Student's t - Test

Course: M.A. Geography (Sem.-3)

Paper: M-301(CC-10)

Quantitative Techniques and Research Methodology

By

Dr. Md. Nazim

Professor, Department of Geography

Patna College, Patna University

Contact No- 8409110509

Email – dr.nazim2011@gmail.com

Lecture- 4

Concept:

Student's t–test in statistics is a method of testing the hypothesis about the mean of a small sample drawn from a normally distributed population when the population standard deviation is unknown. This distribution was developed by an Irish statistician W.S.Gosset .The distribution developed by Gosset written under his pen name Student and has come to be known as Student's t- distribution in 1908.This distribution should be used when the sample size is less than 30.He found that the existing statistical techniques using large samples are 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 degree of freedom) increases, 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 applied.

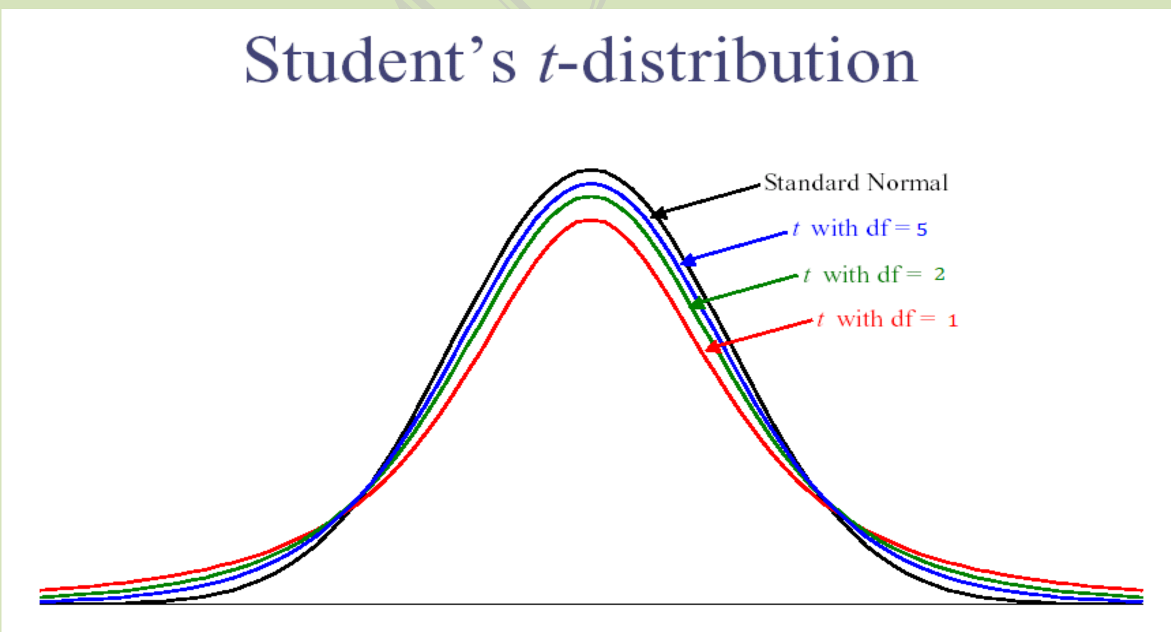
Usually the first step is to formulate a null hypothesis, which states that there is no effective difference between the observed sample mean and the hypothesized or stated population mean – i.e. any measured difference is due only to chance. For an example- in an agricultural study the null hypothesis could be that an application of fertilizer has had no effect on the yield of crop, 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 as 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 is then calculated. I f the observed t –statistic is more extreme than the critical value determined by appropriate reference distribution, the null hypothesis is rejected. The appropriate reference distribution for t-statistic is 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* x *= 79 and standard deviation* s = *10 was drawn at random from a population with mean μ = 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 α = 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 − α.)*

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).*



William Sealy Gosset

developed "t-statistic" and published it under the pseudonym of "Student".

History:

The term "t-statistic" is abbreviated from "hypothesis test statistic". In statistics, the t-distribution was first derived as a posterior distribution in 1876 by Helmert and Lüroth. The t-distribution also appeared in a more general form as Pearson Type IV distribution in Karl Pearson's 1895 paper. However, the T-Distribution, also known as Student's T Distribution gets its name from William Sealy Gosset who first published it in English literature in his 1908 paper titled Biometrika using his pseudonym "Student" because his employer preferred staff to use pen names when publishing scientific papers instead of their real name, so he used the name "Student" to hide his identity. Gosset worked at the Guinness Brewery in Dublin, Ireland, and was interested in the problems of small samples – for example, the chemical properties of barley with small sample sizes. Hence a second version of the etymology of the term Student is that Guinness did not want their competitors to know that they were using the t-test to determine the quality of raw material. Although it was William Gosset after whom the term "Student" is penned, it was actually through the work of Ronald Fisher that the distribution became well known as "Student's distribution" and "Student's t-test".

