M.A. Economics

Semester – II Paper CC-09 (Statistical Methods)

Module V: **Student's t test**

Dr. Sweta Sharan, Assistant Professor, Department of Economics, Magadh Mahila College, Patna University, Email:swetasharan87@gmail.com,Ph:9308489824

Introduction

Initial theoretical work on t-distribution was done by W.S.Gosset in early 1900. This distribution is used when sample size is 30 or less and the population standard deviation is unknown. The t-statistic is defined as:

$$\mathbf{t} = \frac{\bar{X} - \mu}{S} \times \sqrt{n}$$

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

The distribution has been derived mathematically under the assumption of a normally distributed population which has following form:

$$\mathbf{f(t)} = \mathbf{C}\left(1 + \frac{t^2}{\vartheta}\right)^{\frac{-(\vartheta+1)}{2}}$$

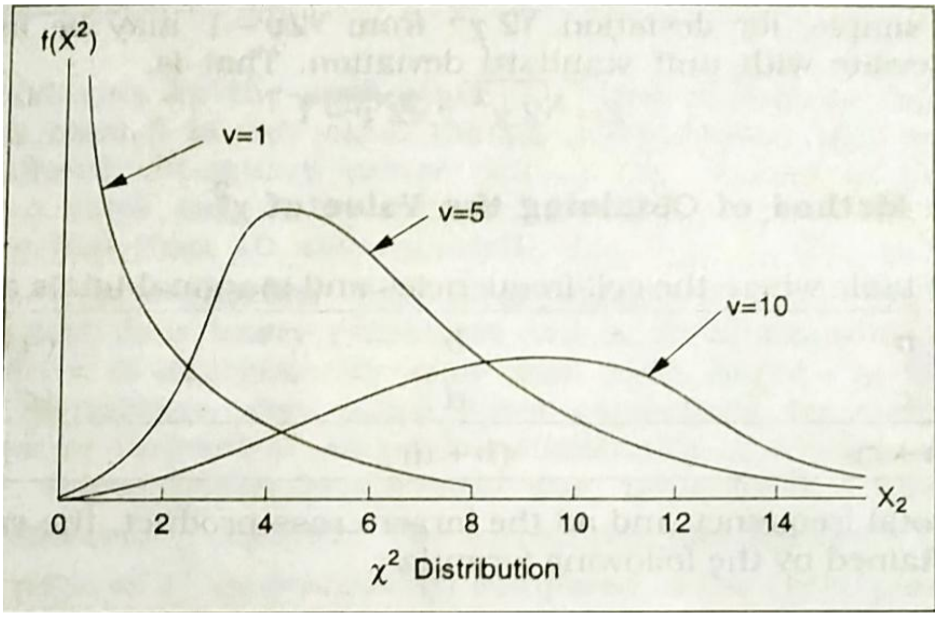
where $\mathbf{t} = \frac{\bar{X} - \mu}{S} \times \sqrt{n}$

C = a constant required to make area under the normal curve equal to unity

ϑ= n – 1, the number of degrees of freedom

Properties of t-distribution

1. The variable ranges from minus infinity to plus infinity
2. The constant c is actually a function of degrees of freedom
3. Just like the normal distribution, t-distribution is symmetrical and has zero mean
4. The variance of t-distribution is greater than one nut approaches one as the number of degree of freedom and thereby the sample size increases. Further it can be shown that for an infinite number of degrees of freedom, t-distribution and normal distribution are exactly equal. Hence a widely practised rule of thumb says that samples of size n > 30 may be considered large and standard normal distribution may be appropriately used as an approximation to t-distribution. This can be shown graphically as well:



The graph shows two distinct characteristics of t-distribution:

i) A t-distribution is lower at the mean and higher at the tails than a normal distribution
ii) The t-distribution has proportionately greater area in its tails than the normal distribution

Applications of t-distribution

1. **To test the significance of means of a random sample:** In determining whether the mean of a sample drawn from a normal population deviates significantly from a stated value, when the variance of the population is unknown, following statistic is used

$$t = \frac{\bar{X} - \mu}{S} \times \sqrt{n}$$

where $\bar{X}$ = mean of the sample

μ = the actual or hypothetical mean of the population

n = the sample size

S = Standard deviation of the sample = $\sqrt{\frac{\sum(X - \bar{X})^2}{n-1}}$ **or S =** $\sqrt{\frac{\sum d^2 - n(\bar{d})^2}{n-1}}$

d = deviations from assumed mean

Thumb Rule: If the calculated value of $|t|$exceeds the critical or table value at a given level of significance it is concluded that the difference between $\bar{X}$ and μ is significant at that level of significance and vice-versa.

