T-test for two dependent samples – some practical notes

Article in Slovak Journal of Sport Science · July 2021

DOI: 10.24040/sjss.2021.7.1.48-55

1 author:

Erika Liptáková
Technical University of Košice

T-test for two dependent samples – some practical notes

Erika Liptáková[1]

[1]Department of Applied Mathematics and Business Informatics, Faculty of Economics, Technical University of Košice, Slovakia

Abstract

This paper focuses on some important issues of the t-test for two dependent samples (paired t-test). Emphasis is placed on the assumptions of the use of the test, on the correct formulation of the tested hypotheses and on the selection of the correct method in testing the averages of differences in pairwise observations. This is extended by the calculation of the confidence interval for the mean difference and calculation of the effect size by Cohend´s d. The paper also offers a demonstration of two different solutions for the same example by means of correctly and incorrectly chosen statistical test.

Key words: statistics, dependent samples, paired t-test

DOI: https://doi.org/10.24040/sjss.2021.7.1.48-55

Introduction

T-test for two dependent samples is a test used frequently in the field of medicine as well as in the sports sector. It compares the average values of two measurements performed on the same individual, object, or related units. For example, such paired measurements can be:

- measurements performed at two different time periods (e.g. pre-test and post-test scores with a certain intervention performed between these tests);
- measurement performed under two different conditions (e.g. passing a test under "control" and "experimental" conditions);
- measurements made from two halves or two sides of the subject or experimental unit (e.g. vision measurements in the subject's left and right eye).

The purpose of the test is to determine whether there are statistically significant differences between paired observations, that means whether there is a statistically significant difference in the observed variable between two time points, two conditions or two measurements (1).

Assumptions

The t-test for two dependent samples works with interval or ratio data. This test belongs to parametric procedures, and it is based on the following assumptions:

- The sample of *n* subjects has been randomly selected from the population it represents.
- The differences between paired observations need to be

approximately normally distributed. It also follows that there are not significant outliers in the differences between the two related samples (2).

1.1 Assumption about normality

The paired t-test assumes that the differences D between two related samples X and Y (calculation: $D_i = X_i - Y_i; i = 1, 2, \dots, n$) have a normal distribution. Therefore, it is not recommended to use the test when there is an extreme value among the differences. This assumption may be impaired due to the nature of data, for example, the differences may be proportional to the values (change in length can be proportional to the length, etc.). In these cases, instead of the difference $Xi - Yi$, a different degree of deviation can be used, for example the ratio Xi / Yi, relative deviation *(Xi –Yi) / Yi*, etc. The aim is to obtain values with symmetric and localized distribution. It would be very helpful to draw data and try different variants.

1.2 Robustness

The t-test is robust due to the slight deviations from normal, which means that it shall not be invalidated if the data distribution slightly differs from normal. In general, robustness:

- is lower for tests against one-sided alternatives;
- decreases with the level of significance α;
- grows with file size n.

Computational background of paired t-test

2.1 Hypotheses

Hypotheses can be written in two different ways that express the same idea and are equivalent mathematically.

The first way of expression of the hypotheses is as follows (2):

$H_0: \mu_1 = \mu_2$ (the paired population means are equal)

$H_1: \mu_1 \neq \mu_2$ (the paired population means are not equal)

where:

$-\ \mu_1$ is the population mean of the first paired variable

$-\ \mu_2$ is the population mean of the second paired variable.

Alternative hypothesis may be also entered as one-sided:

$H_1: \mu_1 > \mu_2$ or $H_1: \mu_1 < \mu_2$

The second way of expression of the hypotheses is more appropriate for paired samples X and Y because the focus is on differences between paired observations:

$H_0: \mu(X - Y) = 0$ (the mean difference between the paired observations is equal to 0)

$H_1: \mu(X - Y) \neq 0$ (the mean difference between the paired observations is not equal to 0)

Alternative hypothesis may be also entered as one-sided:

$H_1: \mu(X - Y) > 0$ or $H_1: \mu(X - Y) < 0$.