Gosset had been hired owing to Claude Guinness's policy of recruiting the best graduates from Oxford and Cambridge to apply biochemistry and statistics to Guinness's industrial processes. Gosset devised the t-test as an economical way to monitor the quality of stout. The t-test work was submitted to and accepted in the journal Biometrika and published in 1908. Company policy at Guinness forbade its chemists from publishing their findings, so Gosset published his statistical work under the pseudonym "Student" (see Student's t-distribution for a detailed history of this pseudonym, which is not to be confused with the literal term student).

Guinness had a policy of allowing technical staff leave for study (so-called "study leave"), which Gosset used during the first two terms of the 1906–1907 academic year in Professor Karl Pearson's Biometric Laboratory at University College London. Gosset's identity was then known to fellow statisticians and to editor-in-chief, Karl Pearson.

Uses:

Among the most frequently used t-tests are:

- A one-sample location test of whether the mean of a population has a value specified in a null hypothesis.
- A two-sample location test of the null hypothesis such that the means of two populations are equal. All such tests are usually called **Student's t-tests**, though strictly speaking that name should only be used if the variances of the two populations are also assumed to be equal; the form of the test used when this assumption is dropped is sometimes called Welch's t-test. These tests are often referred to as "unpaired" or "independent samples" t-tests, as they are typically applied when the statistical units underlying the two samples being compared are non-overlapping.

Assumptions:

Most test statistics have the form $t = Z/s$, where Z and s are functions of the data.

Z may be sensitive to the alternative hypothesis (i.e., its magnitude tends to be larger when the alternative hypothesis is true), whereas s is a scaling parameter that allows the distribution of t to be determined.

As an example, in the one-sample t-test

where X is the sample mean from a sample X_1, X_2, …, X_n, of size n, s is the standard error of the mean,   is the estimate of the standard deviation of the population, and μ is the population mean.

The assumptions underlying a t-test in its simplest form are that

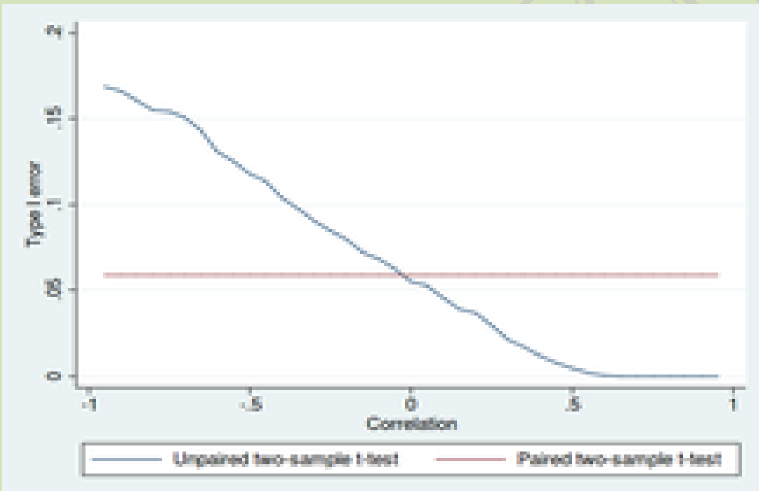
- *X follows a normal distribution with mean μ and variance σ^2/n*
- *$s^2 (n-1)/\sigma^2$ follows a χ^2 distribution with n − 1 degrees of freedom. This assumption is met when the observations used for estimating s^2 come from a normal distribution (and i.i.d for each group).*
- *Z and s are independent.*

In the t-test comparing the means of two independent samples, the following assumptions should be met:

