WIKIPEDIA
The Free Encyclopedia

Contents

move to sidebar hide

Student's *t*-test

A ***t*-test** is any statistical hypothesis test in which the test statistic follows a Student's *t*-distribution under the null hypothesis. It 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 (typically, the scaling term is unknown and therefore a nuisance parameter). When the scaling term is estimated based on the data, the test statistic—under certain conditions—follows a Student's *t* distribution. The *t*-test's most common application is to test whether the means of two populations are different.

History

William Sealy Gosset, who developed the "*t*-statistic" and published it under the pseudonym of "Student"

The term "*t*-statistic" is abbreviated from "hypothesis test statistic".[1] In statistics, the t-distribution was first derived as a posterior distribution in 1876 by Helmert[2][3][4] and Lüroth.[5][6][7] The t-distribution also appeared in a more general form as Pearson Type IV distribution in Karl Pearson's 1895 paper.[8] However, the T-Distribution, also known as Student's t-distribution, gets its name from William Sealy Gosset who first published it in English in 1908 in the scientific journal Biometrika using the pseudonym "Student"[9][10] because his employer preferred staff to use pen names when publishing scientific papers.[11] 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 (see Student's *t*-distribution for a detailed history of this pseudonym, which is not to be confused with the literal term *student*). 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"[12] 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.[13] 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.[14]

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.[15] Gosset's identity was then known to fellow statisticians and to editor-in-chief Karl Pearson.[16]

Uses

The most frequently used *t*-tests are one-sample and two-sample tests:

- 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.[17]

Assumptions

Most test statistics have the form $t = \frac{Z}{s}$, 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

$$t = \frac{Z}{s} = \frac{\bar{X} - \mu}{\widehat{\sigma}/\sqrt{n}}$$

where $\overline{X}$ is the sample mean from a sample $X_1, X_2, \ldots, X_n$, of size n, s is the standard error of the mean, $\widehat{\sigma}$ is the estimate of the standard deviation of the population, and μ is the population mean.

The assumptions underlying a *t*-test in the simplest form above are that:

- $\overline{X}$ follows a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$
- $s^2(n-1)/\sigma^2$ 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.[18]

- 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.[19] 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.[20]

Most two-sample *t*-tests are robust to all but large deviations from the assumptions.[21]

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 χ^2 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. That is as sample size n increases:

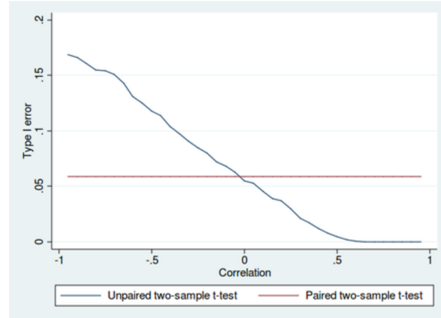
- $\sqrt{n}(\bar{X}-\mu) \xrightarrow{d} N\left(0, \sigma^2\right)$ as per the Central limit theorem.
- $s^2 \xrightarrow{p} \sigma^2$ as per the Law of large numbers.
- $\therefore \frac{\sqrt{n}(\bar{X}-\mu)}{s} \xrightarrow{d} N(0,1)$

Unpaired and paired two-sample *t*-tests

Two-sample *t*-tests for a difference in means involve independent samples (unpaired samples) or paired samples. Paired *t*-tests are a form of blocking, and have greater power (probability of avoiding a type II error, also known as a false negative) 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.[22] 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, and one variable from each of the two populations is 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.

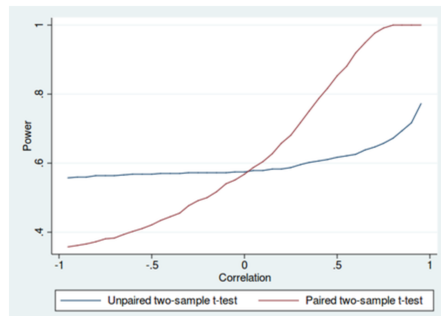


Type I error of unpaired and paired two-sample t-tests as a function of the correlation. The simulated random numbers originate from a bivariate normal distribution with a variance of 1. The significance level is 5% and the number of cases is 60.

