Sri SAI RAM ENGINEERING COLLEGE
INSTITUTE OF TECHNOLOGY
West Tambaram, Chennai - 44

Sairam INSTITUTIONS

SAIRAM DIGITAL RESOURCES

YEAR II SEM IV

MA8391

PROBABILITY AND STATISTICS

Department of Information Technology

UNIT III

PROBABILITY AND RANDOM VARIABLES

3.5 Testing Hypothesis based on t and F test for mean and variance

SCIENCE & HUMANITIES

4 QUALITY EDUCATION

Definition: Degrees of freedom.

The number of independent variate used to compute the test statistic is called as the degrees of freedom.

State the applications of t-distribution.

The t-distribution is used to test the significance of difference between

a) The mean of a small sample and the mean of the population

b) The means of two populations

c) The coefficient of correlation in the small sample and that in the population which is assumed as zero.

Write down the test statistic for checking the equality of two population means using small samples.

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

What is the degree of freedom to be considered while testing for equality of two means using small samples?

$$n_1 + n_2 - 2$$

What is the assumption before applying t-test for equality of two means?

The two populations from which a sample each have been taken, have the same variance.

Is F -test, one-tailed or two-tailed?

Yes. One- tailed

Write any two properties of sampling distribution of " t ".

1. The probability curve of the t- distribution is similar to the standard normal curve, and is symmetric at bell - shaped and asymptotic to the t - axis.

2. For sufficiently large values of γ (degrees of freedom), the t - distribution tends to the standard normal distribution.

3. 3. Mean of the t - distribution is Zero and the variance is $\frac{\gamma}{\gamma - 2}$, if $\gamma > 2$ and is greater than 1 but it tends to 1 as $\gamma \to \infty$.

The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches?

Solution:

$$n = 10\ ,\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{10}(70+67+62+68+61+68+70+64+64+66) = 66$$

$$s^2 = \left(\frac{1}{n}\sum x_i^2\right) - \left(\bar{x}\right)^2 = \left(\frac{1}{10}\left(70^2+67^2+62^2+68^2+61^2+68^2+70^2+64^2+64^2+66^2\right)\right) - 66^2$$

$$= 9$$

$H_0 : \mu = 64 \qquad H_1 : \mu > 64$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{66-64}{\frac{3}{\sqrt{10-1}}} = 2$$

$|t| = 2$

From t - table, for $\gamma = 9$, $t_{5\%} = 1.833$

since $|t| > t_{5\%}$.hence H_0 is rejected and H_1 is accepted

Hence it is reasonable to believe that the average height is greater than 64 inches.

A sample of 10 boys had the following IQ's: 70, 120, 110, 101, 88, 83, 95, 98, 100 and 107. Test whether the population IQ may be 100.

SOLUTION:

$$n = 10\ , \bar{x} = \frac{1}{n}\sum x_i = \frac{1}{10}(70+120+110+101+88+83+95+98+100+107) = 97.2$$

$$s^2 = \left(\frac{1}{n}\sum x_i^2\right) - \left(\bar{x}\right)^2$$

$$= \left(\frac{1}{10}\left(70^2+120^2+110^2+101^2+88^2+83^2+95^2+98^2+100^2+107^2\right)\right) - 66^2 = 183.36$$

$H_0 : \mu = 100 \qquad H_1 : \mu \neq 100$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{97.2 - 100}{\frac{13.541}{\sqrt{10-1}}} = -0.620$$

$|t| = 2$

From t - table, for $\gamma = 9$, $t_{5\%} = 2.262$

since $|t| < t_{5\%}$ .hence H_0 is *accepted* and H_1 is *rejected*

Hence it is reasonable to believe that the population IQ is 100.

A machinist is expected to make engine parts with axle diameter of 1.75 cm. A random sample of 10 parts shows a mean diameter 1.85 cm and a S.D. of 0.1 cm. on the basis of this sample, would you say that the work of the machinist is inferior?

SOLUTION:

$n = 10,\ \overline{x} = 1.85, s = 0.1\ and\ \mu = 1.75$

$H_0 : \overline{x} = \mu \qquad H_1 : \overline{x} \neq \mu$

$$t = \frac{\overline{x} - \mu}{\dfrac{s}{\sqrt{n-1}}} = \frac{1.85 - 1.75}{\dfrac{0.1}{\sqrt{10-1}}} = 3$$

$|t| = 3$

From t - table, for $\gamma = 9,\ t_{5\%} = 2.262\ \ and\ \ t_{1\%} = 3.25$

Therefore H_0 is rejected and H_1 is accepted at 5% LOS, but H_0 is accepted and H_1 is rejected at 1% LOS. That is, at 5% LOS, the work of the machinist can be assumed to be inferior, but at 1% LOS, the work cannot be assumed to be inferior.

Test made on the breaking strength of 10 pieces of a metal wire gave the result:

578, 572, 570, 568, 572, 570, 570, 572, 596 and 584 kg.

Test if the mean breaking strength of the wire can be assumed as 577 Kg.

