Exercise.7
Students's t test – paired and independent t test

Test for single Mean (n<30)

1. Form the null hypothesis

$Ho: \mu=\mu_o$

(i.e) There is no significance difference between the sample mean and the population mean ie., $\mu=\mu_o$

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

$$t_{cal} = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right| \sim t_{(n-1)}\ d.f$$

Where $\bar{x} = \frac{\sum x_i}{n}$

Where s = $\sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$

5. Find the table value

$t_{tab}=t_{(0.05,(n-1)}$ d.f

1. Inference

If $t_{cal}<t_{tab}$

(i) We accept the null hypothesis H_0

(ii) There is no significant difference

(or) if $t_{cal}> t_{tab}$

(i)We reject the null hypothesis H_0 (ie) we accept the alternative hypothesis

(ii)There is significant difference between the sample mean and the population mean.

Example 1

Based on field experiments, a new variety green gram is expected to given an yield of 12.0quintals per hectare. The variety was tested on 10 randomly selected farmers 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 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

Level of significance: 5 %

Test statistic

$$t_{cal} = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right| \sim t_{(n-1)}\ d.f$$

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 $t_{cal} = \left| \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right| \sim t_{(n-1)}\ d.f$

$$= t_{cal} = \frac{12.63 - 12}{0.3432} = 1.836$$

Table value

$t_{(0.05,9)} = 2.262$ (two tailed test)

Inference

$t_{cal} < t_{tab}$

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

F-test is used to test the equality of two means

$$F = \frac{S_1^2}{S_2^2} \sim F_{(n_1 - 1, n_2 - 1)}\ d.f \text{ if } S_1^2 > S_2^2$$

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.

Otherwise

$$F = \frac{S_2^2}{S_1^2} \sim F_{(n_2 - 1, n_1 - 1)}\ d.f \text{ if } S_2^2 > S_1^2$$

Inference

$F_{cal} < F_{tab}$

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

Test for equality of two means (Independent Samples)

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

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2)\right|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n1+n2-2)} d.f$$

Where

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

Variances are equal

(a) When the samples have unequal variances and equal number of observations(n1=n2), the test statistic is

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2)\right|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{\left(\frac{n1+n2}{2}-1\right)} d.f$$

(b) When the samples have unequal variances and unequal number of observations (n1≠n2), the test statistic is

$$t = \left| \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \right|$$

This statistic follows neither t nor normal distribution but it follows Behrens-Fisher test. 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 test

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

Where $t_1=t_{(n1-1)}$ d.f

$t_2=t_{(n2-1)}$ d.f

Example 2

A group of 5 patients treated with medicine. A is of weight 42,39,38,60 &41 kgs. Second group of 7 patients from the same hospital treated with medicine B is of weight 38, 42, 56, 64, 68, 69, & 62 kgs. Find whether there is any difference between medicines?

Solution

Ho:., $\mu_1=\mu_2$ (i.e) there is no significant difference between the medicines A and B as regards on increase in weight.

H_1 $\mu_1\neq\mu_2$ (i.e) there is a significant difference between the medicines A and B

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

$$S_1^2 = \frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}}{n1-1} = 82.5$$

$$S_2^2 = \frac{\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}}{n2-1} = 154.33$$

$$\therefore F = \frac{S_2^2}{S_1^2} \sim F_{(n_2-1, n_{1\_}1)} \, d.f \text{ if } S_2^2 > S_1^2$$

$$F_{cal} = \frac{154.33}{32.5} = 1.8707$$

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

$\Rightarrow F_{cal}<F_{tab}$

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

Test statistic

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2)\right|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n1+n2-2)} d.f$$

Where

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

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

Table value

$t_{tab[(5+7-2)=10]}$d.f at 5% l.o.s = 2.228

Inference:

$t_{cal} < t_{tab}$

We accept the null hypothesis H_0

We conclude that the medicines A and also B do not differ significantly.

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 method do not differ with regard to onion yield.

