfrac{1}{7}}} = 9.39$$

Degrees of freedom v = $n_1 + n_2$ -2=13

Tabulated value of t for 13 d.f. at 5% level of significance is $t_{0.05}$ =1.77

Since $|t| > t_{0.05}$. $\therefore$ H_0 is rejected. H_1 is accepted.

Conclusion:

Type I bulbs may be regarded superior to type II bulbs at 5% level of significance.

2. Two independent sample of size 8 and 7 contained the following value:

Sample I	**19**	**17**	**15**	**21**	**16**	**18**	**16**	**14**
Sample II	**15**	**14**	**15**	**19**	**15**	**18**	**16**	

Is the difference between the sample means significant?

Solution:

x_1	$x_1-\overline{x_1}$	$\left(x_1-\overline{x_1}\right)^2$	x_2	$x_2-\overline{x_2}$	$\left(x_2-\overline{x_2}\right)^2$
19	2	4	15	-1	1
17	0	0	14	-2	4
15	-2	4	15	-1	1
21	4	16	19	3	9
16	-1	1	15	-1	1
18	1	1	18	2	4
16	-1	1	16	0	0
14	-3	9			
136	**0**	**36**	**112**	**0**	**20**

$$\overline{x_1}=\frac{\sum x_1}{n}=\frac{136}{8}=17,\ \overline{x_2}=\frac{\sum x_2}{n}=\frac{112}{7}=16$$

$$S^2=\frac{\sum\left(x_1-\overline{x_1}\right)^2+\sum\left(x_2-\overline{x_2}\right)^2}{n_1+n_2-2}=\frac{36+20}{8+7-2}=4.3076\Rightarrow S=2.0754$$

Let $H_0:\overline{x_1}=\overline{x_2}$,

$H_1:\overline{x_1}\neq\overline{x_2}$ (Two tailed test)

Under H_0, the test statistic is $t=\dfrac{\overline{x_1}-\overline{x_2}}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}=\dfrac{17-16}{2.0754\sqrt{\dfrac{1}{8}+\dfrac{1}{7}}}=0.9309$

$|t|=0.9309$

Degrees of freedom v = v = n_1+n_2 -2=13

From the 't' table, v = 13 degrees freedom at 5% level of significance is $t_{0.05}$=2.16

Since $|t|<t_{0.05}$ $\therefore$ H_0 is accepted

<u>Conclusion:</u>

The two sample mean do not differ significantly at 5% level of significance.

3. The following data represent the biological values of protein from cow's milk and buffalo's milk:

Cow's milk	**1.82**	**2.02**	**1.88**	**1.61**	**1.81**	**1.54**
Buffalo's milk	**2.00**	**1.83**	**1.86**	**2.03**	**2.19**	**1.88**

Examine whether the average values of protein in the two samples significantly differ at 5% level.

Solution:

Given $n_1 = n_2 = 6$

H_0: $\mu_1 = \mu_2$ There is no significant difference between the means of the two samples.

H_1: $\mu_1 \neq \mu_2$ There is a significant difference between the means of the two samples.

Test Statistic: $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

x	y	$x - x$ $x - 1.78$	$(x-\bar{x})^2$	$y - y$ $y - 1.965$	$(y-\bar{y})^2$
1.82	2	0.04	0.0016	0.035	0.00123
2.02	1.83	0.24	0.0576	-0.135	0.01823
1.88	1.86	0.1	0.01	-0.105	0.01102
1.61	2.03	-0.17	0.0289	0.065	0.00425
1.81	2.19	0.03	0.0009	0.225	0.0506
1.54	1.88	-0.24	0.0576	-0.085	0.00723
Total 10.68	11.79		0.1566		0.09256

$$\bar{x} = \frac{\sum x}{n_1} = \frac{10.68}{6} = 1.78;\ \bar{y} = \frac{\sum y}{n_2} = \frac{11.79}{6} = 1.965$$

$$S^2 = \frac{1}{6+6-2}\left[0.1566 + 0.09256\right] = (0.1)(0.2492) = 0.0249 \quad \Rightarrow S = 0.1578$$

$$t = \frac{1.78 - 1.956}{(0.1578)\sqrt{\frac{1}{6} + \frac{1}{6}}} = \frac{-0.176}{(0.1578)(0.5774)} = \frac{-0.176}{0.0911} = 1.9319$$

Critical value: The critical value of t at 5% level of significance with degrees of freedom 10 is 2.228

Here calculated value < table value, we accept H_0

(i.e,) The difference between the mean protein values of the two varieties of milk is not significant at 5% level.

