Lecture.10

T-test – definition – assumptions – test for equality of two means-independent and paired t test

Student's t test

When the sample size is smaller, the ratio $Z = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n}}}$ will follow t distribution and not the standard normal distribution. Hence the test statistic is given as $t = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n}}}$ which follows normal distribution with mean 0 and unit standard deviation. This follows a t distribution with (n-1) degrees of freedom which can be written as $t_{(n-1)}$ d.f.

This fact was brought out by Sir William Gossest and Prof. R.A Fisher. Sir William Gossest published his discovery in 1905 under the pen name Student and later on developed and extended by Prof. R.A Fisher. He gave a test known as t-test.

Applications (or) uses

1. To test the single mean in single sample case.
2. To test the equality of two means in double sample case.
   (i) Independent samples(Independent t test)
   (ii) Dependent samples (Paired t test)
3. To test the significance of observed correlation coefficient.
4. To test the significance of observed partial correlation coefficient.
5. To test the significance of observed regression coefficient.

Test for single Mean

1. Form the null hypothesis

$H_o: \mu=\mu_o$

(i.e) There is no significance difference between the sample mean and the population mean

2. Form the Alternate hypothesis

$H_1: \mu \neq \mu_o$ (or $\mu > \mu_o$ or $\mu < \mu_o$)

ie., There is significance difference between the sample mean and the population mean

3. Level of Significance

The level may be fixed at either 5% or 1%

4. Test statistic

$$t = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right|$$ which follows t distribution with (n-1) degrees of freedom

where $\bar{x} = \frac{\sum x_i}{n}$ and $s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$

6. Find the table value of t corresponding to (n-1) d.f. and the specified level of significance.
7. Inference

If $t < t_{tab}$ we accept the null hypothesis H_0. We conclude that there is no significant difference sample mean and population mean

(or) if $t > t_{tab}$ we reject the null hypothesis H_0. (ie) we accept the alternative hypothesis and conclude that there is significant difference between the sample mean and the population mean.

Example 1

Based on field experiments, a new variety of green gram is expected to given a yield of 12.0 quintals per hectare. The variety was tested on 10 randomly selected farmer's fields. The yield (quintals/hectare) were recorded as 14.3,12.6,13.7,10.9,13.7,12.0,11.4,12.0,12.6,13.1. Do the results conform to the expectation?

Solution

Null hypothesis H_0: μ=12.0

(i.e) the average yield of the new variety of green gram is 12.0 quintals/hectare.

Alternative Hypothesis: H_1:$\mu \neq$ 12.0

(i.e) the average yield is not 12.0 quintals/hectare, it may be less or more than 12 quintals / hectare

Level of significance: 5 %

Test statistic:

$$t = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right|$$

From the given data

$$\sum x = 126.3 \qquad \sum x^2 = 1605.77$$

$$\bar{x} = \frac{\sum x}{n} = \frac{126.3}{10} = 12.63$$

$$s = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}} = \sqrt{\frac{1605.77 - 1595.169}{9}} = \sqrt{\frac{10.601}{9}}$$

$$= 1.0853$$

$$\frac{s}{\sqrt{n}} = \frac{1.0853}{\sqrt{10}} = 0.3432$$

Now $t = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right|$

$$t = \frac{12.63 - 12}{0.3432} = 1.836$$

Table value for t corresponding to 5% level of significance and 9 d.f. is 2.262 (two tailed test)

Inference

$t < t_{tab}$

We accept the null hypothesis H_0

We conclude that the new variety of green gram will give an average yield of 12 quintals/hectare.

Note

Before applying t test in case of two samples the equality of their variances has to be tested by using F-test

$$F = \frac{s_1^2}{s_2^2} \sim F_{(n_1 - 1, n_2 - 1)} \, d.f \quad \text{if } s_1^2 > s_2^2$$

or

$$F = \frac{s_2^2}{s_1^2} \sim F_{(n_2 - 1, n_1 - 1)} \, d.f \quad \text{if } s_2^2 > s_1^2$$

where s_1^2 is the variance of the first sample whose size is n_1.

s_2^2 is the variance of the second sample whose size is n_2.

It may be noted that the numerator is always the greater variance. The critical value for F is read from the F table corresponding to a specified d.f. and level of significance

Inference

$F < F_{tab}$

We accept the null hypothesis H_0.(i.e) the variances are equal otherwise the variances are unequal.