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 interpatient 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 $\frac{n}{2} - 1$ 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).[23]



Power 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 and a deviation of the expected value of 0.4. The significance level is 5% and the number of cases is 60.

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.[24] 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 level), then the null hypothesis is rejected in favor of the alternative hypothesis.

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

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

where $\bar{x}$ is the sample mean, s is the sample standard deviation and n is the sample size. The degrees of freedom used in this test are $n - 1$. Although the parent population does not need to be normally distributed, the distribution of the population of sample means $\bar{x}$ is assumed to be normal.

By the central limit theorem, if the observations are independent and the second moment exists, then t will be approximately normal N(0;1).

Slope of a regression line

Suppose one is fitting the model

$$Y = \alpha + \beta x + \varepsilon$$

where x is known, α and β are unknown, ε is a normally distributed random variable with mean 0 and unknown variance σ^2, and Y is the outcome of interest. We want to test the null hypothesis that the slope β is equal to some specified value β_0 (often taken to be 0, in which case the null hypothesis is that x and y are uncorrelated).

Let

$$\begin{aligned} \widehat{\alpha}, \widehat{\beta} &= \text{least-squares estimators}, \\ SE_{\widehat{\alpha}}, SE_{\widehat{\beta}} &= \text{the standard errors of least-squares estimators}. \end{aligned}$$

Then

$$t_{\text{score}} = \frac{\widehat{\beta} - \beta_0}{SE_{\widehat{\beta}}} \sim \mathcal{T}_{n-2}$$

has a t-distribution with $n - 2$ degrees of freedom if the null hypothesis is true. The standard error of the slope coefficient:

$$SE_{\widehat{\beta}} = \frac{\sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

can be written in terms of the residuals. Let

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \left(\widehat{\alpha} + \widehat{\beta}x_i\right) = \text{residuals} = \text{estimated errors},$$
$$\text{SSR} = \sum_{i=1}^{n}\hat{\varepsilon}_i{}^2 = \text{sum of squares of residuals}.$$

Then t_{score} is given by:

$$t_{\text{score}} = \frac{\left(\widehat{\beta} - \beta_0\right)\sqrt{n-2}}{\sqrt{\frac{SSR}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}}.$$

Another way to determine the t_{score} is:

$$t_{\text{score}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

where r is the Pearson correlation coefficient.

The $t_{\text{score, intercept}}$ can be determined from the $t_{\text{score, slope}}$:

$$t_{\text{score,intercept}} = \frac{\alpha}{\beta}\frac{t_{\text{score,slope}}}{\sqrt{s_{\text{x}}^2 + \bar{x}^2}}$$

where $s_{\text{x}}{}^2$ is the sample variance.

Independent two-sample *t*-test

Equal sample sizes and variance

Given two groups (1, 2), this test is only applicable when:

- the two sample sizes are equal;
- it can be assumed that the two distributions have the same variance;

Violations of these assumptions are discussed below.

The t statistic to test whether the means are different can be calculated as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{2}{n}}}$$

where

$$s_p = \sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{2}}.$$

Here s_p is the pooled standard deviation for $n = n_1 = n_2$ and $s_{X_1}^2$ and $s_{X_2}^2$ are the unbiased estimators of the population variance. The denominator of t is the standard error of the difference between two means.

For significance testing, the degrees of freedom for this test is $2n - 2$ where n is sample size.

Equal or unequal sample sizes, similar variances ($\frac{1}{2} < \frac{s_{X_1}}{s_{X_2}} < 2$)

This test is used only when it can be assumed that the two distributions have the same variance. (When this assumption is violated, see below.) The previous formulae are a special case of the formulae below, one recovers them when both samples are equal in size: $n = n_1 = n_2$.

The t statistic to test whether the means are different can be calculated as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_{X_1}^2 + (n_2 - 1)\, s_{X_2}^2}{n_1 + n_2 - 2}}$$

is the pooled standard deviation of the two samples: it is defined in this way so that its square is an unbiased estimator of the common variance whether or not the population means are the same. In these formulae, $n_i - 1$ is the number of degrees of freedom for each group, and the total sample size minus two (that is, $n_1 + n_2 - 2$) is the total number of degrees of freedom, which is used in significance testing.

