Student's t-test:

The t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis.

A t-test is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistics (under certain conditions) follow a Student's t distribution. The t-test can be used, for example, to determine if two sets of data are significantly different from each other.

The t-statistic was introduced in 1908 by William Sealy Gosset, a chemist working for the Guinness brewery in Dublin, Ireland. "Student" was his pen name. 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.

Student's t-test Uses:

Among the most frequently used t-tests are:

1. one-sample location test of whether the mean of a population has a value specified in a null hypothesis.

2. 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.

3. A test of the null hypothesis that the difference between two responses measured on the same statistical unit has a mean value of zero. For example, suppose we measure the size of a cancer patient's tumor before and after a treatment. If the treatment is effective, we expect the tumor size for many of the patients to be smaller following the treatment. This is often referred to as the "paired" or "repeated measures" t-test.

4. A test of whether the slope of a regression line differs significantly from 0.

Assumptions

Most t-test statistics have the form $t = \frac{Z}{s}$, where Z and s are functions of the data. Typically, Z is designed to 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: $t = \frac{Z}{s} = {(\bar{X} - \mu)}\Big/\left(\frac{\sigma}{\sqrt{n}}\right)$

where $\bar{X}$ is the sample mean from a sample $X_1, X_2, \ldots, X_3$, of size n, s is the standard error of the mean, σ is the population standard deviation of the data, and μ is the population mean.

Unpaired and paired two-sample t-tests

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.

1.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. The randomization is not essential here – if we contacted 100 people by phone and obtained each person's age and gender, and then used a two-sample t-test to see whether the mean ages differ by gender, this would also be an independent samples t-test, even though the data are observational.

Example: use t-test between unpaired W_i and Z_i , where:

W_i	3	5	44	82	77	61	30	71	9	55
Z_i	17	26	11	31	5	20	15	12	2	4

$t_{0.05} = 1.833 \; and \; t_{0.01} = 2.821$?

Solve

$H_O : W_i = Z_i$

$H_A : W_i \neq Z_i$

Rank	W_i	W_i^2	Z_i	Z_i^2
1	3	9	17	289
2	5	25	26	676
3	44	1936	11	121
4	82	6724	31	961
5	77	5929	5	25
6	61	3721	20	400
7	30	900	15	225
8	71	5041	12	144
9	9	81	2	4
10	55	3025	4	16
$\sum$	437	27391	143	2861

$\bar{W} = \frac{\sum_{i=1}^{10} W_i}{n} = \frac{437}{10} = 43.7$, $\bar{Z} = \frac{\sum_{i=1}^{10} Z_i}{n} = \frac{143}{10} = 14.3$

$c.f._W = \frac{\left(\sum_{i=1}^{10} W_i\right)^2}{n} = \frac{(437)^2}{10} = 19096.9$

$$c.f._Z = \frac{\left(\sum_{i=1}^{10} Z_i\right)^2}{n} = \frac{(143)^2}{10} = 2044.9$$

$$S_W^2 = W_i^2 - c.f._W = 27391 - 19096.9 = 8294.1$$

$$S_Z^2 = Z_i^2 - c.f._Z = 2861 - 2044.9 = 816.1$$

$$d.f._W = n - 1 = 10 - 1 = 9 \,, d.f._Z = n - 1 = 10 - 1 = 9$$

$$Pooled\ S^2 = \frac{S_W^2 + S_Z^2}{d.f._W + d.f._F} = \frac{8294.1 + 816.1}{9 + 9} = \frac{9110.2}{18} = 506.12$$

$$S_{\bar{W}-\bar{Z}} = \sqrt{\frac{2S^2}{n}} = \sqrt{\frac{2 * 506.12}{10}} = \sqrt{101.2244} = 10.061$$

$$t = \frac{(\bar{W} - \bar{Z})}{S_{\bar{W}-\bar{Z}}} = \frac{(43.7 - 14.3)}{10.061} = 2.922^{**}$$

$$\therefore\ accept\ H_A : W_i \neq Z_i$$

2.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 between-patient variation has now been eliminated.

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"

Example: use t-test between paired x_i (before) and y_i (after), where:

before	120	124	130	127	126	140	135	150	134	127
after	128	131	129	137	130	143	136	156	133	135

Solve

$H_o : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Rank	before	after	Difference	Difference ^{2}
1	120	128	8	64
2	124	131	7	49
3	130	129	- 1	1
4	127	137	10	100
5	126	130	4	16
6	140	143	3	9
7	135	136	1	1
8	150	156	6	36
9	134	133	- 1	1
10	127	135	8	64
Σ	1313	1358	45	341

$$\frac{1}{2} * \propto = \frac{1}{2} * \ 0.05 = 0.025$$

Or

$$\frac{1}{2} * \propto = \frac{1}{2} * \ 0.01 = 0.005$$

d.f. = 10 – 1 = 9

Rejection Region at 0.05 → $\frac{1}{2} * \propto = \frac{1}{2} * \ 0.05 = 0.025$

$t_{0.025\,,\,9} = 2.262$

$$t_c = \frac{\bar{K} - \mu_D - 0}{\frac{S\,\bar{K}}{n}}$$

$$\bar{K} = \frac{\sum K}{n} = \frac{45}{10} = 4.5$$

$$S^2K = \frac{\sum K_i^2 - \frac{(\sum K_i)^2}{n}}{n-1} = \frac{8^2 + 7^2 + \cdots + 8^2 - \frac{(8+7+\cdots+8)^2}{10}}{10-1} = 15.39$$

$$S\bar{K} = \sqrt{\frac{S^2K}{n}} = \sqrt{\frac{15.39}{10}} = 1.24$$

$= 3.63 \therefore\ t_c = \frac{4.5-0}{1.24}$

The decision : since $t_c > t0.025, 9$ then accept Alternative Hypothesis

3. One-sample t-test

In testing the null hypothesis that the population mean is equal to a specified value μ_0, one uses the statistic :

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Example: find the differences between $\bar{Y}_1$ *and* $\bar{Y}_2$ if $n_1 = 15, S_1 = 40\ \bar{Y}_1 = 120$ *and* $n_2 = 22, S_2 = 35\ \bar{Y}_2 = 96$ by t-test, where $t_{0.05} = 2.035$ *and* $t_{0.01} = 2.704$?

Solve

$H_O : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 - \mu_2 \neq 0$

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(120-96)-(0)}{\sqrt{\frac{1600}{15} + \frac{1225}{22}}} = \frac{24}{12.75} = 1.88$$

$d.f = (n_1 + n_2) - 2 = (15 + 22) = 37 - 2 = 35$

$\therefore\ accept\ \mu_1 - \mu_2 = 0\ there\ are\ not\ difference\ between\ \bar{Y}_1\ and\ \bar{Y}_2$