Hypothesis - Small Sample Tests: t - test and F - test

[Academic Script]

Subject	**:**	**Business Economics**
Course	**:**	**B.A., 3rd Semester, Undergraduate**
Paper No.	**:**	**304 (Three Zero Four)**
Paper Title	**:**	**Basic Econometrics**
Unit No.	**:**	**1 (One)**
& Title	**:**	**Sampling and Hypothesis Testing**
Lecture No.	**:**	**3 (Three)**
& Title	**:**	**Hypothesis - Small Sample Tests: t - test and F - test**

Credits

Subject Expert & Presenter

Dr. M. N. Patel

Professor, Dept. of Statistics,

I/c Director,

K. S. School of Business Management,

Gujarat University,

Ahmedabad.

Subject Co-Ordinator

Dr. M. N. Patel

Professor, Dept. of Statistics,

I/c Director,

K. S. School of Business Management,

Gujarat University,

Ahmedabad.

Editing & Graphics

Akash Choudhary

Floor Assistant

Hemant Upadhyaya

Multimedia

Gaurang Sondarva

Camera

Maqbool Chavda

Technician

Kirit Dave

Technical Asst. & Sound Recording

Smita Bhatt

Archana Patel

Helper & Support

Bharat Chauhan

Jagdish Jadeja

Research

Prof. Pooja Khatri

K. S. School of Business Management,

Gujarat University,

Ahmedabad.

Producer

Dr. P. P. Prajapati

Dinesh Goswami

1. **Introduction of Small Sample Tests:** When sample size is small (*n < 30)* the central limit theorem does not assure to assume the distribution of sample statistics as normal. When the sample size is small the special probability distributions are used to determine critical value for the test statistic. Here we must have to assume that the small sample is drawn from normal population and based on it some important distributions and their applications in testing of hypothesis are discussed.

2. t – Distribution:

If $x_1, x_2, \ldots\ldots\ldots, x_n$ are independent observations drawn from a normal population with mean μ and standard deviation σ then the probability distribution of statistic $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$, where $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ is called student's t – distribution or simply t – distribution with n-1 degrees of freedom. Degrees of freedom indicates the number of independent terms in statistic, as the restriction on statistic increases the degrees of freedom (df) decreases by that much restrictions.

Important properties of t – distribution:

Few important properties are as under

i.It is a continuous distribution with range -∞ to ∞.

ii. Probability curve of t – distribution is symmetric about its mean.

iii. Its mean is always zero.

iv. For large value of sample size it follows normal distribution.

Applications of t – distribution:

Some important applications of t – distribution are as under

i. To test the significance of the mean

ii. To test the significance of two means.

iii. To test the effectiveness of treatment given to data values(Paired t-test)

Let us see the above applications by considering suitable illustrations.

i. To test the significance of the mean:

This test is used to test the hypothesis that there is no significant difference between sample and population mean or the sample has been drawn from the normal population with specified mean.

Illustration 1: A sample of 4 observations from a normal population gives mean 1.75 and variance 0.6875. Test the hypothesis that the population mean is 2.

Solution:

Step 1: Here

$n = 4, \ \overline{x} = 1.75, S^2 = 0.6875, \ \therefore S = 0.8292, \ \mu = 2$ and $\alpha = 0.05, \ df = n - 1 = 3.$

Step 2: $H_0 : \mu = 2$ vs $H_1 : \mu \neq 2$

Step 3: $|Difference| = |\bar{x} - \mu| = 0.25$

$$SE = \frac{S}{\sqrt{n-1}} = \frac{0.8292}{\sqrt{3}}$$

Under H_o the test statistic is

$$|t| = \frac{|Difference|}{S.E} = 0.52$$

Step 4: Since the given test is two tailed test hence the critical value for 5% level of significant with 3 df is obtained from the tables of t - distribution and is 3.183.

Since $0.52 < 3.183$ hence the null hypothesis is not rejected.

Step 5: At 5% level of significance we do not reject null hypothesis hence we conclude that the population mean is 2.

ii. To test the significance of two means:

This test is used to test the hypothesis that there is no significant difference between two population means or two sample has been drawn from the normal population with equal mean.

Illustration 2: The following information is obtained for two samples drawn from two normal populations:

Sample	Size	Mean	S.D
1	10	12	3.162
2	12	15	5.115

Can we say that the two population means are equal?