Test for equality of two Means (Independent Samples)

Given two sets of sample observation $x_{11}, x_{12}, x_{13} \ldots x_{1n}$, and $x_{21}, x_{22}, x_{23} \ldots x_{2n}$ of sizes n1 and n2 respectively from the normal population.

1. Using F-Test , test their variances

   (i) **Variances are Equal**

   Ho:., $\mu_1 = \mu_2$

   H_1 $\mu_1 \neq \mu_2$ (or $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$)

Test statistic

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2)\right|}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where the combined variance

$$s^2 = \frac{\left[\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}\right] + \left[\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}\right]}{n_1 + n_2 - 2}$$

The test statistic t follows a t distribution with (n1+n2-2) d.f.

(ii) Variances are unequal and n1=n2

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2)\right|}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

It follows a t distribution with $\left(\frac{n1 + n2}{2}\right) - 1$d.f.

(i) Variances are unequal and n1≠n2

$$t = \left|\frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right|$$

This statistic follows neither t nor normal distribution but it follows Behrens-Fisher d distribution. The Behrens – Fisher test is laborious one. An alternative simple method has been suggested by Cochran & Cox. In this method the critical value of t is altered as t_w (i.e) weighted t

$$t_w = \frac{t_1\left(\frac{s_1^2}{n1}\right) + t_2\left(\frac{s_2^2}{n2}\right)}{\frac{s_1^2}{n1} + \frac{s_2^2}{n2}}$$

where $t_1$is the critical value for t with (n1-1) d.f. at a dspecified level of significance and t2 is the critical value for t with (n2-1) d.f. at a dspecified level of significance and

Example 2

In a fertilizer trial the grain yield of paddy (Kg/plot) was observed as follows

Under ammonium chloride 42,39,38,60 &41 kgs

Under urea 38, 42, 56, 64, 68, 69,& 62 kgs.

Find whether there is any difference between the sources of nitrogen?

Solution

Ho: $\mu_1 = \mu_2$ (i.e) there is no significant difference in effect between the sources of nitrogen.

H_1: $\mu_1 \neq \mu_2$ (i.e) there is a significant difference between the two sources

Level of significance = 5%

Before we go to test the means first we have to test their variances by using F-test.

F-test

Ho:., $\sigma_1^2 = \sigma_2^2$

H1:., $\sigma_1^2 \neq \sigma_2^2$

$$s_1^2 = \frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}}{n1 - 1} = 82.5$$

$$s_2^2 = \frac{\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}}{n2 - 1} = 154.33$$

$$\therefore F = \frac{s_2^2}{s_1^2} \sim F_{(n_2 - 1, n_{1-} 1)} \; d.f \text{ if } s_2^2 > s_1^2$$

$$F = \frac{154.33}{32.5} = 1.8707$$

$F_{tab}(6,4)$ d.f. = 6.16

$\Rightarrow F < F_{tab}$

We accept the null hypothesis H_0. (i.e) the variances are equal.

Use the test statistic

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2)\right|}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$s^2 = \frac{\left[\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}\right] + \left[\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}\right]}{n_1 + n_2 - 2} = \frac{330 + 926}{10} = 125.6$$

$$t = \frac{|44 - 57|}{\sqrt{125.6\left(\frac{1}{7} + \frac{1}{75}\right)}} = 1.98$$

The degrees of freedom is 5+7-2= 10. For 5 % level of significance, table value of t is 2.228

Inference:

$t < t_{tab}$

We accept the null hypothesis H_0

We conclude that the two sources of nitrogen do not differ significantly with regard to the grain yield of paddy.

Example 3

The summary of the results of an yield trial on onion with two methods of propagation is given below. Determine whether the methods differ with regard to onion yield. The onion yield is given in Kg/plot.

Method I	**Method II**
n1=12	n2=12
$\bar{x}_1 = 25.25$	$\bar{x}_2 = 28.83$
SS1=186.25	SS2=737.6667
$s_1^2 = 16.9318$	$s_2^2 = 67.0606$

Solution

Ho:., $\mu_1=\mu_2$ (i.e) the two propagation methods do not differ with regard to onion yield.

