UNIT – III 1

6. TESTS OF SIGNIFICANCE
(Small Samples)

6.0 Introduction:

In the previous chapter we have discussed problems relating to large samples. The large sampling theory is based upon two important assumptions such as

(a) The random sampling distribution of a statistic is approximately normal and

(b) The values given by the sample data are sufficiently close to the population values and can be used in their place for the calculation of the standard error of the estimate.

The above assumptions do not hold good in the theory of small samples. Thus, a new technique is needed to deal with the theory of small samples. A sample is small when it consists of less than 30 items. ($n < 30$)

Since in many of the problems it becomes necessary to take a small size sample, considerable attention has been paid in developing suitable tests for dealing with problems of small samples. The greatest contribution to the theory of small samples is that of Sir William Gosset and Prof. R.A. Fisher. Sir William Gosset 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 popularly known as 't-test'.

6.1 t - statistic definition:

If $x_1, x_2, \ldots\ldots x_n$ is a random sample of size n from a normal population with mean μ and variance σ^2, then Student's t-statistic is defined as $t = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$

where $\bar{x} = \dfrac{\sum x}{n}$ is the sample mea and $S^2 = \frac{1}{n-1}\Sigma(x - \overline{x})^2$

is an unbiased estimate of the population variance σ^2 It follows student's t-distribution with $\nu = n - 1$ d.f

6.1.1 Assumptions for students t-test:

1. The parent population from which the sample drawn is normal.
2. The sample observations are random and independent.
3. The population standard deviation σ is not known.

6.1.2 Properties of t- distribution:

1. t-distribution ranges from $-\infty$ to ∞ just as does a normal distribution.
2. Like the normal distribution, t-distribution also symmetrical and has a mean zero.
3. t-distribution has a greater dispersion than the standard normal distribution.
4. As the sample size approaches 30, the t-distribution. approaches the Normal distribution

Applications of t-distribution:

The t-distribution has a number of applications in statistics, of which we shall discuss the following in the coming sections:

(i) t-test for significance of single mean, population variance being unknown.

(ii) t-test for significance of the difference between two sample means, the population variances being equal but unknown.

(a) Independent samples

(b) Related samples: paired t-test

6.2 Test of significance for Mean:

We set up the corresponding null and alternative hypotheses as follows:

$H_0: \mu = \mu_0$; There is no significant difference between the sample mean and population Mean.

$H_1: \mu \neq \mu_0$ ($\mu < \mu_0$ (or) $\mu > \mu_0$)

Level of significance:

5% or 1%

Calculation of statistic:

Under H_0 the test statistic is

$$t_0 = \left| \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \right| \quad \text{or} \quad \left| \frac{\bar{x} - \mu}{s / \sqrt{n-1}} \right|$$

$$\text{where } \bar{x} = \frac{\sum x}{n} \text{ is the sample mean}$$

$$\text{and} \quad S^2 = \frac{1}{n-1} \sum (x - \bar{x})^2 \quad \text{(or)} \quad s^2 = \frac{1}{n} \sum (x - \bar{x})^2$$

Expected value :

$$t_e = \left| \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \right| \sim \text{student's t-distribution with (n-1) d.f}$$

Inference :

If $t_0 \leq t_e$ it falls in the acceptance region and the null hypothesis is accepted and if $t_0 > t_e$ the null hypothesis H_0 may be rejected at the given level of significance.

Example 1:

Certain pesticide is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to weigh (in kg) as follows: 48 46 45 49 45 50 49 52 44 45

Test if the average packing can be taken to be 50 kg.

Solution:

Null hypothesis:

$H_0 : \mu = 50$ kgs in the average packing is 50 kgs.

Alternative Hypothesis:

$H_1 : \mu \neq 50$kgs (Two -tailed)

Level of Significance:

Let $\alpha = 0.05$

Calculation of sample mean and S.D

X	d = x –48	d^2
50	2	4
49	1	1
52	4	16
44	–4	16
45	–3	9
48	0	0
46	–2	4
45	–3	9
49	+1	1
45	–3	9
Total	–7	69

$$\bar{x} = A + \frac{\sum d}{n}$$

$$= 48 + \frac{-7}{10}$$

$$= 48 – 0.7 = 47.3$$

$$S^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right]$$

$$= \frac{1}{9}\left[69 - \frac{(7^2)}{10}\right]$$

$$= \frac{64.1}{9} = 7.12$$

Calculation of Statistic:

