t-test

Population: refer to any collection of individuals or of their attributes or of results of operations which can be numerically specified.

Example: height, age, prices of rice, life of electric bulbs etc.,

A population may be finite or infinite elements.

Example:

(a) Marks of students in a class will constitute a finite population

(b) Pressure at every point of the atmosphere will constitute an infinite population

Sample: A finite subset of a population is a sample and the number of elements of a sample is the size of the sample and the process of selection of such samples is called sampling.

Parameters and Statistics: The statistical constants of the population such as the mean, the variance etc., are known as parameters. The statistical concepts of the sample computed from the members or observations of the sample to estimate the parameters of the population from which the sample has been drawn, are known as statistics.

Statistical inference:

Confidence intervals are one of the two most common types of statistical inference. Researchers use a confidence interval when their goal is to estimate a population parameter. The second common type of inference, called a test of significance, which has a different goal: to access the evidence provided by data about some claim concerning a population.

A **test of significance** is a formal procedure for comparing observed data with a claim (also called a hypothesis), the truth of which is being assessed. The claim is a statement about a parameter, like the population proportion p or the population mean μ.

The first step while conducting the test of significance is to state the hypothesis.

Null Hypothesis: The claim test by the statistical test is called null hypothesis or it is a statement of no difference.

Alternative Hypothesis: Negation of null hypothesis statement is the alternative hypothesis.

This alternative hypothesis is one-sided(single-tailed), if it states the parameter is larger or smaller than the null hypothesis value.

This alternative hypothesis is two-sided if it that the parameter is different from the null- hypothesis value.

Note: The test presents 4 – step process.

1. What is the practical question that requires a statistical test.?
2. Formulate by identifying the parameter and then you state the null and alternative hypothesis.
3. Solve: Carry out the test in three phases.
   (i) Check the conditions for the test you plan to use.
   (ii) Calculate the test statistic.
   (iii) Find the p-value.
4. Conclude: Return to the practical question to describe your result in this setting.

The t-test of significance (for small samples) Student's t:

(n= sample size <= 30)

Let $x_1,x_2,x_3,\ldots x_n$ be a random sample of sixe n from a normal population with mean μ and variance σ^2. The student's t test is defined in the statistics as

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

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$

the unbiased estimate of the population variance σ^{2}, and t follows the student t distribution with $\nu = n-1$ degrees of freedom.

TEST 1: t- TEST OF SIGNIFICANCE FOR SINGLE MEAN:

Under the null hypothesis H_0: The sample has been drawn from the population mean μ or there is no significant diffierence between the sample mean $\bar{x}$ and the population mean μ and the test statistics is

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \quad or \quad t = \frac{\bar{x} - \mu}{S/\sqrt{(n-1)}} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$ where

with degrees of freedom (n-1) . At given level of significance α, and the degrees of freedom (n-1) we refer the t-table value t_α (two- tailed or one- tailed test) If $|t| < t_\alpha$, the null hypothesis is accepted. If $|t| > t_\alpha$ H_0 is rejected

Applications of t-distribution:

(i) Sample means $\bar{x}$ differ from the population mean μ.

(ii) To find difference between two sample means.

(iii) Observed correlation coefficient and sample regression coefficient.

(iv) Observed multiple correlation coefficient and regression coefficient.

Problem 1: Sandal powder is packed into packets by a machine. A random sample of 12 packets is drawn and their weights are found to be (in kgs) 0.49,0.48,0.47, 0.49, 0.48, 0.50, 0.51, 0.49, 0.48, 0.50, 0.51, 0.48. Test if the average packing can be taken as 0.5 kg.

Null Hypothesis - H_0: $\mu = 0.5$

Alternative Hypothesis –H_1: $\mu \neq 0.5$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5.88}{12} = 0.49$$

$$s^2 = \frac{\sum x^2}{n} - \bar{x}^2 = [0.2401 \times 3 + 4 \times 0.2304 + 0.2209 + 2 \times 0.25 + 2 \times 0.2601] \times \left(\frac{1}{12}\right) - (0.49)^2$$

$$= 0.00015$$

$$s = 0.012$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n-1}} = \frac{0.49 - 0.5}{0.012 / \sqrt{11}} = 2.7633$$

Now ν=d.f = n-1 = 11.

At 5% level of significance for ν=11, t value from table is 2.20

Since the calculated value of t is greater than the table value, i.e., 2.7633 > 2.20, H_0 is rejected. That is average packing cannot be taken as 0.5 kg.

2. The average breaking strength of steel rods is specified to be 17.5 (in units of 1000 kg). To test this, of 14 rods tested and gave the following results:15, 18, 16, 21, 19, 21, 17,17, 15, 17, 20, 19, 17, 18. Is the result of the experiment is significant? Also obtain the 95% confidence interval for the average breaking strength.

H_0: Average breaking strength is not significant.

H_1: Average breaking strength is significant.

Now ν=d.f = n-1 = 13.

μ=17.5

$$\bar{x} = \frac{\sum x_i}{n} = \frac{250}{14} = 17.857142$$

$$s^2 = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{4514}{14} - 17.857142^2 = 3.5510514$$

$$s = 1.8844233$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n-1}} = \frac{17.857142 - 17.5}{0.5226449} = 0.6833358$$

At 5% level of significance for ν=13, t value from table is 2.16

The Calculated value is less than the table value. Therefore the difference is not significant.

95% confidence limits are , $\bar{x} - 2.16 \times \frac{s}{\sqrt{n-1}} < \mu < \bar{x} + 2.16 \times \frac{s}{\sqrt{n-1}}$

$$17.857142 - 2.16 \times 0.5226449 < \mu < 17.857142 + 2.16 \times 0.5226449$$

$16.728 < \mu < 18.986$.

3. The mean weekly sales of a powder in Super market was 146.3 tins. After a special advertising campaign the mean weekly sales in22 branches in a week increased to 153.7 tins and showed a S.D of 17.2 tins. Was the advertising campaign successful?

H_0: μ=146.3 (The campaign was not successful)

H_1: μ > 146.3 (right tailed)

$$\bar{x} = 153.7$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n-1}} = \frac{153.7 - 146.3}{17.2 / \sqrt{21}} = 1.9715$$

At 5% level of significance for $\nu=21$, t value from table is 1.72(single-tailed)

Calculated t value > table value of t

Reject H0. Hence the campaign was successful in promoting sales.

Test –II t- test for difference of means of two samples from a normal population:

Suppose we want to test if two independent samples $x_1, x_2, x_3, \ldots x_{n1}$ and $y_1, y_2, y_3 \ldots y_{n2}$ of sizes n_1 and n_2 have been drawn from two normal populations with mean μ_1 and μ_2 respectively.

Then the test statistic is given by

7

is an unbiased estimated of the population variance σ^2