SMALL SAMPLE TEST-1

Prof. Nandini Rai
KJSCE

TEST OF SIGNIFICANCE FOR SMALL SAMPLES

If the samples are large (≥ 30) then the sampling distribution of a statistic is normal. But if the samples are small (< 30) then the above result does not hold good and for estimation of the parameter as well as for testing a hypothesis we cannot use the methods used for Large samples.

TEST OF SIGNIFICANCE FOR SMALL SAMPLES

If we take a large number of samples of small (< 30) size, calculate the mean of each samples, plot the frequencies and obtain the frequency curve we will find that the resulting sampling distribution of the mean is not normal but is the student's t – distribution.

STUDENT'S T- DISTRIBUTION

Theoretical work on t – distribution was done by Irish Statistician W.S. Gosset. W.S. Gosset was working with Guinness Brewery in Dublin which did not allow its employees to publish their research work under their own names. So he adopted the pen name "Student" and published his research work under that name in early period of 20th century. Hence, this distribution is known as **Student's *t* - distribution** or simply ***t* - distribution**.

The t – distribution is used when

(i) the sample size is 30 or less and

(ii) Population standard deviation is not known.

The “t – statistic” is defined as,

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \quad where, S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

The curve is given by

$$y = C\left(1 + \frac{t^2}{v}\right)^{-\frac{(v+1)}{2}}$$

$$where, t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$C =$ constant required to make the area under the curve unity.

$v = n - 1$, the number of degrees of freedom.

- Then t – distribution has been derived mathematically **under the hypothesis that parent population is distributed normally.**

PROPERTIES OF T - DISTRIBUTION

As the normal curve, this curve also extends from $-\infty \; to + \infty$.

The constant C depends upon the v (pronounces as ‘nu’), degrees of freedom.

Like the normal distribution, the t – distribution also is symmetrical and has a mean zero.

The variance of t – distribution is greater than unity and approaches unity as the number of degrees of freedom and therefore the size of the sample becomes large.

- Curve

ASSUMPTIONS FOR T - TEST

Sample are drawn from **normal population** and they are random.

For testing the equality of means of two populations their **variances are assumed to be equal.**

T - TABLE

An interval estimate of the population mean is given by $\bar{x} \pm t\ \hat{\sigma}_{\bar{x}}$.

It should be noted that t table gives the probability that population parameter ***will not lie*** within the desired confidence interval i.e the parameter will lie outside the confidence interval.

T - TABLE

For making an estimate, say, at 95% confidence level we consult the column under the head 0.05 (100% - 95% = 5% = 0.05) of the t – table.
t-distribution table

Similarly for 98%, 99% confidence level we consult the column under the head 0.02 and 0.01.

DISTRIBUTION OF SAMPLE MEAN

If $\bar{x}$ is the sample mean and μ is the population mean then

$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ where $S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$

follows Student's t – distribution with $n-1$ degrees of freedom.

DEGREE OF FREEDOM

Degree of freedom means the number of values we are free to choose.

Suppose the sum of three numbers is 15, How many numbers we are free to choose such that the sum is 15?

Certainly not all the three.

We can choose two numbers at our will but the third will be given by 15 – (sum of the two chosen numbers).

Thus, we are free to choose only two numbers. Hence, the degrees of freedom here are two.

USES OF T - DISTRIBUTION

The t – distribution has a wide number of applications. Some important of them are:

- To estimate the population mean μ from the sample mean $\bar{x}$.
- To test the hypothesis that the population mean is μ with the help of the sample mean $\bar{x}$.
- To test the hypothesis that two populations have the same mean with the help of sample means.

Remark: We know that the sample variance is given by $s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$

and an unbiased estimates S of standard deviation σ is given by

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$\therefore ns^2 = \sum(x_i - \bar{x})^2 = (n-1)S^2$$

$$\therefore \frac{S^2}{n} = \frac{s^2}{(n-1)} \quad \therefore \frac{S}{\sqrt{n}} = \frac{s}{\sqrt{n-1}}$$

Hence, $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ ....(1)

OR $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$......................(2)

FOR SMALL SAMPLES

If the standard deviation σ of the parent population is known, then $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is a S.N.V.

FOR SMALL SAMPLES

- If the standard deviation of the parent population is not known and
- the parent population is normal and
- when sample values $x_1, x_2, \ldots \ldots \ldots \ldots . x_n$ are given or
- an unbiased estimates S of standard deviation σ is given, then
- $t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ is a t – distribution where

$$S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$$

FOR SMALL SAMPLES

- If the standard deviation of the parent population is not known and
- if the parent population is normal and
- sample standard deviation s is given, then
- $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$ is a t- distribution where

$$s^2 = \frac{\sum(x_i-\bar{x})^2}{n}$$

TESTING THE HYPOTHESIS THAT THE POPULATION MEAN IS μ

- To test the hypothesis that the population mean is μ when the sample is small we follow the steps as for large sample but use the t – distribution instead of normal distribution

EXAMPLE-1

- A sample of size 10 has mean 40 and standard deviation 10. Construct, the 99% confidence interval for the population mean.
- SOLUTION:
- **Step 1:** For 99% confidence level and for $10 - 1 = 9$ degrees of freedom, the column under head 0.01 $(100\% - 99\% = 1\% = 0.01)$ gives the critical value $t_\alpha = 3.25$

t-distribution table

Step 2: Since the population standard deviation is not know and sample is small

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n-1}} = \frac{10}{\sqrt{10-1}} = \frac{10}{3} = 3.33$$

