Hypothesis Testing – Small Samples

- TEST OF SIGNIFICANCE FOR SMALL SAMPLES:
  - When the size of the sample is less than 30, then the sample is called small sample.
  - The tests used in case of large samples are not applicable for small samples because the assumptions on which they are based do not hold good for small samples.
  - In small samples, population standard deviation is not known and as such it is estimated from the random sample drawn from the population.

- **THE ‘t’ DISTRIBUTION OR STUDENT’S ‘t’ DISTRIBUTION:**
  - **It is very important an**d useful test of significance for the small samples
  - The 't test is mainly **applied in testing the si**gnificance of the:

    (i) mean of the sample,

    (ii) difference between the means of two independent/dependent samples,

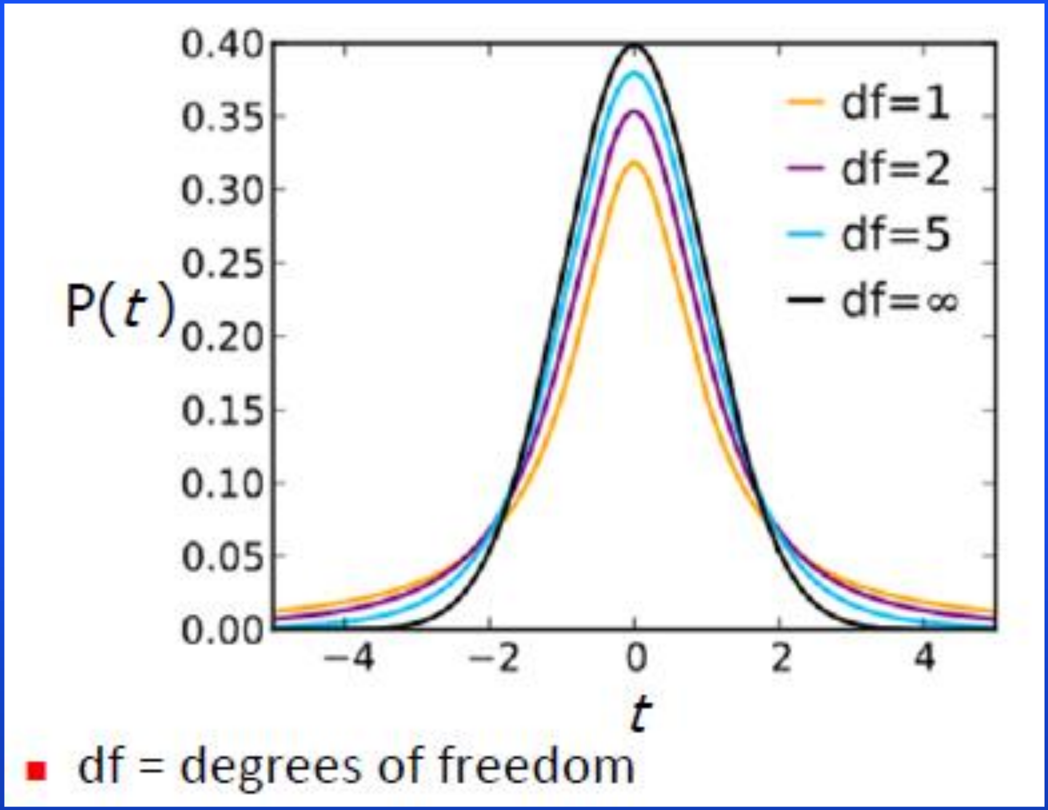
    (iii) observed coefficient of correlation
  - Here the p**opulation standard deviation is not known**

Degree of Freedom

- For a fixed value of the mean the number of free choices is called degree of freedom
- For example, in a distribution 2, 3, 5, 8, 7 the mean is 5. To have a distribution containing 5 values with mean 5, 4 values can be independently chosen but the 5th value has to be taken in such a way that the mean is 5. So, the degree of freedom is 4
- We have defined degree of freedom for a fixed value of mean but in certain situations like Chi-square test, degree of freedom is calculated in different way.

❖ Properties of t-distribution

- **t-distribution is unimodal** distribution
- **The probability distribution curve is symmetrical about** the line $t = 0$.
- **It is bell shaped curve just like a normal curve with its tail a little higher above the abscissa than the normal curve. Its spread increases as degree of freedom ‘*k*’ increases.**



❖ Properties of t-distribution

- t-distribution has only one parameter k, the degree of freedom
- The constants of t distribution are as follows:

  Mean = 0 for $k \geq 2$

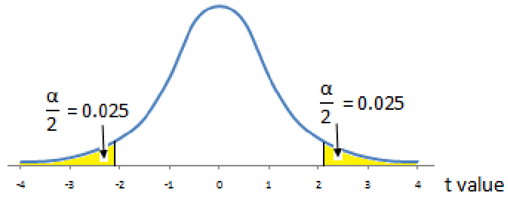
  Variance $\sigma^2 = \frac{k}{k-2}$ for $k \geq 3$
- The area under t-distribution curve for $t < t_0$, is determined by the equation

$$P(t < t_0) = \int_{-\infty}^{t_0} f(t)dt$$

  Students need not integrate actually for the area as the tables of area under the curve for different values of *t* are available and vice versa, (see table for student's '*t*' distribution).
- t-distribution tends to normal distribution as k increases. For practical purposes, t is taken as equivalent to the normal distribution provided k > 30. t-distribution has tremendous utility in testing of hypothesis about one population mean or about equality of two population means when standard deviation of population is not known.