H_1 $\mu_1\neq\mu_2$ (i.e) the two propagation method 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:., $\sigma_1^2\neq\sigma_2^2$

$$S_1^2 = \frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}}{n1-1} = 16.9318$$

$$S_2^2 = \frac{\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}}{n2-1} = 67.0606$$

$$\therefore F = \frac{S_2^2}{S_1^2} \sim F_{(n_2-1, n_1-1)}\ d.f \text{ if } S_2^2 > S_1^2$$

$$F_{cal} = \frac{67.0606}{16.9318} = 3.961$$

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

$\Rightarrow F_{cal}>F_{tab}$

We reject the null hypothesis H_0.(i.e) the variances are unequal.

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

$$t = \frac{\left|(\bar{x}_1 - \bar{x}_2)\right|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{\left(\frac{n1+n2}{2}-1\right)} d.f$$

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_{cal}=1.353

table value

$$t_{\left(\frac{n1+n2}{2}-1\right)} = t_{\left(\frac{12+12}{2}-1\right)} = t_{11}\ d.f \text{ at } 5\% \text{ l.os} = 2.201$$

Inference:

$t_{cal}<t_{tab}$

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 related the rubber percentage of two types of rubber plants, where the sample have been drawn independently. Test for their mean difference.

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 significance difference between the two samples.

H_1 $\mu_1 \neq \mu_2$ (i.e) there is a significance difference between the two samples.

Level of significance = 5%

Here

n1=16	n2=12
$\sum x_1 = 90.09$	$\sum x_2 = 80.92$
$\bar{x}_1 = 5.63$	$\bar{x}_2 = 6.7431$
$\sum x_1^2 = 513.085$	$\sum x_2^2 = 561.64$

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

$$S_1^2 = \frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n1}}{n1-1} = 0.388$$

$$S_2^2 = \frac{\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n2}}{n2-1} = 1.452$$

$$\therefore F = \frac{S_2^2}{S_1^2} \sim F_{(n_2-1, n_{1\_}1)}\ d.f \text{ if } S_2^2 > S_1^2$$

$$F_{cal} = \frac{1.452}{0.388} = 3.742$$

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

$\Rightarrow F_{cal}>F_{tab}$

We reject the null hypothesis H_0.(i.e) the variances are unequal.

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

$$t = \left|\frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\right| \sim t_w$$

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

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

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

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

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

Inference

$t_{cal} > t_{tab}$

We rejectthe null hypothesis H_0

(i.e) there is significant difference between the two rubber plants.

Equality of two means (Dependant samples)
Paired t test
Null hypothesis

$H_0 : \mu_1 = \mu_2$ ie the increments are just by chance

Alternative Hypothesis

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

test statistic

$$t = \frac{|\bar{d}|}{S/\sqrt{n}} \sim t(n-1)d.f$$

Where $\bar{d} = \frac{\sum di}{n}$, $S = \sqrt{\frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1}}$

$di = Xi - Yi$ (i = 1, 2, ……n)

Example 5

In certain food experiment to compare two types of baby foods A and B, the following results of increase in weight (lbs) we observed in 8 children as follows.

Food A(x)	49	53	51	52	47	50	52	53
Food B(y)	52	55	52	53	50	54	54	53

Examine the significance of increase in weight of children due to food B.

Solution

$H_0 : \mu_1 = \mu_2$, there is no significant difference between the two foods.

$H_1 : \mu_1 \neq \mu_2$, there is significant difference between the two foods.

Level of significance = 5%

test statistic

$$t = \frac{|\bar{d}|}{S/\sqrt{n}} \sim t(n-1)d.f$$

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**

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

$$S = \sqrt{\frac{\sum di^2 - \frac{\left(\sum di\right)^2}{n}}{n-1}} = 1.3093$$

$$t_{cal} = \frac{|-2|}{1.3093/\sqrt{8}} = 4.32$$

Table value:

t(8-1) d.f at 5% l.o.s = 2.365

Inference:

$t_{cal} > t_{tab}$

We reject the null hypothesis H_0

(i.e) there is significant difference between the two foods A and B.