Equal or unequal sample sizes, unequal variances ($s_{X_1} > 2s_{X_2}$ or $s_{X_2} > 2s_{X_1}$)

This test, also known as Welch's t-test, is used only when the two population variances are not assumed to be equal (the two sample sizes may or may not be equal) and hence must be estimated separately. The t statistic to test whether the population means are different is calculated as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{\Delta}}}$$

where

$$s_{\bar{\Delta}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Here s_i^2 is the unbiased estimator of the variance of each of the two samples with n_i = number of participants in group i (i = 1 or 2). In this case $(s_{\bar{\Delta}})^2$ is not a pooled variance. For use in significance testing, the distribution of the test statistic is approximated as an ordinary Student's t-distribution with the degrees of freedom calculated using

$$\mathrm{d.f.} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1} + \frac{\left(s_2^2/n_2\right)^2}{n_2-1}}.$$

This is known as the Welch–Satterthwaite equation. The true distribution of the test statistic actually depends (slightly) on the two unknown population variances (see Behrens–Fisher problem).

Exact method for unequal variances and sample sizes

The test[25] deals with the famous Behrens–Fisher problem, i.e., comparing the difference between the means of two normally distributed populations when the variances of the two populations are not assumed to be equal, based on two independent samples.

The test is developed as an exact test that allows for **unequal sample sizes** and **unequal variances** of two populations. The exact property still holds even with small **extremely small and unbalanced sample sizes** (e.g. $n_1 = 5, n_2 = 50$).

The Te statistic to test whether the means are different can be calculated as follows:

Let $X = [X_1, X_2, \ldots, X_m]^T$ and $Y = [Y_1, Y_2, \ldots, Y_n]^T$ be the i.i.d. sample vectors ($m > n$) from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ separately.

Let $(P^T)_{n \times n}$ be an $n \times n$ orthogonal matrix whose elements of the first row are all $1/\sqrt{n}$, similarly, let $(Q^T)_{n \times m}$ be the first n rows of an $m \times m$ orthogonal matrix (whose elements of the first row are all $1/\sqrt{m}$).

Then $Z := (Q^T)_{n \times m} X/\sqrt{m} - (P^T)_{n \times n} Y/\sqrt{n}$ is an n-dimensional normal random vector.

$$Z \sim N((\mu_1 - \mu_2, 0, \ldots, 0)^T, (\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}) I_n).$$

From the above distribution we see that

$$Z_1 - (\mu_1 - \mu_2) \sim N(0, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}),$$

$$\frac{\sum_{i=2}^n Z_i^2}{n-1} \sim \frac{\chi_{n-1}^2}{n-1} \times (\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$$

$$Z_1 - (\mu_1 - \mu_2) \perp \sum_{i=2}^n Z_i^2.$$

$$T_e := \frac{Z_1 - (\mu_1 - \mu_2)}{\sqrt{(\sum_{i=2}^n Z_i^2)/(n-1)}} \sim t_{n-1}.$$

Dependent *t*-test for paired samples

This test is used when the samples are dependent; that is, when there is only one sample that has been tested twice (repeated measures) or when there are two samples that have been matched or "paired". This is an example of a paired difference test. The *t* statistic is calculated as

$$t = \frac{\bar{X}_D - \mu_0}{s_D/\sqrt{n}}$$

where $\bar{X}_D$ and s_D are the average and standard deviation of the differences between all pairs. The pairs are e.g. either one person's pre-test and post-test scores or between-pairs of persons matched into meaningful groups (for instance drawn from the same family or age group: see table). The constant μ_0 is zero if we want to test whether the average of the difference is significantly different. The degree of freedom used is $n-1$, where n represents the number of pairs.