Example 1: A random sample of size 16 has 53 as mean. The sum of the squares of deviation taken from mean is 135. Can this sample be regarded as taken from the population having 56 as mean? The level of significance is 5%.

Solution: Let us take the hypothesis (H_0) that there is no significant difference between the sample mean and hypothetical population mean. Applying t-test:

$$t = \frac{\bar{X} - \mu}{S} \times \sqrt{n}$$

$\bar{X} = 53$, $\mu = 56$, $n = 16$, $\sum(X - \bar{X})^2 = 135$

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}} = \sqrt{\frac{135}{15}} = 3$$

$$t = \frac{|53 - 56|}{3}\sqrt{16} = \frac{3 \times 4}{3} = 4$$

$$\vartheta = 16 - 1 = 15$$

Table value of t-statistic for degrees of freedom at 5% level of significance equal to 15 is 2.13.

The calculated value of t is more than table value, thus the hypothesis is rejected. Hence it is concluded that the sample does not comes from the population having 56 as its mean.

2. **Testing difference between means of two samples (independent samples):** If there are two independent samples of size n_1 and n_2 with means $\bar{X}_1$ and $\bar{X}_2$ and standard deviation S_1 and S_2 one may be interested in knowing that whether the samples came from same normal population or not. To carry out this test following statistic is used:

$$\mathbf{t} = \frac{\bar{X}_1 - \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

where $\bar{X}_1$ = mean of the first sample

$\bar{X}_2$ = mean of the second sample

n_1 = number of observations in first sample

n_2 = number of observations in second sample

S = combined standard deviation

The value of S is calculated as follows:

$$\mathbf{S} = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

In case when assumed mean is used instead of actual mean, combined standard deviation is calculated using the following formula:

$$\mathbf{S} = \sqrt{\frac{\sum(X_1 - A_1)^2 + \sum(X_2 - A_2)^2 - n_1(\bar{X}_1 - A_1)^2 - n_2(\bar{X}_2 - A_2)^2}{n_1 + n_2 - 2}}$$

Where, $\bar{X}_1$ = actual mean of the first sample

$\bar{X}_2$ = actual mean of the second sample

A_1 = assumed mean of first sample

A_1 = assumed mean of second sample

In case number of observations and standard deviation of the two samples are given, pooled estimate of standard deviation can be obtained as follows:

$$S = \sqrt{\frac{(n_1 - 1)\, S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}}$$

Example 2: Two types of drugs were used on 5 and 7 patients for reducing their weight. Drug A was imported and Drug B was indigenous. The decrease in the weight after using the drugs for six months was as follows:

Drug A	10	12	13	11	14	-	-
Drug B	8	9	12	14	15	10	9

Is there a significant difference in the efficacy of the two drugs? If not, which drug should one buy?

Solution: Let us take the hypothesis that there is no significant difference in the efficacy of the two drugs. Applying t-test:

$$t = \frac{\overline{X_1} - \overline{X_2}}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

X_1	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	X_2	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
10	-2	4	8	-3	9
12	0	0	9	-2	4
13	1	1	12	1	1
11	-1	1	14	3	9
14	2	4	15	4	16
			10	-1	1
			9	-2	4
$\sum X_1 = 60$		$\sum(X_1 - \bar{X}_1)^2 = 10$	$\sum X_2 = 77$		$\sum(X_2 - \bar{X}_2)^2 = 44$

$$\overline{X_1} = \frac{\sum X_1}{n_1} = \frac{60}{5} = 12$$

$$\overline{X_2} = \frac{\sum X_2}{n_2} = \frac{77}{7} = 11$$

$$S = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{10+44}{5+7-2}} = \sqrt{\frac{54}{10}} = 2.324$$

$$t = \frac{\overline{X_1} - \overline{X_2}}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$= \frac{12-11}{2.324} \times \sqrt{\frac{5 \times 7}{5+7}} = \frac{1.708}{2.324} = 0.735$$

