Paired sample t-test
(t-test for related data)

We have been considering cases where we wish to test whether two sample means are significantly different. Up to now we have considered cases where the two samples are independent, random samples – there is no association between the samples. However, we shall now consider the case where one value in one of the samples is related to another value in the second sample. The values are paired and the size of each sample is the same. We also are concerned with the case where the common sample variance is not known and has to be estimated. In this case the test is to apply is the t-test for related samples. [1]

Let (X_1, X_2) represent a pair of values where X_1 is drawn from the first sample and X_2 is drawn from the second sample. Then the difference for each pair is $D= X_1- X_2$.

Let $\overline{d}$ be the mean of the differences and S^2 the biased variance.

We assume that the mean of the difference is normally distributed

$$D \sim N\left(\overline{d}, \frac{\hat{\sigma}^2}{n}\right)$$

where $\hat{\sigma}^2$ is the unbiased estimate of the variance of the difference and n is the common sample size.

We estimate $\hat{\sigma}^2$ in the usual way

$$\hat{\sigma}^2 = s^2 = \frac{n}{n-1} S^2$$

Then the appropriate test statistic is:

$$t_{test} = \frac{\overline{\mathrm{d}}}{\sqrt{\frac{\sigma^2}{n}}}$$

The critical value is drawn from the Student's t-distribution with degrees of freedom $v = n-1$.

We are in a position to immediately illustrate this result.

[1] If you are not a "mathematician" but would like a simpler "one-step" version of the formula, go to the last section of these notes.

Example

It is believed that airfares between London and Budapest have fallen between 1997 and 1998. Travel agencies were sampled at random and their cheapest return air fare between these two celebrated cities was noted for the first week of January 1997 and 1998. The results were:

Agency	1997 fare	1998 fare
A	389	375
B	342	312
C	351	363
D	299	305
E	305	285
F	318	295

By means of a paired-sample t-test at the 5% significance level determine whether the belief that airfares have fallen is valid.

Let the 1997 fare be X and the 1998 fare be Y.
Let μ_x, μ_y be the main air fares in 1997 and1998 respectively.

H0: $\mu_x = \mu_y$
H1: $\mu_x > \mu_y$ one-tailed $\alpha = 5\%$

Let $D = X–Y$, then we test these hypotheses by their equivalents.

H_0 $\bar{D} = 0$
H_1 $\bar{D} > 0$
We need to estimate $\bar{\mathrm{D}}$ from the data and estimate the variance of the differences.

Agency	X	Y	$d=X–Y$	d^2
A	389	375	14	196
B	342	312	30	900
C	351	363	-12	144
D	299	305	-6	36
E	305	285	20	400
F	318	295	23	529
			$\sum d = 69$	$\sum d^2 = 2205$

$$\bar{d} = \frac{\sum d}{n} = \frac{69}{6} = 11.5$$

$$S^2 = \frac{\sum d^2}{n} - \left(\bar{d}\right)^2 = 235.25$$

$$\hat{\sigma}^2 = s^2 = \frac{n}{n-1} S^2 = \frac{6}{5} \text{ x } 235.25 = 282.3$$

$$t_{test} = \frac{\bar{d}}{\sqrt{\frac{s^2}{n}}} = \frac{11.5}{\sqrt{\frac{282.3}{6}}} = 1.667 \ \ (3\ d.p.)$$

$$v = n - 1 = 6 - 1 = 5$$

$$t_{critical} = 2.015 \text{ for } v = 5 \ \ p = 0.950$$

$$t_{test} < t_{critical}$$

$\therefore$ Accept H_0, reject H_1

Air-fares have not significantly fallen.

One-step formula

The t-test for related data is frequently used by "non-mathematicians" to test hypotheses formed in the context of psychology, geography and the social sciences among other subject areas. For the benefit of these students a "one-step" version of the formula is

$$t_{test} = \frac{\sum d}{\sqrt{\frac{N\sum d^2 - \left[\sum d\right]^2}{N-1}}}$$

To explain the variables used in this formula.

The test starts with values drawn from two samples – these values are paired off because of some relationship between them. Let X_1 represent one value in a pair and X_2 represent the other. Then the difference between these two pairs is

$$d = X_1 - X_2$$

This is where the d comes from in the above formula. There will be N pairs. We illustrate this with the same example of the main text. The symbol $\sum$ means "sum" or "add up".

Example

It is believed that airfares between London and Budapest have fallen between 1997 and 1998. Travel agencies were sampled at random and their cheapest return air fare between these two celebrated cities was noted for the first week of January 1997 and 1998. The results were:

Agency	1997 fare	1998 fare
A	389	375
B	342	312
C	351	363
D	299	305
E	305	285
F	318	295

By means of a paired-sample t-test at the 5% significance level determine whether the belief that airfares have fallen is valid.

Solution

Firstly, find the differences between the pairs of values; find the squares of these differences and sum accordingly, as follows

Agency	X	Y	$d=X-Y$	d^2
A	389	375	14	196
B	342	312	30	900
C	351	363	-12	144
D	299	305	-6	36
E	305	285	20	400
F	318	295	23	529
			$\sum d = 69$	$\sum d^2 = 2205$

Then substitute into the formula

$$t_{test} = \frac{\sum d}{\sqrt{\frac{N\sum d^2 - \left[\sum d\right]^2}{N-1}}}$$

Here, $\sum d = 69$ and $\sum d^2 = 2205,$ there are 6 pairs of data, so $N = 6,$ hence

$$t_{test} = \frac{69}{\sqrt{\frac{6\times 2205 - 69^2}{5}}} = 1.677\left(3.d.p.\right)$$

This gives the test value for the sample data. We must compare it with a critical value drawn from a table of such values. We are told to test the hypothesis at the 5% level, so that determines which column (or row) to use. The other row (or column) is for the degrees of freedom. In this test the degrees of freedom is one less than the number in the sample.

Hence, the number of degrees of freedom is 5, since the number of pairs in the sample is 6. From tables,

$t_{critical} = 2.015.$

Then

$t_{test} < t_{critical}$

$\therefore$ We reject the hypothesis that air-fares have fallen signficantly.