Example of matched pairs

Pair	Name	Age	Test
1	John	35	250
1	Jane	36	340
2	Jimmy	22	460
2	Jessy	21	200

Example of repeated measures

Number	Name	Test 1	Test 2
1	Mike	35%	67%
2	Melanie	50%	46%
3	Melissa	90%	86%
4	Mitchell	78%	91%

Worked examples

Let A_1 denote a set obtained by drawing a random sample of six measurements:

$$A_1 = \{30.02,\ 29.99,\ 30.11,\ 29.97,\ 30.01,\ 29.99\}$$

and let A_2 denote a second set obtained similarly:

$$A_2 = \{29.89,\ 29.93,\ 29.72,\ 29.98,\ 30.02,\ 29.98\}$$

These could be, for example, the weights of screws that were chosen out of a bucket.

We will carry out tests of the null hypothesis that the means of the populations from which the two samples were taken are equal.

The difference between the two sample means, each denoted by $\overline{X}_i$, which appears in the numerator for all the two-sample testing approaches discussed above, is

$$\bar{X}_1 - \bar{X}_2 = 0.095.$$

The sample standard deviations for the two samples are approximately 0.05 and 0.11, respectively. For such small samples, a test of equality between the two population variances would not be very powerful. Since the sample sizes are equal, the two forms of the two-sample *t*-test will perform similarly in this example.

Unequal variances

If the approach for unequal variances (discussed above) is followed, the results are

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \approx 0.04849$$

and the degrees of freedom

$$\text{d.f.} \approx 7.031.$$

The test statistic is approximately 1.959, which gives a two-tailed test *p*-value of 0.09077.

Equal variances

If the approach for equal variances (discussed above) is followed, the results are

$$s_p \approx 0.08396$$

and the degrees of freedom

$$\text{d.f.} = 10.$$

The test statistic is approximately equal to 1.959, which gives a two-tailed *p*-value of 0.07857.

Related statistical tests

Alternatives to the *t*-test for location problems

The *t*-test provides an exact test for the equality of the means of two i.i.d. normal populations with unknown, but equal, variances. (Welch's *t*-test is a nearly exact test for the case where the data are normal but the variances may differ.) For moderately large samples and a one tailed test, the *t*-test is relatively robust to moderate violations of the normality assumption.[26] In large enough samples, the t-test asymptotically approaches the *z*-test, and becomes robust even to large deviations from normality.[18]

If the data are substantially non-normal and the sample size is small, the *t*-test can give misleading results. See Location test for Gaussian scale mixture distributions for some theory related to one particular family of non-normal distributions.

When the normality assumption does not hold, a non-parametric alternative to the *t*-test may have better statistical power. However, when data are non-normal with differing variances between groups, a t-test may have better type-1 error control than some non-parametric alternatives.[27] Furthermore, non-parametric methods, such as the Mann-Whitney U test discussed below, typically do not test for a difference of means, so should be used carefully if a difference of means is of primary scientific interest.[18] For example, Mann-Whitney U test will keep the type 1 error at the desired level alpha if both groups have the same distribution. It will also have power in detecting an alternative by which group B has the same distribution as A but after some shift by a constant (in which case there would indeed be a difference in the means of the two groups). However, there could be cases where group A and B will have different distributions but with the same means (such as two distributions, one with positive skewness and the other with a negative one, but shifted so to have the same means). In such cases, MW could have more than alpha level power in rejecting the Null hypothesis but attributing the interpretation of difference in means to such a result would be incorrect.

In the presence of an outlier, the t-test is not robust. For example, for two independent samples when the data distributions are asymmetric (that is, the distributions are skewed) or the distributions have large tails, then the Wilcoxon rank-sum test (also known as the Mann–Whitney *U* test) can have three to four times higher power than the *t*-test.[26][28][29] The nonparametric counterpart to the paired samples *t*-test is the Wilcoxon signed-rank test for paired samples. For a discussion on choosing between the *t*-test and nonparametric alternatives, see Lumley, et al. (2002).[18]

One-way analysis of variance (ANOVA) generalizes the two-sample *t*-test when the data belong to more than two groups.