4. The following data relate to the marks obtaind by 11 students in 2 test, one held at the beginning of a year and the other at the end of the year intensive coaching.

Test 1	**19**	**23**	**16**	**24**	**17**	**18**	**20**	**18**	**21**	**19**	**20**
Test 2	**17**	**24**	**20**	**24**	**20**	**22**	**20**	**20**	**18**	**22**	**19**

Do the data indicate that the students have benefited by coaching?

Solution:

The given data relate to the marks obtained in 2 tests by the same set of students. Hence the marks in the 2 set can be regarded as correlated.

We use t-test for paired values.

Let $H_0 : \overline{x_1} = \overline{x_2}$,

$H_1 : \overline{x_1} < \overline{x_2}$ (one tailed test)

x_1	x_2	d = x_1 - x_2	$d^2 = \left(\overline{x}_1 - \overline{x}_2\right)^2$	d - d	$\left(d - \overline{d}\right)^2$
19	17	2	4	3	9
23	24	-1	1	0	0
16	20	-4	16	-3	9
24	24	0	0	1	1
17	20	-3	9	-2	4
18	22	-4	16	-3	9
20	20	0	0	1	1
18	20	-2	4	-1	1
21	18	3	9	4	16
19	22	-3	9	-2	4
20	19	1	1	2	4
		-11			**58**

$$\bar{d} = \frac{\sum d}{n} = -\frac{-11}{11} = 1 \quad S^2 = \frac{\sum\left(d - \bar{d}\right)^2}{n} = \frac{58}{11} = 5.272$$

the test statistic is $t = \dfrac{\bar{d}}{S/\sqrt{n-1}} = \dfrac{-1}{\sqrt{5.272}/\sqrt{10}} = -1.377 \Rightarrow |t| = 1.377$

from the table, v = n-1 = 10 (d.f.), $t_{0.05}$ =1.812

Since $|t| < t_{0.05}$ $\therefore$ H_0 is accepted

<u>Conclusion:</u>

The students have not benefitted by coaching.

5. **Ten Persons were appointed in the officer cadre in an office. Their performance was noted by giving a test and the marks were recorded out of 100.**

Employee	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Before training	**80**	**76**	**92**	**60**	**70**	**56**	**74**	**56**	**70**	**56**
After training	**84**	**70**	**96**	**80**	**70**	**52**	**84**	**72**	**72**	**50**

By applying the t-test, can it be concluded that the employees have been benefited by the training?

Solution:

Null Hypothesis H_0: $\mu_1 = \mu_2$ i.e., the employees have not been benefited by the training.

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$ i.e., the employees have been benefited by the training.

Level of significance: $\alpha = 5\% = 0.05$ (one tailed test)

Test Statistic : $t = \dfrac{\bar{d}}{\dfrac{S}{\sqrt{n}}}$

where $S^2 = \dfrac{1}{n-1}\sum(d-\bar{d})^2$ & $\bar{d} = \dfrac{\sum d}{n}$

Employees	**Before**	**After**	**d**	$(d-\bar{d})^2$
A	80	84	-4	0
B	76	70	6	100
C	92	96	-4	0
D	60	80	-20	256
E	70	70	0	16
F	56	52	4	64
G	74	84	-10	36
H	56	72	-16	144
I	70	72	-2	4
J	50	50	6	100
Total			44	44.4

$$\bar{d} = \frac{\sum d}{n} = \frac{-40}{10} = -4$$

$$S^2 = \frac{1}{n-1}\sum(d-\bar{d})^2 = \frac{1}{9}(720) = 80$$

$$t = \frac{\bar{d}}{\dfrac{S}{\sqrt{n}}} = \frac{-4}{8.94/\sqrt{10}} = -1.414 \Rightarrow |t| = 1.414$$

Critical value : The critical value of tat 5% level of significance with degrees of freedom 9 is 1.83

Conclusion:

Here calculated value < table value.

so we accept H_0

Hence the employees have not been benefited by the training.