Degree of freedom (ϑ) = $n_1 + n_2 - 2$ = 5 + 7 – 2 = 10

Critical or table value at ϑ = 10 is 2.228

The calculated value of t is less than the table value of t, the hypothesis is accepted. Hence there is no significance in the efficacy of two drugs. Since drug B is indigenous and there is no difference in the efficacy of two drugs one should use drug B.

3. **Testing difference between means of two samples (Dependent samples):** Two samples are said to be dependent when elements in one sample are related to those in the other in any significant manner. When samples are dependent they comprise the same number of elementary units. Degree of freedom in this case is (n-1). The t-test based on paired observations is defined using the following formula:

$$t = \frac{\bar{d} - 0}{S} \times \sqrt{n} \text{ or } t = \frac{\bar{d}\sqrt{n}}{S}$$

where, $\bar{d}$ = the mean of the differences

S = the standard deviation of the differences

Further S is calculated using the formula:

$$S = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n-1}}$$

Example 3: To verify whether a course in accounting improved performance a similar test was to 12 participants both before and after the course. The original marks recorded in alphabetical order of the participants – were 44,40,61,52,32,44,70,41,67,72,53 and 72. After the course, the marks were in the same order, 53,38,69,57,46,39,73,48,73,74,60 and 78. Was the course useful?

Solution: Let us assume the hypothesis that there is no difference in the marks obtained before and after the course. Applying t-test:

$$t = \frac{\bar{d}\sqrt{n}}{S}$$

Participants	Before (1^{st} test)	After (2^{nd} test)	d = (2^{nd} – 1^{st})	d^2
A	44	53	9	81
B	40	38	-2	4
C	61	69	8	64
D	52	57	5	25
E	32	46	14	196
F	44	39	-5	25
G	70	73	3	9
H	41	48	7	49
I	67	73	6	36
J	72	74	2	4
K	53	60	7	49
L	72	78	6	36
			$\sum d = 60$	$\sum d^2 = 578$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{60}{12} = 5$$

$$S = \frac{\Sigma(d - \bar{d})^2}{n-1} = \sqrt{\frac{578 - 12(5)^2}{12-1}} = \sqrt{\frac{278}{11}} = 5.03$$

$$t = \frac{5 \times \sqrt{12}}{5.03} = \frac{5 \times 3.464}{5.03} = 3.443$$

$$\vartheta = n - 1 = 12 - 1 = 11$$

For degree of freedom 11, table value of t-statistic is 2.201. Since calculated value is greater than table value, the hypothesis is rejected. Hence the course is useful.

4. **Testing the significance of an observed correlation coefficient**: In order to test whether the correlation coefficient of a population is zero i.e. to find out whether the variables in the population are uncorrelated, following test is applied:

$$t = \frac{r}{\sqrt{1 - r^2}} \times \sqrt{n - 2}$$

Where t is based on (n – 2) degrees of freedom. If the calculated value of t exceeds the table or critical value we conclude that r is significant at given level of significance and vice-versa.

Example 4: A random sample of 27 pairs of observation from a normal population gives a correlation coefficient of 0.42. Is it likely that the variables in the population are uncorrelated?

Solution: Taking the hypothesis that there is no significant difference in the sample correlation and correlation in the population. Applying t-test:

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2}$$

Given r = 0.42, n = 27

$$t = \frac{0.42}{\sqrt{1-(0.42)^2}} \times \sqrt{27-2}$$

$$t = \frac{.042}{0.908} \times 5 = 2.31$$

Degrees of freedom (ϑ) = n – 2 = 27 – 2 = 25

For ϑ = 25, table value of t-statistic is 1.708. Calculated value of t is more than the table value. Hence the hypothesis is rejected. So it is unlikely that the variables in the population are uncorrelated.