Under H_0 the test statistic is :

$$t_0 = \left| \frac{\bar{x} - \mu}{\sqrt{S^2 / n}} \right|$$

$$= \left| \frac{47.3 - 50.0}{\sqrt{7.12/10}} \right|$$

$$= \frac{2.7}{\sqrt{0.712}} = 3.2$$

Expected value:

$$t_e = \left| \frac{\bar{x} - \mu}{\sqrt{S^2/n}} \right| \text{ follows t distribution with (10–1) d.f}$$

$$= 2.262$$

Inference:

Since $t_0 > t_e$, H_0 is rejected at 5% level of significance and we conclude that the average packing cannot be taken to be 50 kgs.

Example 2:

A soap manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop was 140 dozens. After the campaign, a sample of 26 shops was taken and the mean sales was found to be 147 dozens with standard deviation 16. Can you consider the advertisement effective?

Solution:

We are given

$n = 26$; $\bar{x} = 147$dozens; $s = 16$

Null hypothesis:

H_0: $\mu = 140$ dozens i.e. Advertisement is not effective.

Alternative Hypothesis:

H_1: $\mu > 140$kgs (Right -tailed)

Calculation of statistic:

Under the null hypothesis H_0, the test statistic is

$$t_0 = \left| \frac{\bar{x} - \mu}{S/\sqrt{n-1}} \right|$$

$$= \left| \frac{147 - 140}{16/\sqrt{25}} \right| = \frac{7 \times 5}{16} = 2.19$$

Expected value:

$$t_e = \left|\frac{\bar{x}-\mu}{s/\sqrt{n-1}}\right| \quad \text{follows t-distribution with } (26-1) = 25 \text{ d.f}$$

$$= 1.708$$

Inference:

Since $t_0 > t_e$, H_0 is rejected at 5% level of significance. Hence we conclude that advertisement is certainly effective in increasing the sales.

6.3 Test of significance for difference between two means:

6.3.1 Independent samples:

Suppose we want to test if two independent samples have been drawn from two normal populations having the same means, the population variances being equal. Let $x_1, x_2, \ldots x_{n_1}$ and $y_1, y_2, \ldots\ldots y_{n_2}$ be two independent random samples from the given normal populations.

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ i.e. the samples have been drawn from the normal populations with same means.

Alternative Hypothesis:

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

Test statistic:

Under the H_0, the test statistic is

$$t_0 = \left|\frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right|$$

$$\text{where } \bar{x} = \frac{\Sigma x}{n_1}\ ; \ \bar{y} = \frac{\Sigma y}{n_2}$$

$$\text{and } S^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2\right] = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

Expected value:

$$t_e = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ follows t-distribution with $n_1 + n_2 - 2$ d.f

Inference:

If the $t_0 < t_e$ we accept the null hypothesis. If $t_0 > t_e$ we reject the null hypothesis.

Example 3:

A group of 5 patients treated with medicine 'A' weigh 42, 39, 38, 60 and 41 kgs: Second group of 7 patients from the same hospital treated with medicine 'B' weigh 38, 42 , 56, 64, 68, 69 and 62 kgs. Do you agree with the claim that medicine 'B' increases the weight significantly?

Solution:

Let the weights (in kgs) of the patients treated with medicines A and B be denoted by variables X and Y respectively.

Null hypothesis:

$H_0 : \mu_1 = \mu_2$

i.e. There is no significant difference between the medicines A and B as regards their effect on increase in weight.

Alternative Hypothesis:

$H_1 : \mu_1 < \mu_2$ (left-tail) i.e. medicine B increases the weight significantly.

Level of significance : Let $\alpha = 0.05$

Computation of sample means and S.Ds

Medicine A

X	$x - \bar{x}$ ($\bar{x} = 46$)	$(x - \bar{x})^2$
42	-4	16
39	-7	49
48	2	4
60	14	196
41	-5	25
230	**0**	**290**

$$\bar{x} = \frac{\sum x}{n_1} = \frac{230}{5} = 46$$

Medicine B

Y	$y-\bar{y}$ ($\bar{y}=57$)	$(y-\bar{y})^2$
38	–19	361
42	–15	225
56	–1	1
64	7	49
68	11	121
69	12	144
62	5	25
399	**0**	**926**

$$\bar{y} = \frac{\sum y}{n_2} = \frac{399}{7} = 57$$

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum(x-\bar{x})^2 + \sum(y-\bar{y})^2\right]$$

$$= \frac{1}{10}[290 + 926] = 121.6$$

5 + 7 - 2

Calculation of statistic:

Under H_0 the test statistic is

$$t_0 = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{46-57}{\sqrt{121.6\left(\frac{1}{5}+\frac{1}{7}\right)}}$$

$$= \frac{11}{\sqrt{121.6 \times \frac{12}{35}}}$$

$$= \frac{11}{6.57} = 1.7$$

Expected value:

$$t_e = \left| \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$ follows t-distribution with (5+7–2) =10 d.f

$= 1.812$

Inference:

Since $t_0 < t_e$ it is not significant. Hence H_0 is accepted and we conclude that the medicines A and B do not differ significantly as regards their effect on increase in weight.

Example 4:

Two types of batteries are tested for their length of life and the following data are obtained:

	No of samples	Mean life (in hrs)	Variance
Type A	9	600	121
Type B	8	640	144

Is there a significant difference in the two means?

Solution:

We are given

$n_1=9$; $\bar{x}_1=600$hrs; $s_1^2=121$; $n_2=8$; $\bar{x}_2=640$hrs; $s_2^2=144$

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ i.e. Two types of batteries A and B are identical i.e. there is no significant difference between two types of batteries.

Alternative Hypothesis:

$H_1 : \mu_1 \neq \mu_2$ (Two- tailed)

Level of Significance:

Let $\alpha = 5\%$

Calculation of statistics:

Under H_0, the test statistic is

$$t_0 = \left| \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$

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

$$= \frac{9 \times 121 + 8 \times 144}{9 + 8 - 2}$$

$$= \frac{2241}{15} = 149.4$$

$$\therefore \quad t_0 = \left| \frac{600 - 640}{\sqrt{149.4\left(\frac{1}{9} + \frac{1}{8}\right)}} \right|$$

$$= \frac{40}{\sqrt{149.4\left(\frac{17}{72}\right)}} = \frac{40}{5.9391} = 6.735$$

Expected value:

$$t_e = \left| \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$ follows t-distribution with $9+8-2 = 15$ d.f

$= 2.131$

Inference:

Since $t_0 \geq t_e$ it is highly significant. Hence H_0 is rejected and we conclude that the two types of batteries differ significantly as regards their length of life.

6.3.2 Related samples –Paired t-test:

In the t-test for difference of means, the two samples were independent of each other. Let us now take a particular situations where

(i) The sample sizes are equal; i.e., $n_1 = n_2 = n$(say), and

(ii) The sample observations $(x_1, x_2, \ldots\ldots x_n)$ and $(y_1, y_2, \ldots\ldots y_n)$ are not completely independent but they are dependent in pairs.

That is we are making two observations one before treatment and another after the treatment on the same individual. For example a business concern wants to find if a particular media of promoting sales of a product, say door to door canvassing or advertisement in papers or through T.V. is really effective. Similarly a pharmaceutical company wants to test the efficiency of a particular drug, say for inducing sleep after the drug is given. For testing of such claims gives rise to situations in (i) and (ii) above, we apply paired t-test.

Paired – t –test:

Let $d_i = X_i - Y_i$ $(i = 1, 2, \ldots\ldots n)$ denote the difference in the observations for the i^{th} unit.

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) $\mu_1 < \mu_2)$

Calculation of test statistic:

$$t_0 = \left| \frac{\bar{d}}{S/\sqrt{n}} \right|$$

where $\bar{d} = \dfrac{\sum d}{n}$ and $S^2 = \dfrac{1}{n-1}\sum(d-\bar{d})^2 = \dfrac{1}{n-1}\left[\sum d^2 - \dfrac{(\sum d)^2}{n}\right]$

Expected value:

$$t_e = \left| \frac{\bar{d}}{S/\sqrt{n}} \right| \text{ follows t-distribution with } n-1 \text{ d.f}$$

Inference:

By comparing t_0 and t_e at the desired level of significance usually 5% or 1%, we reject or accept the null hypothesis.

Example 5:

To test the desirability of a certain modification in typists desks, 9 typists were given two tests of as nearly as possible the same nature, one on the desk in use and the other on the new type. The following difference in the number of words typed per minute were recorded:

Typists	A	B	C	D	E	F	G	H	I
Increase in number of words	2	4	0	3	−1	4	−3	2	5

Do the data indicate the modification in desk promotes speed in typing?

Solution:

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ i.e. the modification in desk does not promote speed in typing.