Student's t Distribution Table

For example, the t value for 18 degrees of freedom is 2.101 for 95% confidence interval (**2-Tail** $\alpha = 0.05$).



	90%	95%	97.5%	99%	99.5%	99.95%	**1-Tail Confidence Level**
	80%	90%	95%	98%	99%	99.9%	**2-Tail Confidence Level**
	0.100	0.050	0.025	0.010	0.005	0.0005	**1-Tail Alpha**
df	0.20	0.10	0.05	0.02	0.01	0.001	**2-Tail Alpha**
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192	
2	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991	
3	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240	
4	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103	
5	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688	
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588	
7	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079	
8	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413	
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809	
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869	
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370	
12	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178	
13	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208	
14	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405	
15	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728	
16	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150	
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651	
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216	
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834	
20	1.3253	1.7247	2.0860	2.5280	2.8453	3.8495	
21	1.3232	1.7207	2.0796	2.5176	2.8314	3.8193	
22	1.3212	1.7171	2.0739	2.5083	2.8188	3.7921	
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676	
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.7454	
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.7251	
26	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066	
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.6896	
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.6739	
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.6594	
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.6460	

❖ Applicatiopn of t-distribution

- To test the significance of the mean of a random sample
- To test the difference between mean of two independent samples
- To test the difference between mean of two dependent samples (Paired t-Test)
- To test the significance of an observed correlation coefficient

❖ To Test the Significance of the Mean of a Random Sample

- Suppose a random sample $x_1, x_2, \dots, x_n$ of size $n(n \geq 2)$ has been drawn from a normal population whose variance σ^2 is unknown. On the basis of this random sample the aim is to test
- Ho: There is no significant difference between the sample mean $\bar{x}$ and the population mean μ

  i.e. $\mu = \mu_0$
- Test Statistics

  $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ where $\bar{x}$ is the mean of sample

  And $s^2 = \frac{1}{n-1}\sum(x - \bar{x})^2$ with degree of freedom $n - 1$

- The table giving the value of *t* required for significance at various levels of probability and for different degree area called t-tables which are given in statistical tables by Fishers and Yates.
- The t-distribution has a different values for each degree of freedom and when the degrees of freedom are infinitely large, the t-distribution is equivalent to normal distribution and the probabilities shown in the normal distribution tables are applicable
- The computed value is compared with the tabulated value at 5% or 1% levels of significance and at $(n - 1)$ degree of freedom and accordingly the null hypothesis is accepted or rejected.

Example 1

A random sample of size 20 from a normal population has mean 42 and standard deviation of 5. Test the hypothesis that the population mean is 45. Use 5% level of significance.

Example Solution:

Here $n = 20,\ \bar{x} = 42$

$\mu = 45, s = 5$

Let us assume null hypothesis that There is no significant difference between the sample mean and population mean

H_0: $\mu = 45$

H_1: $\mu \neq 45$ (Two tailed test)

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{42-45}{5/\sqrt{20}} = -2.683$$

α = 0.05

t Value : 2. 683

As tabulated value of t at 5% level for 19 degree of freedom is $t_{0.05} = 2.09$. As the calculated value of $|t| = 2.683 > t_{0.05}$ for 19 degree of freedom Ho is rejected

Decision: Reject Ho at α = 0.05(At 5%)

Conclusion:

There is significant difference between the sample mean and population mean

Example 2

The lifetime of electric bulbs for a random sample of 10 from a large consignment gave the following data:

Item	1	2	3	4	5	6	7	8	9	10
Life in ‘000 hours	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average lifetime of bulb is 4000 hours?

Example Solution:

Here $\mathrm{n} = 10, \quad \mu = 4000$

Let us assume null hypothesis that There is no significant difference between the sample mean and population mean

H_0: $\mu = 4000$ hrs

H_1: $\mu \neq 4000$ hrs (Two tailed test)

Applying t- test

Here $\bar{x} = \frac{\sum x}{n} = \frac{44}{10} = 4.4$

x	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6
$x - \bar{x}$	- 0.2	0.2	-0.5	-0.3	0.8	-0.6	-0.5	-0.1	0	1.2
$(x - \bar{x})^2$	0.04	0.04	0.25	0.09	0.64	0.36	0.25	0.01	0	1.44

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{3.12}{9}} = 0.589$$

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{4.4-4}{0.589/\sqrt{10}} = 2.123$$

As tabulated value of t at 5% level for 9 degree of freedom is

$t_{0.05} = 2.26$. As the calculated value of $|t| = 2.123 < t_{0.05}$ for 9 degree

of freedom Ho is Accepted

α = 0.05, t Value :
$\mathbf{t_{0.05} = 2.26}$

Decision: Accept Ho at α = 0.05(At 5%)

Conclusion:

The average life time of bulbs could be 4000 hrs

Example 3

An automobile tyre manufacturer claims that the average life of a particular grade of tyre is more than 20,000 km when used under normal conditions. A random sample of 16 tyres was tested and a mean and standard deviation of 22,000 km and 5,000 km, respectively were computed. Assuming the life of the tyres in km to be approximated normally distributed, decide whether the manufacturer's claim is valid

Example 4

The nine items of a sample had the following values:

45, 47, 50, 52, 48, 47, 49, 53, 51

Does the mean of nine items differ significantly from the assumed population mean of 47.5?