**Bio-instrumentation & Biostatistics**

**Dr. Bibha Verma**
**PG Department of Zoology**
**Maharaja College Ara**

**PG Sem II, CC-6, Unit -IV**

# Student's t-test

## Introduction

Biostatistics is contraction of biology and statistics, sometimes referred to as biometry or biometrics, is the application of statistics to a wide range of topics in biology. The science of biostatistics encompasses the design of biological experiments, especially in medicine and agriculture; the collection, summarization, and analysis of data from those experiments; and the interpretation of, and inference from, the results. Hypothesis testing determines the validity of the assumption with a view to choose between two conflicting hypotheses about the value of a population parameter. Hypothesis testing helps to decide on the basis of a sample data, whether a hypothesis about the population is likely to be true or false. Statisticians have developed several tests of hypotheses (also known as the tests of significance) for the purpose of testing of hypotheses which can be classified as:- (a) Parametric tests or standard tests of hypotheses, and (b) Non-parametric tests or distribution-free test of hypotheses. Parametric tests usually assume certain properties of the parent population from which we draw samples. Assumptions like observations come from a normal population, sample size is large, assumptions about the population parameters like mean, variance, etc., must hold good before parametric tests can be used. But there are situations when the researcher cannot or does not want to make such assumptions. In such situations we use statistical methods for testing hypotheses which are called non-parametric tests because such tests do not depend on any assumption about the parameters of the parent population. Besides, most non-parametric tests assume only nominal or ordinal data, whereas parametric tests require measurement equivalent to at least an interval scale.

**Student's t-test**, in statistics, a method of testing hypotheses about the mean of a small sample drawn from a normally distributed population when the population standard deviation is unknown.

In 1908 William Sealy Gosset, an Englishman publishing under the pseudonym Student, developed the *t*-test and *t* distribution. (Gosset worked at the Guinness brewery in Dublin and found that existing statistical techniques using large samples were not useful for the small sample sizes that he encountered in his work.) The *t* distribution is a family of curves in which the number of degrees of freedom (the number of independent observations in the sample minus one) specifies a particular curve. As the sample size (and thus the degrees of freedom) increases, the *t* distribution approaches the bell shape of the standard normal distribution. In practice, for tests involving the mean of a sample of size greater than 30, the normal distribution is usually applied.

It is usual first to formulate a null hypothesis, which states that there is no effective difference between the observed sample mean and the hypothesized or stated population mean—i.e., that any measured difference is due only to chance. In an agricultural study, for example, the null hypothesis could be that an application of fertilizer has had no effect on crop yield, and an experiment would be performed to test whether it has increased the harvest. In general, a *t*-test may be either two-sided (also termed two-tailed), stating simply that the means are not equivalent, or one-sided, specifying whether the observed mean is larger or smaller than the hypothesized mean. The test statistic *t* is then calculated. If the observed *t*-statistic is more extreme than the critical value determined by the appropriate reference distribution, the null hypothesis is rejected. The appropriate reference distribution for the *t*-statistic is the *t* distribution. The critical value depends on the significance level of the test (the probability of erroneously rejecting the null hypothesis).

For example, suppose a researcher wishes to test the hypothesis that a sample of size $n = 25$ with mean $\bar{x} = 79$ and standard deviation $s = 10$ was drawn at random from a population with mean $\mu = 75$ and unknown standard deviation. Using the formula for the *t*-statistic, $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ the calculated *t* equals 2. For a two-sided test at a common level of significance $\alpha = 0.05$, the critical values from the *t* distribution on 24 degrees of freedom are $-2.064$ and $2.064$. The calculated *t* does not exceed these values, hence the null hypothesis cannot be rejected with 95 percent confidence. (The confidence level is $1 - \alpha$.)

A second application of the *t* distribution tests the hypothesis that two independent random samples have the same mean. The *t* distribution can also be used to construct confidence intervals for the true mean of a population (the first application) or for the difference between two sample means (the second application).

**Use of t-test**

In medical research, t-test is among the three or four most commonly use statistical tests. The purpose of a t-test is to compare the means of a continuous variable in two research samples in orders to determine whether or not the difference between the two observed means exceeds the difference that would be expected by chance from random samples.

**Some important terms used in student t –test**

**Degree of Freedom**

The term degree of freedom refers to the number of observations that are free to vary. The degrees of freedom for any tests are considered to be the total sample size minus 1 degree of freedom for each mean that is calculated. In the student's t-test, 2 degree of freedom are lost because two means are calculated one mean for each group whose means are to be composed. The general formula for the degrees of freedom for the student's two group t-test is $N_1 + N_2 - 2$, where $N_1$ is the sample size in the first group and $N_2$ is the sample size in the second group. (Jakel James et.al. 2001, Armstrong N et.al. 2006)

**The t- distribution**

The t-distribution was described by William Goasset, who used the pseudonym "student" when he wrote the description. The normal distribution is the z distribution. The t distribution looks similar to it, except that its tails are somewhat wider and its peak is slightly less high, depending on the sample size. The t-distribution is necessary because when sample size. The t distribution is necessary because when sample sizes are small, the observed estimates of mean and variance are subject to considerable error. The larger the sample size is the smaller the errors are and the more the t size distribution looks like the normal distribution. In the case of an infinite sample size, the two distributions are identical. For practical purposes, when the combined sample size of the two groups being compared is larger than 120, the difference between the normal distribution and the t distribution is negligible. (Jakel James et.al. 2001, Le Chap T 1992)

**References:**


https://www.pharmatutor.org/articles/review-student-t-test

https://www.statisticshowto.com/probability-and-statistics/t-test/

https://www.britannica.com/science/Students-t-test

https://www.investopedia.com/terms/t/t-test.asp