A design which includes both paired observations and independent observations

When both paired observations and independent observations are present in the two sample design, assuming data are missing completely at random (MCAR), the paired observations or independent observations may be discarded in order to proceed with the standard tests above. Alternatively making use of all of the available data, assuming normality and MCAR, the generalized partially overlapping samples t-test could be used.[30]

Multivariate testing

A generalization of Student's *t* statistic, called Hotelling's *t*-squared statistic, allows for the testing of hypotheses on multiple (often correlated) measures within the same sample. For instance, a researcher might submit a number of subjects to a personality test consisting of multiple personality scales (e.g. the Minnesota Multiphasic Personality Inventory). Because measures of this type are usually positively correlated, it is not advisable to conduct separate univariate *t*-tests to test hypotheses, as these would neglect the covariance among measures and inflate the chance of falsely rejecting at least one hypothesis (Type I error). In this case a single multivariate test is preferable for hypothesis testing. Fisher's Method for combining multiple tests with *alpha* reduced for positive correlation among tests is one. Another is Hotelling's T^2 statistic follows a T^2 distribution. However, in practice the distribution is rarely used, since tabulated values for T^2 are hard to find. Usually, T^2 is converted instead to an F statistic.

For a one-sample multivariate test, the hypothesis is that the mean vector ($\boldsymbol{\mu}$) is equal to a given vector ($\boldsymbol{\mu}_0$). The test statistic is Hotelling's t^2:

$$t^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$$

where n is the sample size, $\bar{\mathbf{x}}$ is the vector of column means and $\mathbf{S}$ is an $m \times m$ sample covariance matrix.

For a two-sample multivariate test, the hypothesis is that the mean vectors ($\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$) of two samples are equal. The test statistic is Hotelling's two-sample t^2:

$$t^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{\text{pooled}}{}^{-1}\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right).$$

Software implementations

Many spreadsheet programs and statistics packages, such as QtiPlot, LibreOffice Calc, Microsoft Excel, SAS, SPSS, Stata, DAP, gretl, R, Python, PSPP, MATLAB and Minitab, include implementations of Student's *t*-test.

Language/Program	Function	Notes
Microsoft Excel pre 2010	`TTEST(array1, array2, tails, type)`	See [1] (https://support.office.com/en-US/article/TTEST-function-1696FFC1-4811-40FD-9D13-A0EAAD83C7AE)
Microsoft Excel 2010 and later	`T.TEST(array1, array2, tails, type)`	See [2] (https://support.office.com/en-US/article/TTEST-function-d4e08ec3-c545-485f-962e-276f7cbed055)
Apple Numbers	`TTEST(sample-1-values, sample-2-values, tails, test-type)`	See [3] (https://www.apple.com/au/mac/numbers/compatibility/functions.html)
LibreOffice Calc	`TTEST(Data1; Data2; Mode; Type)`	See [4] (https://help.libreoffice.org/Calc/Statistical_Functions_Part_Five#TTEST)
Google Sheets	`TTEST(range1, range2, tails, type)`	See [5] (https://support.google.com/docs/answer/6055837)
Python	`scipy.stats.ttest_ind(a, b, equal_var=True)`	See [6] (http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html)
MATLAB	`ttest(data1, data2)`	See [7] (http://www.mathworks.com/help/stats/ttest.html)
Mathematica	`TTest[{data1,data2}]`	See [8] (https://reference.wolfram.com/language/ref/TTest.html)
R	`t.test(data1, data2, var.equal=TRUE)`	See [9] (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/t.test.html)
SAS	`PROC TTEST`	See [10] (https://web.archive.org/web/20131029195637/http://www.sas.com/offices/europe/belux/pdf/academic/ttest.pdf)
Java	`tTest(sample1, sample2)`	See [11] (http://commons.apache.org/proper/commons-math/apidocs/org/apache/commons/math4/stat/inference/TTest.html)
Julia	`EqualVarianceTTest(sample1, sample2)`	See [12] (https://juliastats.org/HypothesisTests.jl/stable/parametric/#HypothesisTests.EqualVarianceTTest)
Stata	`ttest data1 == data2`	See [13] (https://www.stata.com/manuals/rttest.pdf)