H_1 $\mu_1 \neq \mu_2$ (i.e) the two propagation methods differ with regard to onion yield.

Level of significance = 5%

Before we go to test the means first we have to test their variability using F-test.

F-test

Ho: $\sigma_1^2=\sigma_2^2$

H1: $\sigma_1^2\neq\sigma_2^2$

$$s_1^2 = \frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}}{n1-1} = 16.9318$$

$$s_2^2 = \frac{\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}}{n2-1} = 67.0606$$

$$\therefore F = \frac{s_2^2}{s_1^2} \sim F_{(n_2-1, n_{1-}1)} \; d.f \;\; \text{if } s_2^2 > s_1^2$$

$$F = \frac{67.0606}{16.9318} = 3.961$$

$F_{tab}(11,11)$ d.f. = 2.82

$\Rightarrow F > F_{tab}$

We reject the null hypothesis H_0.we conclude that the variances are unequal.

Here the variances are unequal with equal sample size then the test statistic is

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2)\right|}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$s^2 = \frac{\left[\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}\right] + \left[\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}\right]}{n_1 + n_2 - 2}$$

$$s^2 = \frac{SS1 + SS2}{n_1 + n_2 - 2} = \frac{186.25 + 737.6667}{12 + 12 - 2} = 41.9962$$

$$t = \frac{25.25 - 28.83}{\sqrt{41.9962\left(\frac{1}{12} + \frac{1}{12}\right)}} = \frac{3.58}{\sqrt{6.9994}} = 1.353$$

t =1.353

The table value for $\left(\frac{12+12}{2}-1\right)$=11 d.f. at 5% level of significance is 2.201

Inference:

$t < t_{tab}$

We accept the null hypothesis H_0

We conclude that the two propagation methods do not differ with regard to onion yield.

Example 4

The following data relate the rubber yield of two types of rubber plants, where the sample have been drawn independently. Test whether the two types of rubber plants differ in their yield.

Type I	6.21	5.70	6.04	4.47	5.22	4.45	4.84	5.84	5.88	5.82	6.09	5.59
	6.06	5.59	6.74	5.55								

Type II	4.28	7.71	6.48	7.71	7.37	7.20	7.06	6.40	8.93	5.91	5.51	6.36

Solution

Ho:., $\mu_1=\mu_2$ (i.e) there is no significant difference between the two rubber plants.

H_1 $\mu_1\neq\mu_2$ (i.e) there is a significant difference between the two rubber plants.

Level of significance = 5%

Here

$$n1=16 \qquad n2=12$$

$$\sum x_1 = 90.09 \qquad \sum x_2 = 80.92$$

$$\bar{x}_1 = 5.63 \qquad \bar{x}_2 = 6.7431$$

$$\sum x_1^2 = 513.085 \qquad \sum x_2^2 = 561.64$$

Before we go to test the means first we have to test their variability using F-test.

F-test

Ho:., $\sigma_1^2=\sigma_2^2$

H1:., $\sigma_1^2\neq\sigma_2^2$

$$s_1^2 = \frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}}{n1-1} = 0.388$$

$$s_2^2 = \frac{\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}}{n2-1} = 1.452$$

$$\therefore F = \frac{s_2^2}{s_1^2} \sim F_{(n_2 - 1, n_{1-}1)}\ d.f.\ \text{if } s_2^2 > s_1^2$$

$$F = \frac{1.452}{0.388} = 3.742$$

$F_{tab}(11,15)$ d.f.=2.51

$\Rightarrow F > F_{tab}$

We reject the null hypothesis H_0. Hence, the variances are unequal.

Here the variances are unequal with unequal sample size then the test statistic is

$$t = \left| \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right|$$

$$t = \frac{(5.63 - 6.7431_2)}{\sqrt{\frac{0.388}{16} + \frac{1.452}{12}}} = 2.912$$

$$t_w = \frac{t_1\left(\frac{S_1^2}{n1}\right) + t_2\left(\frac{S_2^2}{n2}\right)}{\frac{S_1^2}{n1} + \frac{S_2^2}{n2}}$$

$t_1=t_{(16-1)}$ d.f.=2.131

$t_2=t_{(12-1)}$ d.f .=2.201

$$t_w = \frac{2.131\left(\frac{0.388}{16}\right) + 2.201\left(\frac{1.452}{12}\right)}{\frac{0.388}{16} + \frac{1.425}{12}} = 2.187$$

Inference:

$t > t_w$

We reject the null hypothesis H_0. We conclude that the second type of rubber plant yields more rubber than that of first type.