Step 3: The confidence interval is $\bar{x} \pm t_\alpha \hat{\sigma}_{\bar{x}}$

i.e. $40 \pm 3.25(3.33)$

i.e. 40 ± 10.82

$\therefore$ 99% confidence interval $[29.18, 50.82]$

EXAMPLE - 2

- A sample of size 12 from a normal population gave $\bar{x} = 15.8$ and ${s_x}^2 = 10.3$. Find 99% interval of population mean .
- SOLUTION:
- **Step 1:** For 99% confidence level and for $12 - 1 =$11 degrees of freedom, the column under 0.01 gives the critical value
- $t_\alpha = 3.106$

- **Step 2:** Since the population standard deviation is not know and sample is small
- $\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n-1}}$
- $= 0.9676$
- **Step 3:** The confidence interval is $\bar{x} \pm t_{\alpha}\hat{\sigma}_{\bar{x}}$
- $\therefore$ 99% confidence interval $[12.795, 18.805]$

EXAMPLE-3

- A random sample of 16 values from a normal population showed a mean of 41.6 inches and the sum of the squares of deviations from this mean equal to 135. Obtain 95% and 99% fiducial limits for the mean.

EXAMPLE-3

- **Step 1:** For 95% confidence level and for $16-1=15$ degrees of freedom, the column under $0.05(100\%-95\%=5\%=0.05)$ gives the critical value $t_\alpha = 2.131$
- For 99% confidence level, for 15 degrees of freedom the column under 0.01 gives the critical value $t_\alpha = 2.947$.

- **Step 2:** since the sample is small,
- $s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n}} = \sqrt{\frac{135}{16}} = 2.9047$
- $\therefore\ \hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n-1}} = 0.75$
- **Step 3:** The confidence interval is $\bar{x} \pm t_\alpha \hat{\sigma}_{\bar{x}}$
- $\therefore$ **95% confidence interval** $[\mathbf{40.002, 43.198}]$
- $\therefore$ **99% confidence interval** $[\mathbf{39.39, 43.81}]$

EXAMPLE-4

- A random sample of size 16 from a normal population showed a mean of 103.75 cm and sum of squares of deviations from the mean $843.75\ \mathrm{cm}^2$. Can we say that the population has a mean of 108.75 cm?
- SOLUTION:
- We first calculate sample standard deviation

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n} = \frac{843.75}{16} = 52.73$$

- **The null hypothesis** $H_0: \mu = 108.75$
- **Alternative Hypothesis** $H_a: \mu \neq 108.75$
- **Calculation of test statistic:**
- $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} = -2.67$
- $\therefore\ |t| = 2.67$

- **Level of significance:** $\alpha = 0.05$

- **Critical value**: The value of t_α for 5% level of significance and degree of freedom $v = 16 - 1 = 15$ from the table is 2.131
- **Decision**: Since the computed value of $|t| = 2.67$ is greater than the table value $t_\alpha = 2.131$, the null hypothesis is rejected
- $\therefore$ **We cannot say that the population mean is $\mathbf{108.75}$**

EXAMPLE-5

- 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 sale was found to be 147 dozens with standard deviation of 16. Can you consider the advertisement effective?

SOLUTION

- **The null Hypothesis** $H_0: \mu = 140$
- **Alternative Hypothesis** $H_a: \mu > 140$
- **Calculation of test statistic**: Since the sample is small and SD of the population is not know, we use t −distribution
- $t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} = \frac{147-140}{16/\sqrt{26-1}} = \frac{7}{3.2} = 2.19$
- **Level of significance**: $\alpha = 0.05$

- **Critical value**: The value of t_α for 5% level of significance and degree of freedom $v = 26 - 1 = 25$ from the table is
- $t = 1.708$
- **Decision**: Since the computed value of $|t| = 2.19$ is greater than the critical value $t_\alpha = 1.708$, the null hypothesis is rejected
- $\therefore$ **The advertisement have increased the average sales.**

EXAMPLE-6

- Nine items of a sample had the following values $45, 47, 50, 52, 48, 47, 49, 53, 51$. Does the mean of 9 items differ significantly from the assumed population mean 47.5?
- SOLUTION:
- We first calculate sample standard deviation
- $s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$

x_i									
$x_i - \overline{x}$									
$(x_i - \overline{x})^2$									

- $s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$ = 6.099
- **The null hypothesis** $\boldsymbol{H_0}: \mu = 47.5$
- **Alternative Hypothesis** $\boldsymbol{H_a}: \mu \neq 47.5$
- **Calculation of test statistic:** $t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$
- $\therefore |t| = 1.84$
- **Level of significance:** $\alpha = 0.05$

- **Critical value**: The value of t_α for 5% level of significance for $v = 9 - 1 = 8$ degree of freedom is 2.306
- **Decision**: Since the calculated value of $|t| = 1.84$ is less than the table value $t_\alpha = 2.306$, the null hypothesis is accepted
- **The mean of nine items does not differ significantly from assumed population mean $\mathbf{47.5}$**