Alternative Hypothesis:

$H_1 : \mu_1 < \mu_2$ (Left tailed test)

Level of significance: Let $\alpha = 0.05$

Typist	d	d^2
A	2	4
B	4	16
C	0	0
D	3	9
E	1	1
F	4	16
G	3	9
H	2	4
I	5	25
	$\Sigma d = 16$	$\Sigma d^2 =$ [illegible]

$$\bar{d} = \frac{\sum d}{n} = \frac{16}{9} = 1.778$$

$$S = \sqrt{\frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right]}$$

$$= \sqrt{\frac{1}{8}\left[84 - \frac{(16)^2}{9}\right]} = \sqrt{6.9} = 2.635$$

Calculation of statistic:

Under H_0 the test statistic is

$$t_0 = \left|\frac{\bar{d}.\sqrt{n}}{S}\right| = \frac{1.778 \times 3}{2.635} = 2.024$$

Expected value:

$t_e = \left|\frac{\bar{d}.\sqrt{n}}{S}\right|$ follows t- distribution with $9 - 1 = 8$ d.f

$= 1.860$

Inference:

When $t_0 < t_e$ the null hypothesis is accepted. The data does not indicate that the modification in desk promotes speed in typing.

Example 6:

An IQ test was administered to 5 persons before and after they were trained. The results are given below.

Candidates	I	II	III	IV	V
IQ before training	110	120	123	132	125
IQ after training	120	118	125	136	121

Test whether there is any change in IQ after the training programme (test at 1% level of significance)

Solution:

Null hypothesis:

$H_0 : \mu_1 = \mu_2$ i.e. there is no significant change in IQ after the training programme.

Alternative Hypothesis:

$H_1 : \mu_1 \neq \mu_2$ (two tailed test)

Level of significance :

$\alpha = 0.01$

x	110	120	123	132	125	Tot.
y	120	118	125	136	121	.
d = x−y	−10	2	−2	−4	4	−10
d^2	100	4	4	16	16	140

$$\bar{d} = \frac{\sum d}{n} = \frac{-10}{5} = -2$$

$$S^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{(\sum d)^2}{n}\right]$$

$$= \frac{1}{4}\left[140 - \frac{100}{5}\right] = 30$$

Calculation of Statistic:

Under H_0 the test statistic is

$$t_0 = \left|\frac{\bar{d}}{S/\sqrt{n}}\right|$$

$$= \left|\frac{-2}{\sqrt{30/5}}\right|$$

$$= \frac{2}{2.45}$$

$$= 0.816$$

Expected value:

$$t_e = \left|\frac{\bar{d}}{\sqrt{S^2/n}}\right|$$ follows t-distribution with $5-1=4$ d.f

$$= 4.604$$

Inference:

Since $t_0 < t_e$ at 1% level of significance we accept the null hypothesis. We therefore, conclude that there is no change in IQ after the training programme.

7/03/2022

t-test for an observed sample correlation co-efficient :-

UNIT-III Continuation !

we have set up the null and Alternative hypothesis

Null hypothesis $H_0 : \rho = 0$ against

Alternative hypothesis $H_1 : \rho \neq 0$

level of significance : $\alpha = 5\%$ (or) 1%

Under the H_0, the test statistic is given by

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2} \sim t(n-2)\,df$$

Here r is the sample correlation of co-efficient.

Problem:-

28 (1). A random sample of 27 pairs of observations from a normal population gave a correlation co-efficient of 0.6 is the significant of correlation co-efficient in the population.

Solution:-

Let us consider the Null hypothesis H_0 to test the hypothesis the correlation co-efficient is the significant of correlation in the population that is

$$H_0 : \rho = 0 \text{ vs } H_1 : \rho \neq 0$$

under the H_0, the test statistic is given by,

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2} \sim t(n-2)\,df$$

1/03/2022

$$t_0 = \frac{0.6}{\sqrt{1-(0.6)^2}} \times \sqrt{27-2}$$

$$= \frac{0.6}{\sqrt{0.64}} \times \sqrt{25}$$

$$= \frac{0.6}{0.8} \times 5$$

$$t_0 = 3.75$$

Table value is 2.06 for 25 df for 5% LOS

Inference:-

since calculated value greater than Table value that is $3.75 > 2.06$ therefore reject the null hypothesis at 5% level of significance. that is we may be conclude that the correlation co-efficient is not the significant of correlation in the population.