Equality of two means (Dependant samples)

Paired t test

In the t-test for difference between two means, the two samples were independent of each other. Let us now take particular situations where the samples are not independent.

In agricultural experiments it may not be possible to get required number of homogeneous experimental units. For example, required number of plots which are similar in all; characteristics may not be available. In such cases each plot may be divided into two equal parts and one treatment is applied to one part and second treatment to another part of the plot. The results of the experiment will result in two correlated samples. In some other situations two observations may be taken on the same experimental unit. For example, the soil properties before and after the application of industrial effluents may be observed on number of plots. This will result in paired observation. In such situations we apply paired t test.

Suppose the observation before treatment is denoted by x and the observation after treatment is denoted by y. for each experimental unit we get a pair of observation(x,y). In case of n experimental units we get n pairs of observations : (x1,y1), (x2,y2)…(xn,yn). In order to apply the paired t test we find out the differences (x1- y1), (x2-y2),..,(xn-yn) and denote them as d1, d2,…,dn. Now d1, d2…form a sample . we apply the t test procedure for one sample (i.e) $t = \frac{|\bar{d}|}{\sqrt{s^2/n}}$

where $\bar{d} = \frac{\sum di}{n}$, $s^2 = \frac{\sum di^2 - \frac{\left(\sum di\right)^2}{n}}{n-1}$

the mean $\bar{d}$ may be positive or negative. Hence we take the absolute value as $|\bar{d}|$. The test statistic t follows a t distribution with (n-1) d.f.

Example 5

In an experiment the plots where divided into two equal parts. One part received soil treatment A and the second part received soil treatment B. each plot was planted with sorghum. The sorghum yield (kg/plot) was absorbed. The results are given below. Test the effectiveness of soil treatments on sorghum yield.

Soil treatment A	49	53	51	52	47	50	52	53
Soil treatment B	52	55	52	53	50	54	54	53

Solution

H_0: $\mu_1 = \mu_2$, there is no significant difference between the effects of the two soil treatments

H_1: $\mu_1 \neq \mu_2$, there is significant difference between the effects of the two soil treatments

Level of significance = 5%

Test statistic

$$t = \frac{|\bar{d}|}{\sqrt{s^2 / n}}$$

x	y	d=x-y	d^2
49	52	-3	9
53	55	-2	4
51	52	-1	1
51	52	-1	1
47	50	-3	16
50	54	-4	16
52	54	-2	4
53	53	0	0
Total		-16	44

$$\bar{d} = \frac{\sum di}{n} = \frac{-16}{8} = -2,$$

$$s^2 = \frac{\sum di^2 - \frac{\left(\sum di\right)^2}{n}}{n-1} = 1.7143$$

$$t = \frac{|-2|}{\sqrt{1.7143/8}} = 4.32$$

Table value of t for 7 d.f. at 5% l.o.s is 2.365

Inference:

$t > t_{tab}$

We reject the null hypothesis H_0. We conclude that the is significant difference between the two soil treatments between A and B. Soil treatment B increases the yield of sorghum significantly,

Questions

1. The test statistic $F = \frac{s_1^2}{s_2^2}$ is used for testing

(a) $H_0: \mu_1 = \mu_2$ (b) $H_0: \sigma_1^2 = \sigma_2^2$
(c) $H0: \sigma_1 = \sigma_2$ (d) $H_0: \sigma_2 = \sigma_{0_2}$

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

2. In paired t test with n observations in each group the degrees of freedom is
(a) n (b) n-1 (c) n-2 (d) n+1

Ans: n-1

3. Student t- test is applicable in case of small samples.
Ans: True
4.F test is also known as variance ratio test.

Ans: True

5. In case of comparing the equality of two variances the greater variance should be taking in the numerator.

Ans: True

6. While comparing the means of two independent samples the variances of the two samples will be always equal.

Ans: False

7. Define t statistic.

8. Define F statistic.

9. Explain the procedure of testing the equality of two variances.

10. How to compare the means of two independent small samples.