SOLUTION:

$$Given \;\; n = 10$$

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$= \frac{1}{10}(578+572+570+568+572+570+570+572+596+584) = 582$$

$$s^2 = \left(\frac{1}{n}\sum x_i^2\right) - \left(\bar{x}\right)^2$$

$$= \left(\frac{1}{10}\left(578^2+572^2+570^2+568^2+572^2+570^2+570^2+572^2+596^2+584^2)\right)\right) - 582^2$$

$$= 8.26$$

$$H_0 : \bar{x} = \mu \qquad H_1 : \bar{x} \neq \mu$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{575.2 - 577}{\frac{8.26}{\sqrt{10-1}}} = -0.65$$

$$|t| = 0.65$$

From t - table, for $\gamma = 9$, $t_{5\%} = 2.262$

since $|t| < t_{5\%}$ .hence H_0 is *accepted* and H_1 is *rejected*

Therefore the mean breaking strength of the wire can be assumed as 577 kg at 5% LOS.

Test if the difference in the means is significant for the following data:

Sample I : 76 68 70 43 94 68 33

Sample II: 40 48 92 85 70 76 68 22

SOLUTION:

$$n_1 = 7,\ \overline{X_1} = \frac{1}{7}(76+68+70+43+94+68+33)$$

$$= 64.571$$

$$S_1^2 = \frac{1}{7}\left(76^2+68^2+70^2+43^2+94^2+68^2+33^2\right)-(64.571)^2$$

$$= 358.872$$

$$n_2 = 8,\ \overline{X_2} = \frac{1}{8}(40+48+92+85+70+76+68+22) = 62.625$$

$$S_2^2 = \frac{1}{8}\left(40^2+48^2+92^2+85^2+70^2+76^2+68^2+22^2\right)-(62.625)^2$$

$$= 500.234$$

$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{64.571 - 62.625}{\sqrt{\frac{(7 \times 358.872) + (8 \times 500.234)}{7 + 8 - 2}\left(\frac{1}{7} + \frac{1}{8}\right)}}$$

$$= 0.168$$

From the t-table, ν =13(dof) $t_{5\%}$ = 2.160

Since $|t| < |t_{5\%}|$. Hence H_0 is accepted and H_1 is rejected

Threrfore it is reasonable to accept that there is no significant difference between the means.

The following data relate to the marks obtained by 11 students in two tests, one before and the other after an intensive coaching. Do the data indicate that the students have benefitted by coaching?

Test I:	19	23	16	24	17	18	20	18	21	19	20
Test II:	17	24	20	24	20	22	20	20	18	22	19

SOLUTION:

Given data relate to the marks obtained in two tests by the same set of students.

Hence the marks in the two tests can be regarded as correlated and so

the t-test for paired values should be used

Let $d = x_1 - x_2$

$d = 2, -1, -4, 0, -3, -4, 0, -2, 3, -3, 1$

$\sum d = -11$ and $\sum d^2 = 69$

now $\bar{d} = -1$ and $s = \sqrt{\frac{\sum d^2}{n} - (\bar{d})^2} = 2.296$

$H_0: \bar{d} = 0$ and $H_1: \bar{d} < 0$

$$t = \frac{\bar{d}}{s/\sqrt{n-1}} = \frac{-1}{2.296/\sqrt{10}} = -1.38$$

From the t-table, $\nu = 10$(dof) $t_{5\%} = 1.812$

Since $|t| < |t_{5\%}|$. Hence H_0 is accepted and H_1 is rejected

Threrfore it is reasonable to believe that the student have not benifited by coaching.

In one sample of 8 observations, the sum of the squares of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observations it was 102.6. Test whether this difference is significant at 5% level.

SOLUTION:

Given

$$n_1 = 8, \quad \sum\left(X - \bar{X}\right)^2 = 84.4, \; s_1^2 = \frac{\sum\left(X - \bar{X}\right)^2}{n_1}$$

$$n_2 = 10, \quad \sum\left(Y - \bar{Y}\right)^2 = 84.4, \; s_2^2 = \frac{\sum\left(Y - \bar{Y}\right)^2}{n_2}$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{\sum\left(X - \bar{X}\right)^2}{n_1 - 1} = \frac{84.4}{7} = 12.06$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{\sum\left(Y - \bar{Y}\right)^2}{n_2 - 1} = \frac{102.6}{9} = 11.4$$

$H_0 : \sigma_1^2 = \sigma_2^2 \qquad H_1 : \sigma_1^2 \neq \sigma_2^2$

$$F = \frac{S_1^2}{S_2^2} = \frac{12.06}{11.4} = 1.058$$

Number of degrees of freedom = $(n_1 - 1, n_2 - 1) = (7,9)$

$F(7,9)$ at 5% $LOS = 3.29$

the calculated value of F < the table value of F. Hence accept H_0.

Therefore the population variances are equal.

In one sample of 10 observations, the sum of the squares of the deviations of the sample values from the sample mean was 120 and in another sample of 12 observations it was 314. Test whether this difference is significant at 5% level of significance.