6. The weight gains in pounds under two systems of feeding of calves of 10 pairs of identical twins is given below.

Twin pair	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Weight gains under System A	**43**	**39**	**39**	**42**	**46**	**43**	**38**	**44**	**51**	**43**
Sytem B	**37**	**35**	**34**	**41**	**39**	**37**	**37**	**40**	**48**	**36**

Discuss whether the difference between the two systems of feeding is significant.

Solution:

Null Hypothesis H_0: $\mu_1 = \mu_2$ i.e., there is no significance difference between the two system of feedings

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$ i.e., there is significance difference between the two systemsof feedings.

Level of significance: $\alpha = 5\% = 0.05$ (Two tailed test)

Test Statistic : $t = \dfrac{\bar{d}}{\dfrac{S}{\sqrt{n}}}$

where $S^2 = \dfrac{1}{n-1}\sum(d-\bar{d})^2$ & $\bar{d} = \dfrac{\sum d}{n}$

Twin Pair	System A x	System B y	$d = x - y$	$(d-\bar{d})^2$
1	43	37	6	2.56
2	39	35	4	0.16
3	39	34	5	0.36
4	42	41	1	11.56
5	46	39	7	6.76
6	43	37	6	2.56
7	38	37	1	11.56
8	44	40	4	0.16
9	51	48	3	1.96
10	43	36	7	6.76
Total			44	44.4

$$\bar{d} = \frac{\sum d}{n} = \frac{44}{10} = 4.4$$

$$S^2 = \frac{1}{n-1}\sum(d-\bar{d})^2 = \frac{1}{9}(44.4) = 4.93 \qquad \Rightarrow S = 2.08$$

$$t = \frac{d}{\frac{S}{\sqrt{n}}} = \frac{4.4}{2.08/\sqrt{10}} = 6.68$$

Critical value : The critical value of tat 5% level of significance with degrees of freedom 9 is 2.62

Conclusion:

Here calculated value < table value.

so we accept H_0

Hence there is no significance difference between the two systems of feedings.

II **<u>F-test</u>**

(i) To test whether if there is any significant difference between two estimates of population variance

(ii) To test if the two sample have come from the same population.

We use F-test:

The test statistic is given by $F = \frac{S_1^2}{S_2^2}, \; if \; S_1^2 > S_2^2$

Where $S_1^2 = \frac{n_1 s_1^2}{n_1 - 1}$ [n_1 is the first sample size] and $S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$ [n_2 is the second sample size]

The degrees of freedom $(v_1, v_2) = (n_1 - 1 \; n_2 - 1)$

Note :

1. If $S_1^2 < S_2^2$ then $F = \frac{S_2^2}{S_1^2}$ (always F > 1)

2. To test whether two independent samples have been drawn from the same normal population, we have to test

i) Equality of population means using t-test or z-test, according to sample size.

ii) Equality of population variances using F-test

<u>Problem:</u>

1. **A sample of size 13 gave an estimated population variance of 3.0, while another sample of size 15 gave an estimate of 2.5. Could both sample be from population with the same variance?**

<u>Solution:</u>

Given $n_1 = 13$, $n_2 = 15$, $S_1^2 = 3.0$, $S_2^2 = 2.5$

Let $H_0 : S_1^2 = S_2^2$ (the two samples have been drawn from populations with same variance}

$H_1 : S_1^2 \neq S_2^2$

The test statistics is $F = \frac{S_1^2}{S_2^2} = \frac{3}{2.5} = 1.2$

From the table, with degrees of freedom v = $(n_1 - 1 \; n_2 - 1)$ = (12, 14)

$F_{0.05} = 2.53$ Since $F < F_{0.05} \therefore H_0$ is accepted

<u>Conclusion:</u>

The two sample could have come from two normal population with the same variance.