Learning Exercise

1. 10 samples of leaves of the plant are chosen at random from a large population and their weight in grams are found to be as follows

   63 63 64 65 66 69 69 70 70 71

   From this data mean weight in universe is 65 gm. Can we assume this mean weight?

2. A health status survey in a few villages revealed that the normal serum protein value of children in that locality is 7.0 g/100ml. A group of 16 children, who received high protein food for a period of 6 months had serum protein values shown below. Can we consider that the mean serum protein level of these who were fed on high protein diet is different from that of the general population.

Children	1	2	3	4	5	6	7	8	9	10
Protein level g %	7.1	7.70	8.2	7.56	7.05	7.08	7.21	7.25	7.36	6.59
Children	11	12	13	14	15	16				
Protein level g %	6.85	7.9	7.27	6.56	7.93	8.5				

3. The following data related to the rate of diffusion of CO_2 through two series of different porosity, find out whether the diffusion rate same for both sides.

Diffusion through fine soil (x_1)	20	31	31	23	28	23	26	27	26	17	17	25
Diffusion through coarse soil (x_2)	19	30	32	28	15	26	35	18	25	27	35	34

4. A new variety of cotton was evolved by a breed. In order to compare its yielding ability with that of a ruling variety, an experiment was conducted in Completely Randomised Design. The yield (kg/plot) was observed. The summary of the results are given below. Test whether the new variety of cotton gives higher yield than the ruling variety.

New Variety	$n_1 = 9$	$\bar{x}_1$ =28.2	S_1^2=5.4430
Ruling Variety	$n_2 = 11$	$\bar{x}_2$ =25.9	S_2^2=1.2822

5. The iron contents of fruits before and after applying farm yard manure were observed as follows.

Fruit No:	1	2	3	4	5	6	7	8	9	10
Before Applying	7.7	8.5	7.2	6.3	8.1	5.2	6.5	9.4	8.3	7.5
After Applying	8.1	8.9	7.0	6.1	8.2	8.0	5.8	8.9	8.7	8.0

Is there any Significant difference between the mean iron contents in the fruits before & after the farm yarn manure?

1. 10 samples of leaves of the plant are chosen at random from a large population and their weight in grams are found to be as follows

63	63	64	65	66	69	69	70	70	71

From this data mean weight in universe is 65 gm. Can we assume this mean weight?

2. A health status survey in a few villages revealed that the normal serum protein value of children in that locality is 7.0 g/100ml. A group of 16 children, who received high protein food for a period of 6 months had serum protein values shown below. Can we consider that the mean serum protein level of these who were fed on high protein diet is different from that of the general population.

Children	1	2	3	4	5	6	7	8	9	10
Protein level g %	7.1	7.70	8.2	7.56	7.05	7.08	7.21	7.25	7.36	6.59
Children	11	12	13	14	15	16				
Protein level g %	6.85	7.9	7.27	6.56	7.93	8.5				

3., The following data related to the rate of diffusion of CO_2 through two series of different porosity, find out whether the diffusion rate same for both sides.

Diffusion through fine soil (x_1)	20	31	31	23	28	23	26	27	26	17	17	25
Diffusion through coarse soil (x_2)	19	30	32	28	15	26	35	18	25	27	35	34

4. A new variety of cotton was evolved by a breed. In order to compare its yielding ability with that of a ruling variety, an experiment was conducted in Completely Randomised Design. The yield (kg/plot) was observed. The summary of the results are given below. Test whether the new variety of cotton gives higher yield than the ruling variety.

New Variety	$n_1 = 9$	$\bar{x}_1$ =28.2	S_1^2=5.4430
Ruling Variety	$n_2 = 11$	$\bar{x}_2$ =25.9	S_2^2=1.2822

5. The iron contents of fruits before and after applying farm yard manure were observed as follows.

Is there any Significant difference between the mean iron contents in the fruits before & after the farm yarn manure?