SOLUTION:

Given

$$n_1 = 10,\quad \sum\left(X-\overline{X}\right)^2 = 120,\ s_1^2 = \frac{\sum\left(X-\overline{X}\right)^2}{n_1}\ ;\quad n_2 = 12,\ \sum\left(Y-\overline{Y}\right)^2 = 314,\ s_2^2 = \frac{\sum\left(Y-\overline{Y}\right)^2}{n_2}$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1-1} = \frac{\sum\left(X-\overline{X}\right)^2}{n_1-1} = \frac{120}{9} = 13.33\ ;\quad S_2^2 = \frac{n_2 s_2^2}{n_2-1} = \frac{\sum\left(Y-\overline{Y}\right)^2}{n_2-1} = \frac{314}{11} = 28.55$$

$$H_0 : \sigma_1^2 = \sigma_2^2 \qquad H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$F = \frac{S_2^2}{S_1^2} = \frac{28.55}{13.33} = 2.14$$

Number of degrees of freedom = $(n_2-1,\ n_1-1) = (11,\ 9)$

$F(11,9)$ *at* 5% $LOS = 3.11$

the calculated value of F < the table value of F. Hence accept H_0.

We conclude that the sample might have come from two populations having the same variance

Two independent samples of size 9 and 7 from a normal populations had the following values of the variables:

Sample I	18	13	12	15	12	14	16	14	15
Sample II	16	19	13	16	18	13	15		

Do the estimates of the population variance differ significantly at 5% level?

SOLUTION:

X	Y	X2	Y2
18	16	324	256
13	19	169	361
12	13	144	169
15	16	225	256
12	18	144	324
14	13	196	169
16	15	256	225
14		196	
15		225	
129	110	1879	1760

$$\overline{X} = \frac{\sum X}{n_1} = \frac{129}{9} = 14.33 \quad ; \quad \overline{Y} = \frac{\sum Y}{n_2} = \frac{110}{7} = 15.71$$

$$s_1^2 = \frac{\sum X^2}{n_1} - \left(\overline{X}\right)^2 = \frac{1879}{9} - (14.33)^2 = 3.33 \quad ; \quad s_2^2 = \frac{\sum Y^2}{n_2} - \left(\overline{Y}\right)^2 = \frac{1760}{7} - (15.71)^2 = 4.49$$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{9 \times 3.33}{8} = 3.75 \quad ; \quad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{7 \times 4.49}{6} = 5.24$$

$$H_0 : \sigma_1^2 = \sigma_2^2 \qquad H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$F = \frac{S_2^2}{S_1^2} = \frac{5.24}{3.75} = 1.39$$

Number of degrees of freedom $= (n_2 - 1 \, , \, n_1 - 1) = (6 \, , \, 8)$

$F(6\,,8)$ at 5% $LOS = 3.58$

the calculated value of F < the table value of F. Hence accept H_0.

We conclude that the difference is not significant

Two random samples gave the following results:

	Size	mean	Sum of squares of deviation from mean
Sample I	10	15	90
Sample II	12	14	108

Test whether the samples have come from the same normal population.

SOLUTION:

Given

$n_1=10$ $\quad$ $n_2=12$

$\bar{x}=15$ $\quad$ $\bar{y}=14$

$\sum\left(x_i-\bar{x}\right)^2=90$ $\quad$ $\sum\left(y_i-\bar{y}\right)^2=108$

$$S_1^2=\frac{1}{n_1-1}\sum\left(x_i-\bar{x}\right)^2=\frac{1}{9}(90)=10$$

$$S_2^2=\frac{1}{n_2-1}\sum\left(y_i-\bar{y}\right)^2=\frac{1}{11}(108)=9.82$$

F-Test

$H_0: \sigma_1^2 = \sigma_2^2$ $\quad$ $:H_1 \; \sigma_1^2 \neq \sigma_2^2$

$$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.82} = 1.018$$

no. of degrees freedom = (9,11)

At 5% level of significance with (9,11) degrees of freedom, $F_{5\%} = 2.92$

$F < F_{5\%}$, H_0 is accepted and H_1 is rejected.

T-test

$H_0: \mu_1 = \mu_2$ $\quad$ $H_1 : \mu_1 \neq \mu_2$

$s_1^2 = \frac{1}{n_1}\sum\left(x_i - \bar{x}\right)^2 = \frac{1}{10}(90) = 9$ $\quad$ $s_2^2 = \frac{1}{n_2}\sum\left(y_i - \bar{y}\right)^2 = \frac{1}{12}(108) = 9$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{15-14}{\sqrt{\left(\frac{90+100}{20}\right)\left(\frac{1}{10}+\frac{1}{12}\right)}} = 0.742$$

Number of degrees of freedom = 10 + 12 - 2 = 20 and $t_{5\%}$(20 dof) = 2.228

Now $t < t_{5\%}$, H_0 is accepted and H_1 is rejected.

Hence by both F - test and T- test, we conclude that both the samples are came from same population.