Small sample tests: t-distribution

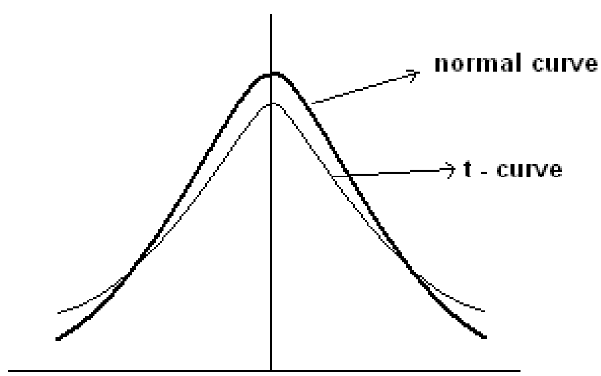
If the sample sizes are small (<30), the above tests (for large samples) do not hold good. Therefore, for estimation of the parameters as well as for testing a hypothesis, we cannot use the above methods.

Student's t-distribution

If we take a large number of samples of small (<30) sizes, calculate the mean of each sample, obtain the frequencies and obtain the frequency curve, we will find that the resulting sampling distribution of the mean is the **student's t-distribution**.

Properties of t distribution

1. The curve extends from – ∞ to + ∞
2. Like the normal distribution the t distribution is also symmetrical and has mean zero.
3. The variance of t distribution is greater than unity and approaches unity when the degrees of freedom (d.f.) i.e the size of the sample becomes large. The degress of freedom (d.f.) is given by $n-1$



t distribution curve

Uses of t distribution

Suppose it is not possible to take a sample of large size (due to lack of time or prohibitive costs), we can use the t distribution to do the 't-test' for testing hypothesis about the population mean/s and for estimating the population mean using the sample mean.

1. **Test of Significance (TOS) of the difference between sample mean and population mean:**

Sample size n $(n < 30)$

Population mean: μ

Sample mean : $\bar{x}$

Population S. D. : σ (always unknown)

Sample S. D. : s

Null Hypothesis : $H_0 : \mu = \mu_0$

Alternate Hypothesis : $H_1 : \mu \neq \mu_0$ (Two tailed)

$H_1 : \mu < \mu_0$ (One tailed (left))

$H_1 : \mu > \mu_0$ (One tailed (right))

Test Statistic

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n-1}} \sim t_{(n-1)} \ d.f$$

This t statistic is said to follow a t distribution with ν = $n - 1$ degrees of freedom. The t distribution has been derived under the hypothesis that the parent population is distributed normally.

<u>**Conclusion of the test**</u>

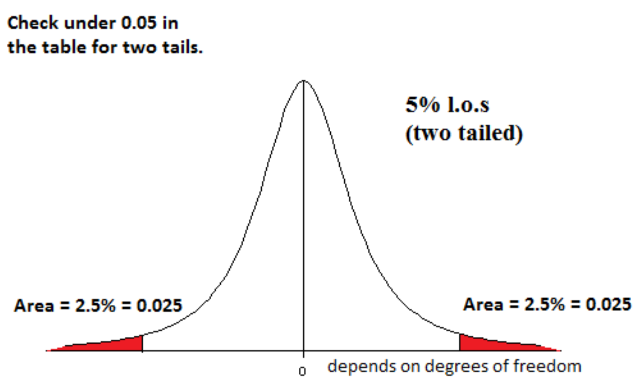
Let LOS = α

If $|t_{cal}| > t_\alpha$, Reject H_0 at α level of significance. Otherwise we have no reason to reject H_0.