Solution:

Step 1: Here n_1 = 10, $\bar{x}_1 = 12\,, S_1 = \ 3.162$

n_2 = 12, $\bar{x}_2 = 15\,,\ S_2 = \ 5.115$

Step 2: $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$

Step 3: difference $|\bar{x}_1 - \bar{x}_2| = \ |12 - 15| = 3$

Under H_o the test statistic is

$$|t| = \frac{|\bar{x}_1 - \bar{x}_2|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$S = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{10(3.162)^2 + 12(5.115)^2}{10+12-2}} = \sqrt{\frac{10(3.162)^2 + 12(5.115)^2}{20}} =$ 4.5494

Hence

$|t| = \frac{|12-15|}{4.5494\sqrt{\frac{1}{10} + \frac{1}{12}}} = \frac{3}{1.9479} =$ 1.5401

Step 4: Since the given test is two tailed test hence the critical value for 5% level of significant with 20df is obtained from the tables of t - distribution and is 2.074. Since $1.5401 < 2.074$ hence the null hypothesis is not rejected.

Step 5: At 5% level of significance we do not reject null hypothesis hence we conclude that the population means are equal.

iii. To test the effectiveness of treatment given to data values(Paired t-test)**.**

This test is used to test the hypothesis that, the treatment given to the observations is ineffective i.e. there is no significant difference in the value of observations before and after giving treatment.

Illustration 3: A data of sales of an item in six shops before and after special sales promotional campaigning are as under:

Shops	1	2	3	4	5	6
Before campaigning	53	28	32	48	50	42
After campaigning	58	32	30	50	56	45

Can the campaigning be regarded as success?

Solution:

Step 1: Here

Shops	Before campaigning	After campaigning	d	d^2
1	53	58	5	25
2	28	32	4	16
3	32	30	- 2	04
4	48	50	2	04
5	50	56	6	36
6	42	45	3	09
		Total	18	94

$$n = 6,\ \bar{d} = \frac{\sum d}{n} = \frac{18}{6} = 3$$

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

$$= \frac{1}{6}\left(94 - \frac{(18)^2}{6}\right)$$

$$= 6.667$$

$$\therefore S = 2.58$$

and $\alpha = 0.05,\ df = n - 1 = 5$

Step 2: $H_0 : Compaigning\ \ is\ ineffective$ vs $H_1 : Compaigning\ is\ \ effective$

Step 3: $|Difference| = |\bar{d}| = 3$

$$SE = \frac{S}{\sqrt{n-1}} = \frac{2.58}{\sqrt{5}}$$

Under H_o the test statistic is

$$|t| = \frac{|Difference|}{S.E} = 2.6$$

Step 4: Since the critical value for 5% level of significant with 5 df is obtained from the tables of t - distribution and is 2.571.Since $2.6 > 2.571$, the null hypothesis is rejected.

Step 5: At 5% level of significance with we reject null hypothesis hence we conclude that the campaigning is successful.

3. F – Distribution:

If $\chi_1^2$is a chi square variate with n_1 degrees of freedom and χ_2^2 is another independent chi square variate with n_2 degrees of freedom then the distribution of the ratio $\frac{\chi_1^2/n_1}{\chi_2^2/n_2}$is called F- distribution with (n_1, n_2) degrees of freedom. To test the equality of means of two small samples we use t – distribution. In applying t – distribution it is assumed that the populations from which the samples are drawn have equal variances. Hence before applying t – test, it is necessary to test the equality of population variances. For that F – distribution is used. Consider the following illustration for the application of F – distribution.

Illustration 4: The following information is obtained for two samples drawn from two normal populations:

Sample	Size	Mean	S.D
1	10	12	3.162
2	12	15	5.115

Can we say that the two population variances are equal?

Solution:

Step 1: Here

$n_1 = 10, \ n_2 = 12,$

$S_1^2 = (3.162)^2 = 9.9982,$

$S_2^2 = (5.115)^2 = 26.1632$

and $\alpha = 0.05$, $df = (n_2 - 1, n_1 - 1) = (11, 9)$

Step 2: $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$

Step 3: The test statistic is (since $S_2^2 > S_1^2$)

$$F = \frac{n_2 S_2^2/(n_2-1)}{n_1 S_1^2/(n_1-1)} = \frac{28.5417}{11.1091} = 2.5692$$

Step 4: Since the critical value for 5% level of significant with (11, 9) df is obtained from the tables of F - distribution and is 3.10

Since $2.5692 < 3.10$ hence the null hypothesis is not rejected.

Step 5: At 5% level of significance we accept null hypothesis hence we conclude that the both the population variances are equal.

We know that for small samples, the significance between two means can be tested by t – test. But in practice we may like to test equality of means of several samples, in such situation t – test is not applicable. In such situations F – test is used to test average variations among various samples which is known as analysis of variance is used. It is a process of determining various factors which influence the value of variance. It is a technique of dividing the total variations of the data into component variations due to various sources. The process of separating only one source of variation from total variation is called one – way analysis of variance (one – way ANOVA) and if we separate two sources of variation then it is called two – way ANOVA. For the application of F – distribution in such cases it is assumed that all samples have been independently drawn from the same population with same variance. Let us consider the following illustrations for the ANOVA.