See also

Mathematics portal

- Conditional change model
- *F*-test
- Noncentral *t*-distribution in power analysis
- Student's *t*-statistic
- *Z*-test
- Mann–Whitney *U* test
- Šidák correction for *t*-test
- Welch's *t*-test
- Analysis of variance (ANOVA)

References

Citations

1. *The Microbiome in Health and Disease* (https://books.google.com/books?id=kiToDwAAQBAJ&pg=PA397). Academic Press. 2020-05-29. p. 397. ISBN 978-0-12-820001-8.
2. Szabó, István (2003), "Systeme aus einer endlichen Anzahl starrer Körper", *Einführung in die Technische Mechanik*, Springer Berlin Heidelberg, pp. 196–199, doi:10.1007/978-3-642-61925-0_16 (https://doi.org/10.1007%2F978-3-642-61925-0_16), ISBN 978-3-540-13293-6
3. Schlyvitch, B. (October 1937). "Untersuchungen über den anastomotischen Kanal zwischen der Arteria coeliaca und mesenterica superior und damit in Zusammenhang stehende Fragen". *Zeitschrift für Anatomie und Entwicklungsgeschichte*. **107** (6): 709–737. doi:10.1007/bf02118337 (https://doi.org/10.1007%2Fbf02118337). ISSN 0340-2061 (https://www.worldcat.org/issn/0340-2061). S2CID 27311567 (https://api.semanticscholar.org/CorpusID:27311567).
4. Helmert (1876). "Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen gleicher Genauigkeit" (https://zenodo.org/record/1424695). *Astronomische Nachrichten* (in German). **88** (8–9): 113–131. Bibcode:1876AN.....88..113H (https://ui.adsabs.harvard.edu/abs/1876AN.....88..113H). doi:10.1002/asna.18760880802 (https://doi.org/10.1002%2Fasna.18760880802).
5. Lüroth, J. (1876). "Vergleichung von zwei Werthen des wahrscheinlichen Fehlers" (https://zenodo.org/record/1424693). *Astronomische Nachrichten* (in German). **87** (14): 209–220. Bibcode:1876AN.....87..209L (https://ui.adsabs.harvard.edu/abs/1876AN.....87..209L). doi:10.1002/asna.18760871402 (https://doi.org/10.1002%2Fasna.18760871402).
6. Pfanzagl J, Sheynin O (1996). "Studies in the history of probability and statistics. XLIV. A forerunner of the t-distribution". *Biometrika*. **83** (4): 891–898. doi:10.1093/biomet/83.4.891. MR 1766040.
7. Sheynin, Oscar (1995). "Helmert's work in the theory of errors". *Archive for History of Exact Sciences*. **49** (1): 73–104. doi:10.1007/BF00374700 (https://doi.org/10.1007%2FBF00374700). ISSN 0003-9519 (https://www.worldcat.org/issn/0003-9519). S2CID 121241599 (https://api.semanticscholar.org/CorpusID:121241599).
8. Pearson, K. (1895-01-01). "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. **186**: 343–414 (374). doi:10.1098/rsta.1895.0010. ISSN 1364-503X
9. "Student" William Sealy Gosset (1908). "The probable error of a mean" (PDF). *Biometrika*. **6** (1): 1–25. doi:10.1093/biomet/6.1.1. hdl:10338.dmlcz/143545. JSTOR 2331554
10. "T Table" (https://www.tdistributiontable.com).
11. Wendl MC (2016). "Pseudonymous fame". Science. 351 (6280): 1406. doi:10.1126/science.351.6280.1406 (https://doi.org/10.1126%2Fscience.351.6280.1406). PMID 27013722 (https://pubmed.ncbi.nlm.nih.gov/27013722)