2. **Two sample of size 9 and 8 give the sums of squares of deviations from their respective means equal to 160 and 91 respectively. Could both samples be from populations with the same variance?**

Solution:

Given $n_1=9,\ n_2=8,\ \sum\left(x-\bar{x}\right)^2=160,\ \sum\left(y-\bar{y}\right)^2=91$

$$S_1^2=\frac{\sum\left(x-\bar{x}\right)^2}{n_1-1}=\frac{160}{8}=20,\ S_2^2=\frac{\sum\left(y-\bar{y}\right)^2}{n_2-1}=\frac{91}{7}=13$$

Let $H_0:\sigma_1^2=\sigma_2^2$ (the two normal populations have the same variance}

$H_1:\sigma_1^2\neq\sigma_2^2$

The test statistics is $F=\frac{S_1^2}{S_2^2}=\frac{20}{13}=1.538$

From the table, with degrees of freedom v = ($n_1-1\ n_2-1$) = (8,7)

$F_{0.05}=3.73$ Since $F<F_{0.05}\therefore H_0$ is accepted

Conclusion:

The two sample could have come from two populations with the same variance.

3. **Two random samples gave the following data:**

Sample	Size	Mean	Variance
I	**8**	**9.6**	**1.2**
II	**11**	**16.5**	**2.5**

Cane we conclude that the two samples have been drawn from the same normal population?

Solution:

The two samples have been drawn from the same normal population we have to check

(i) the variance of the population do not differ significantly by F-test.

(ii) the sample means do not differ significantly by t-test.

(i) F-test:

Given $n_1=8,\ n_2=11,\ s_1^2=1.2,\ s_2^2=2.5,\ \bar{x}_1=9.6,\ \bar{x}_2=16.5$

$$S_1^2=\frac{n_1s_1^2}{n_1-1}=\frac{8(1.2)}{7}=1.37\quad S_2^2=\frac{n_2s_2^2}{n_2-1}=\frac{11(2.5)}{10}=2.75$$

Let $H_0:\sigma_1^2=\sigma_2^2$

$H_1:\sigma_1^2\neq\sigma_2^2$

The test statistics is $F=\frac{S_2^2}{S_1^2}$ $(\text{since } S_1^2<S_2^2)$

$$=\frac{2.75}{1.37}=2.007$$

From the table, $F_{0.05}\left(n_2-1,n_1-1\right)=F_{0.05}(10,7)=3.63$

Since $F<F_{0.05}\therefore H_0$ is accepted

(ii) t-test:(Equality of means)

Let $H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

Under H_0 , the test statistic is $t = \dfrac{\overline{x_1} - \overline{x_2}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$,

$$where\ S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{8(1.2) + 11(2.5)}{8 + 11 - 2}} = 1.4772$$

$$t = \frac{9.6 - 16.5}{1.4772\sqrt{\dfrac{1}{8} + \dfrac{1}{11}}} = -10.0525 \ \Rightarrow |t| = 10.0525$$

From the table ,with degrees of freedom $n_1 + n_2$ -2=17, $t_{0.05}$ =2.110

$\sin ce\ |t| > t_{0.05}\ \therefore H_0$ is rejected ie. $\mu_1 \neq \mu_2$

Conclusion:

$\therefore$ The two samples could not have been drawn from the same normal population.

4. Two independent samples of 5 and 6 items respectively had the following values of the following values of the variable:

Sameple1:	**21**	**24**	**25**	**26**	**27**	
Sameple2:	**22**	**27**	**28**	**30**	**31**	**36**

Can you say that the two samples came from the same population?

Solution:

Let $H_0 : \sigma_1^2 = \sigma_2^2$ and $\mu_1 = \mu_2$ (the two samples have been drawn from the same population)

$H_1 : \sigma_1^2 \neq \sigma_2^2$ and $\mu_1 \neq \mu_2$

(i) F-test : (Equality of variance)

x_1	$x_1 - \overline{x_1}$	$\left(x_1 - \overline{x_1}\right)^2$	x_2	$x_2 - \overline{x_2}$	$\left(x_2 - \overline{x_2}\right)^2$
21	-3.6	12.96	22	-7	49
24	-0.6	0.36	27	-2	4
25	0.4	0.16	28	-1	1
26	1.4	1.96	30	1	1
27	2.4	5.76	31	2	4
			36	7	49
123		21.2	174		108

$$\bar{x}_1 = \frac{\sum x_1}{n} = \frac{123}{5} = 24.6, \bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{174}{6} = 29$$

$$s_1^2 = \frac{\sum\left(x - \bar{x}\right)^2}{n_1 - 1} = \frac{21.2}{4} = 5.3,\ s_2^2 = \frac{\sum\left(x_2 - \bar{x}^2\right)^2}{n_2 - 1} = \frac{108}{5} = 21.6$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{5(5.3)}{4} = 6.625 \quad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{6(21.6)}{5} = 25.92$$