Illustration 5: The following information is obtained by using three fertilizers in different plots.

Fertilizers	Yield (Qtl/acre)			
A	1	4	3	3
B	6	5	4	2
C	7	3	5	6

Test the hypothesis that there is no significant difference between the mean productivities due to different fertilizers.

Solution:

Here only one source is to be determined so it is called one way analysis of variance.

Step 1: Here

Fertilizers	Yield (Qtl/acre)				Total
A	1	4	3	3	11
B	6	5	4	2	17
C	7	3	5	6	21
					49

Grand total $G = 49$

Correction Factor $cf = \frac{G^2}{N} = \frac{(49)^2}{12} = 200.08$

Total Sum of Squares

$$TSS = \sum\sum x_{ij}^2 - cf$$
$$= (1^2 + 4^2 + 3^2 + ... + 6^2) - 200.08$$
$$= 34.92$$

Sum of Squares due to fertilizer (Row Sum of Squares)

$$RSS = \frac{\sum R_j^2}{k} - cf$$

$$= \frac{(11)^2 + (17)^2 + (21)^2}{4} - 200.08$$

$$= 12.67$$

Error Sum of Squares

$$ESS = TSS - RSS$$

$$= 34.92 - 12.67$$

$$= 22.25$$

Step 2:

H_0:There is no significant difference in the mean productivities due to different fertilizers

Step 3:

ANOVA Table

Source	SS	df	MSS	F_c	F_t
Due to fertilizers	12.67	3 – 1 = 2	6.335	2.56	4.26
Due to Error	22.25	11 – 2 = 9	2.47		
Total	34.92	12 – 1 = 11			

Step 4: Since the critical value for 5% level of significant with (2,9) df is obtained from the tables of F - distribution and is 4.26

Since $2.56 < 4.26$ hence the null hypothesis is not rejected.

Step 5: At 5% level of significance with wedo not reject the null hypothesis hence we conclude that the variations due to fertilizers is insignificant.

Illustration 6: The following information is obtained for the productions of three operators on 4 machines.

Operators	Machines (production in units)			
A	560	540	580	560
B	580	550	600	590
C	570	560	560	590

Analyze the data for

(i) There is no significant difference among operators.

(ii) There is no significant difference among machines.

Solution:

Here two sources are to be determined so it is called two way analysis of variance. Here we subtract 580 from each observation and then divide it by 10 then we have the following table

Step 1: Here

Operators	Machines				Total
A	-2	-4	0	-2	-8
B	0	-3	2	1	0
C	-1	-2	-2	1	-4
Total	-3	-9	0	0	-12

Grand total $G = -12$

Correction Factor $cf = \frac{G^2}{N} = \frac{(-12)^2}{12} = 12$

Total Sum of Squares

$$TSS = \sum\sum x_{ij}^2 - cf$$
$$= 48 - 12$$
$$= 36$$

Sum of Squares due to operators (Row Sum of Squares)

$$RSS = \frac{\sum R_j^2}{k} - cf$$
$$= \frac{(-8)^2 + (0)^2 + (-4)^2}{4} - 12$$
$$= 8$$

Sum of Squares due to Machine (Column Sum of Squares)

$$CSS = \frac{\sum C_j^2}{m} - cf$$
$$= \frac{(-3)^2 + (-9)^2 + (0)^2 + (0)^2}{3} - 12$$
$$= 18$$

Error Sum of Squares

$$ESS = TSS - RSS - CSS$$
$$= 36 - 8 - 18$$
$$= 10$$

Step 2:

H_0:(i) There is no significant difference among operators

(ii) There is no significant difference among machines

Step 3:

ANOVA Table

Source	SS	df	MSS	F_c	F_t
Due to Operators	8	2	4	2.4	5.14

Due to machines	18	3	6	3.6	4.76
Due to Error	10	6	1.67		
Total	36	11			

Step 4: Since the critical values for 5% level of significant with (2,6) df and (3,6) df are respectively obtained from the tables of F - distribution and are $5.14\ and\ 4.76$.

Since $(a)\ 2.4 < 5.14$ hence the null hypothesis (i) is not rejected and $(b)\ 3.6 < 4.76$ hence the null hypothesis (ii) is not rejected.

Step 5: At 5% level of significance we accept null hypothesis hence we conclude that the variations due to operators and machines are insignificant.

4. Summary:

We have discussed small sample tests , i.e. sample size less then 30, based on t-distribution and F-distribution. The tests based on t-distributions are (i) test for single mean, (ii) test for two means for independent samples, (iii) paired t- test for dependent samples. The applications of F-distributions are (i) testing the equality of two variances, (ii) testing the equality of three or more means of independent populations(ANOVA test). For the above tests we have discussed various types of examples.