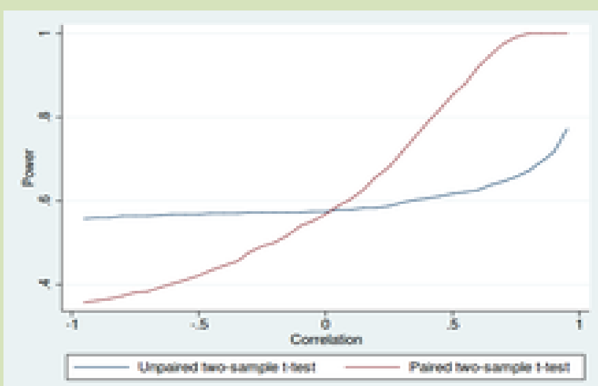
- *The means of the two populations being compared should follow normal distributions. Under weak assumptions, this follows in large samples from the central limit theorem, even when the distribution of observations in each group is non-normal.*
- *If using Student's original definition of the t-test, the two populations being compared should have the same variance (testable using F-test, Levene's test, Bartlett's test, or the Brown–Forsythe test; or assessable graphically using a Q–Q plot). If the sample sizes in the two groups being compared are equal, Student's original t-test is highly robust to the presence of unequal variances. Welch's t-test is insensitive to equality of the variances regardless of whether the sample sizes are similar.*
- *The data used to carry out the test should either be sampled independently from the two populations being compared or be fully paired. This is in general not testable from the data, but if the data are known to be dependent (e.g. paired by test design), a dependent test has to be applied. For partially paired data, the classical independent t-tests may give invalid results as the test statistic might not follow a t distribution, while the dependent t-test is sub-optimal as it discards the unpaired data.*

Most two-sample t-tests are robust to all but large deviations from the assumptions. For exactness, the t-test and Z-test require normality of the sample means, and the t-test additionally requires that the sample variance *follows a scaled [χ² distribution](), and that the sample mean and sample variance be [statistically independent](). Normality of the individual data values is not required if these conditions are met. By the [central limit theorem](), sample means of moderately large samples are often well-approximated by a normal distribution even if the data are not normally distributed. For non-normal data, the distribution of the sample variance may deviate substantially from a χ^2 distribution. However, if the sample size is large, [Slutsky's theorem]() implies that the distribution of the sample variance has little effect on the distribution of the test statistic.*

Unpaired and paired two-sample t-tests:



Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1. The significance level is 5% and the number of cases is 60.



Power of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bi-variate normal distribution with a variance of 1 and a deviation of the expected value of 0.4. The significance level is 5% and the number of cases is 60.

Two-sample t-tests for a difference in mean involve independent samples (unpaired samples) or paired samples. Paired t-tests are a form of blocking, and have greater power than unpaired tests when the paired units are similar with respect to "noise factors" that are independent of membership in the two groups being compared. In a different context, paired t-tests can be used to reduce the effects of confounding factors in an observational study.

Independent (unpaired) samples:

The independent samples t-test is used when two separate sets of independent and identically distributed samples are obtained, one from each of the two populations being compared. For example, suppose we are evaluating the effect of a medical treatment and we enroll 100 subjects into our study, then randomly assign 50 subjects to the treatment group and 50 subjects to the control group. In this case, we have two independent samples and would use the unpaired form of the t-test.

Paired samples:

Paired samples t-tests typically consist of a sample of matched pairs of similar units, or one group of units that has been tested twice (a "repeated measures" t-test).

A typical example of the repeated measures t-test would be where subjects are tested prior to a treatment, say for high blood pressure, and the same subjects are tested again after treatment with a blood-pressure-lowering medication. By comparing the same patient's numbers before and after treatment, we are effectively using each patient as their own control. That way the correct rejection of the null hypothesis (here: of no difference made by the treatment) can become much more likely, with statistical power increasing simply because the random inter-patient variation has now been eliminated. However, an increase of statistical power comes at a price: more tests are required, each subject having to be tested twice. Because half of the sample now depends on the other half, the paired *version of Student's t-test has only n/2 − 1 degrees of freedom (with n being the total number of observations). Pairs become individual test units, and the sample has to be doubled to achieve the same number of degrees of freedom. Normally, there are n − 1 degrees of freedom (with n being the total number of observations).*

A paired samples t-test based on a "matched-pairs sample" results from an unpaired sample that is subsequently used to form a paired sample, by using additional variables that were measured along with the variable of interest. The matching is carried out by identifying pairs of values consisting of one observation from each of the two samples, where the pair is similar in terms of other measured variables. This approach is sometimes used in observational studies to reduce or eliminate the effects of confounding factors.

Paired samples t-tests are often referred to as "dependent samples t-tests".

Calculations:

Explicit expressions that can be used to carry out various t-tests are given below. In each case, the formula for a test statistic that either exactly follows or closely approximates a t-distribution under the null hypothesis is given. Also, the appropriate degrees of freedom are given in each case. Each of these statistics can be used to carry out either a one-tailed or two-tailed test.

Once the t value and degrees of freedom are determined, a p-value can be found using a table of values from Student's t-distribution. If the calculated p-value is below the threshold chosen for statistical significance (usually the 0.10, the 0.05, or 0.01 levels), then the null hypothesis is rejected in favor of the alternative hypothesis.