The test statistics is $F = \dfrac{S_2^2}{S_1^2}$ $(\sin ce\ S_1^2 < S_2^2)$

$$= \frac{25.92}{6.625} = 3.912$$

From the table, $F_{0.05}\left(n_2 - 1, n_1 - 1\right) = F_{0.05}(5,4) = 6.26$

Since $F < F_{0.05} \therefore H_0$ is accepted

(ii) <u>t-test</u>:(Equality of means)

Under H_0 , the test statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$,

$$where\ S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{5(5.3) + 6(21.6)}{5 + 6 - 2}} = 4.164$$

$$t = \frac{24.6 - 29}{4.16\sqrt{\dfrac{1}{5} + \dfrac{1}{6}}} = -1.746 \Rightarrow |t| = 1.746$$

From the table ,with degrees of freedom $n_1 + n_2$ -2=9, $t_{0.05}$ =2.262

$\sin ce\ |t| < t_{0.05}\ \therefore H_0$ is accepted ie. $\mu_1 \neq \mu_2$

<u>Conclusion:</u> $\therefore$ The two samples could have been drawn from the same normal population.

5. Two random samples gave the following results:

Sample	Size	Sample mean	Sum of squares of deviations from the mean
1	**10**	**15**	**90**
2	**12**	**14**	**108**

Test whether the samples come from the same normal population at 5% level of significance.

Solution:

A normal population has 2 parameters namely mean μ and variance σ^2. To test if independent samples have been drawn from the same normal population, we have to test

1) Equality of population means using t-test or z-test, according to sample size.

2) Equality of population variances using F-test.

Given $\bar{x} = 15,\ \bar{y} = 14,\ n_1 = 10,\ n_2 = 12, \sum(x - \bar{x})^2 = 90, \sum(y - \bar{y})^2 = 108$

i) t-test to test equality of population means:

Null hypothesis H_0: $\mu_1 = \mu_2$ there is no difference between the two population means.

Alternate Hypothesis H_1: $\mu_1 \neq \mu_2$ there is difference between the two population means.

Level of Significance : $\alpha = 5\% = 0.05$ (Two tailed test)

Test statistic: $t = \dfrac{\bar{x} - \bar{y}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

Where $S^2 = \dfrac{1}{n_1 + n_2 - 2}\left[\sum(x - \bar{x})^2 + \sum(y - \bar{y})^2\right] = \dfrac{1}{10 + 12 - 2}(90 + 108) = 9.9$

$S = \sqrt{9.9} = 3.146$

$$t = \frac{15-14}{3.146\sqrt{\frac{1}{10}+\frac{1}{12}}} = 0.742$$

Critical value: The critical value of t at 5% level of significance with degrees of freedom $n_1 + n_2 - 2 = 10 + 12 - 2 = 20$ is 2.086

Conclusion: calculated value < table value

H_0 is Accepted.

ii) F-test to test equality of populations variances:

Null Hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ The population Variances are equal

Alternative Hypothesis H_1: $\sigma_1^2 \neq \sigma_2^2$ The population Variances are not equal

Level of significance: $\alpha = 5\%$

Test Statistics:

$$F = \frac{S_1^2}{S_2^2}$$

Where $S_1^2 = \frac{1}{n_1 - 1}\sum(x-x)^2 = \frac{1}{10-1}(90) = 10$

$$S_1^2 = \frac{1}{n_1 - 1}\sum(y-y)^2 = \frac{1}{12-1}(108) = 9.818$$

Here $S_1^2 > S_2^2 \quad \therefore F = \frac{S_1^2}{S_2^2} = \frac{10}{9.818} = 1.02$

Critical value: The critical value of F at 5% level of significance with degrees of freedom $(n_1 - 1, n_2 - 1) = (9, 11)$ is 2.90

Here calculated value < table value, we accept H_0

Conclusion: Both null hypothesis $\mu_1 \neq \mu_2$ and $\sigma_1^2 = \sigma_2^2$ are accepted.

Hence we may conclude the two samples are drawn from same normal population.