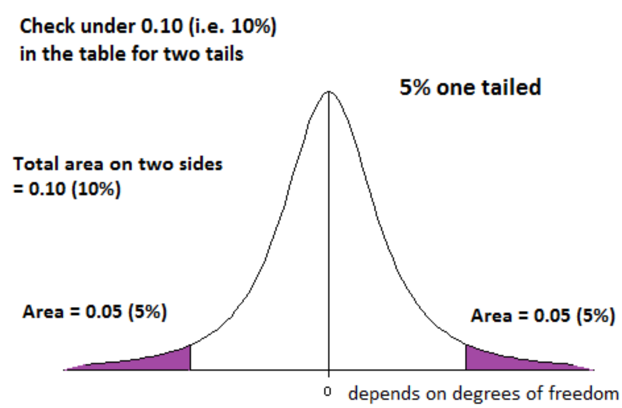
Prof. Anushri Tambe

How to read the table:

There is a different table given for t-distribution, since the curve is different, the areas will also differ from a normal distribution. The rows in the t-distribution table indicate degrees of freedom, and the columns indicate the most commonly used levels of significance. So if you wanted to find the critical value at 5% l.os.(two tailed) and with 9 degrees of freedom (i.e. the size of the sample is 10), then the value from the table is 2.26. This means that when the area in the critical region is 5% (2.5% on each side, since it is two-tailed, the critical value $Z_{\propto} = 2.26$.



Note that the table given to you indicates two-tail areas. In the event of a two-tailed test (depending on the alternative hypothesis), reading the critical values from the table is easy. However, if the test is one tailed, with 5% l.o.s. then you cannot look up the value under 5% in the table because it is the two tailed area, meaning the corresponding one-tailed area is 2.5%. However, if you look up the critical value for the two tailed area 10%, then the area is distributed evenly on both the tails (5% on each side), and the critical value corresponding to this area will be the same as the 5% area on a one-tailed test.



Prof. Anushri Tambe

Solved Examples:

1. 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 of 1.85 cm, with 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: Given $n = 10$ (small sample)

   $$\bar{x} = 1.85$$

   Level of significance: $\alpha = 5\%$

   $$H_0 : \mu = 1.75$$
   $$H_1 : \mu \neq 1.75 \; (two-tailed\, test)$$

   Under H_0, the test statistic is given by

   $$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n-1}} \sim t_{(n-1)} = t_9 \, d.f$$

   $$i.e\; t = \frac{1.85 - 1.75}{0.1/\sqrt{10-1}} = 3$$

   From the tables, $t_{9,0.05} = 2.262$

   Since $|t| > 2.262$ we reject H_0 at 5% level of significance. That is, **the work is inferior at 5% LOS.**

   Remark: But from the tables, $t_{9,0.01} = 3.250$. Since $|t| < 2.262$ we have no reason to reject H_0 at 1% level of significance. That is, **we cannot consider the work to be inferior at 1% LOS.**

Prof. Anushri Tambe

2. A certain injection administered to each of the 12 patients resulted in the following increases of blood pressure: $5,\ 2,\ 8,\ -1,\ 3,\ 6,\ 0,\ -2,\ +1,\ 5,\ 0,\ 4$

Can it be concluded that the injection will be, in general,

accompanied by an increase in BP?

Solution: We have n=12 (small sample).

The mean of the sample is the average increase in B.P.

$$\bar{x} = \frac{1}{n}\sum x = \frac{1}{12}(31) = 2.58$$

The S.D of the sample is given by

$$s^2 = \frac{1}{n}\sum x^2 - \left(\frac{1}{n}\sum x\right)^2 = \frac{1}{12}(185) - (2.58)^2 = 8.76$$

$$\Rightarrow s = 2.96$$

Level of significance: $\alpha = 5\%$

$H_0 : \mu = 0$ (there is no increase in B.P)

$H_1 : \mu > 0$ $(one-tailed\ test)$

Under H_0, the test statistic is given by

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n-1}} \sim t_{(n-1)} = t_{11}\, d.f$$

$$i.e\ t = \frac{2.58 - 0}{2.96/\sqrt{11}} = 2.89$$

From the tables, $t_{11, 0.05}\ (one-tailed) = 1.796$

Since $|t| > 1.796$ we reject H_0 at 5% level of significance. And conclude that, **the injection will, in general, result in increase in B.P**

Exercise: The annual rainfall at a certain place is normally distributed with mean 30mm. If the rainfalls during the past 8 years (in mm) are $31.1,\ 30.7,\ 24.3,\ 28.1,\ 27.9,\ 32.2,\ 25.4$ and 29.1, can we conclude that the average rainfall during the last 8 years is less than the normal rainfall?

Prof. Anushri Tambe