12. Walpole, Ronald E. (2006). *Probability & statistics for engineers & scientists*. Myers, H. Raymond. (7th ed.). New Delhi: Pearson. ISBN 81-7758-404-9. OCLC 818811849 (https://www.worldcat.org/oclc/818811849).
13. O'Connor, John J.; Robertson, Edmund F., "William Sealy Gosset" (https://mathshistory.st-andrews.ac.uk/Biographies/Gosset.html), *MacTutor History of Mathematics archive*, University of St Andrews
14. "The Probable Error of a Mean" (http://seismo.berkeley.edu/~kirchner/eps_120/Odds_n_ends/Students_original_paper.pdf) (PDF). *Biometrika*. **6** (1): 1–25. 1908. doi:10.1093/biomet/6.1.1 (https://doi.org/10.1093%2Fbiomet%2F6.1.1). hdl:10338.dmlcz/143545 (https://hdl.handle.net/10338.dmlcz%2F143545). Retrieved 24 July 2016.
15. Raju, T. N. (2005). "William Sealy Gosset and William A. Silverman: Two 'Students' of Science". *Pediatrics*. **116** (3): 732–5. doi:10.1542/peds.2005-1134 (https://doi.org/10.1542%2Fpeds.2005-1134). PMID 16140715 (https://pubmed.ncbi.nlm.nih.gov/16140715). S2CID 32745754 (https://api.semanticscholar.org/CorpusID:32745754).
16. Dodge, Yadolah (2008). *The Concise Encyclopedia of Statistics* (https://books.google.com/books?id=k2zklGOBRDwC&pg=PA234). Springer Science & Business Media. pp. 234–235. ISBN 978-0-387-31742-7.
17. Fadem, Barbara (2008). *High-Yield Behavioral Science*. High-Yield Series. Hagerstown, MD: Lippincott Williams & Wilkins. ISBN 9781451130300.
18. Lumley, Thomas; Diehr, Paula; Emerson, Scott; Chen, Lu (May 2002). "The Importance of the Normality Assumption in Large Public Health Data Sets" (https://doi.org/10.1146%2Fannurev.publhealth.23.100901.140546). *Annual Review of Public Health*. **23** (1): 151–169. doi:10.1146/annurev.publhealth.23.100901.140546 (https://doi.org/10.1146%2Fannurev.publhealth.23.100901.140546). ISSN 0163-7525 (https://www.worldcat.org/issn/0163-7525). PMID 11910059 (https://pubmed.ncbi.nlm.nih.gov/11910059).
19. Markowski, Carol A.; Markowski, Edward P. (1990). "Conditions for the Effectiveness of a Preliminary Test of Variance". *The American Statistician*. **44** (4): 322–326. doi:10.2307/2684360 (https://doi.org/10.2307%2F2684360). JSTOR 2684360 (https://www.jstor.org/stable/2684360).
20. Guo, Beibei; Yuan, Ying (2017). "A comparative review of methods for comparing means using partially paired data". *Statistical Methods in Medical Research*. **26** (3): 1323–1340. doi:10.1177/0962280215577111 (https://doi.org/10.1177%2F0962280215577111). PMID 25834090 (https://pubmed.ncbi.nlm.nih.gov/25834090). S2CID 46598415 (https://api.semanticscholar.org/CorpusID:46598415).
21. Bland, Martin (1995). *An Introduction to Medical Statistics* (https://books.google.com/books?id=v6xpAAAAMAAJ). Oxford University Press. p. 168. ISBN 978-0-19-262428-4.
22. Rice, John A. (2006). *Mathematical Statistics and Data Analysis* (3rd ed.). Duxbury Advanced.
23. Weisstein, Eric. "Student's t-Distribution" (http://mathworld.wolfram.com/Studentst-Distribution.html). *mathworld.wolfram.com*.
24. David, H. A.; Gunnink, Jason L. (1997). "The Paired *t* Test Under Artificial Pairing". *The American Statistician*. **51** (1): 9–12. doi:10.2307/2684684 (https://doi.org/10.2307%2F2684684). JSTOR 2684684 (https://www.jstor.org/stable/2684684).
25. Chang Wang, Jinzhu Jia. "A New Non-asymptotic t-test for Behrens-Fisher Problems".https://arxiv.org/abs/2210.16473
26. Sawilowsky, Shlomo S.; Blair, R. Clifford (1992). "A More Realistic Look at the Robustness and Type II Error Properties of the *t* Test to Departures From Population Normality". *Psychological Bulletin*. **111** (2): 352–360. doi:10.1037/0033-2909.111.2.352 (https://doi.org/10.1037%2F0033-2909.111.2.352).