There is also a generalized way of writing *H₀*, when there is assuming that there exists some specific difference μ_0 between paired observations: (3)

$H_0: \mu(X - Y) = \mu_0$ (the mean difference is about the claimed value μ_0)

$H_1: \mu(X - Y) \neq \mu_0$ (the mean difference is quite different than the claimed value μ_0)

2.2 Test statistic

To obtain the test statistic, it is first necessary to calculate the differences between the paired observations thus we get the random variable D ($D_i = y_i - x_i$). This is followed by the calculation of the mean and standard deviation. Test statistic has a Student´s t-distribution with *n-1* degrees of freedom and follows the same formula as the one sample t-test (4):

$$t = \frac{\bar{D}}{SE_{\bar{D}}} = \frac{\bar{D}}{s_D/\sqrt{n}}$$

where:

$\bar{D}$, s_D – mean and standard deviation of the differences

n - sample size

$SE_{\bar{D}}$ - standard error of the mean difference.

If the test statistic falls into a critical area, it means that the mean difference is statistically significant.

It should be noted that as the file size *n* increases, so does the value of the test statistic. This in turn leads to an increased likelihood of H_0 rejection.

If any of assumptions mentioned above are clearly violated, the reliability of the test statistics may be compromised.

2.3 Confidence interval for the true mean difference

Sometimes there is a situation that the mean difference seems to be small, nevertheless, the null hypothesis is rejected. In this case, it would be useful to calculate confidence interval for the mean difference. The borders of the interval delimit the region in which the true mean difference is likely to be located. The formula for two-sided γ% confidence interval is (2):

$$\bar{D} \pm t_{\frac{\gamma+1}{2}}(n-1)\,\frac{s_D}{\sqrt{n}}$$

where:

γ – level of confidence, usually 95% (but can also be 90% or 99%)

$t_{\frac{\gamma+1}{2}}(n-1)$ – critical value which is $\left(\frac{\gamma+1}{2}\right)$quantile of Student´s t-distribution.

2.4 Method for computing the power of the test – Cohen´s *d* index

Cohen´s d index can be used for computing the power of the t-test for two dependent samples as well as for the power of the t-test for two independent sample. Cohen´s *d* formula (5):

$$d = \frac{\bar{D}}{s_D}$$

Effect sizes proposed by Cohen, are: 0,2 (small effect), 0,5 (moderate effect) and 0,8 (large effect). It follows that if the average difference between the two dependent samples does not differ by 0,2 standard deviation or more, this difference is trivial, even if it is statistically significant.

1. **Reasons for using paired t-test instead of t-test for two independent samples**

In the case of dependent samples, paired t-test is considered stronger than t-test for two independent samples. This is because there are the same subjects in both samples compared. Thus, it eliminates differences between samples that could have a different cause than what is being tested.

The paired t-test is suitable when there is a relationship between the values of *X* and *Y*. For not very small sample sizes, the paired t-test is stronger than t-test for two independent samples even with small correlations, small degree of linear dependence between the values of *X* and *Y*. It means that the paired t-test has a lower probability of type II error (probability of the non-rejection of a false null hypothesis) than two-sample t-test. Conversely, if there is no correlation between the *X* and *Y* values, the two-sample t-test is stronger than the paired t-test. (2)

2. **When the paired t-test cannot be used and what about it**

A paired t-test can only compare averages for only two related (paired) units. The observed variable, the differences between paired observations, must be continuous and have a normal distribution. This test is not suitable for the following analyses: (1)

- data are not in form of paired observations;
- comparisons between more than two groups;
- a continuous variable that has no normal distribution;
- ordinal data.

In these cases, the following can be used:

Two-sample t-test is used when comparing the values of a continuous normally distributed variable between two independent samples.

The ANOVA method is used when comparing the values of a continuous normally distributed variable between more than two independent samples.

The nonparametric Wilcoxon Signed-Ranks Test is used when comparing two dependent samples for a continuous variable that does not have a normal distribution.

The nonparametric Wilcoxon Signed-Ranks Test is used when comparing two dependent samples for an ordinal variable.

Example: diastolic blood pressure before and after activity

The data used in followed example represent diastolic blood pressure in a sample of 50 women before (Pre-test) and after (Post-test) short physical activity. We are interested in whether the given short activity caused a statistically significant change in the values of diastolic blood pressure. We expect an increase of blood pressure.

2.1 Paired t-test – right choice

There are 50 matched pairs of values of diastolic blood pressure. Tested hypothesis is:

$H_0: \mu(X_{\text{posttest}} - X_{\text{pretest}}) = 0$ … there is no change in diastolic blood pressure.

Alternative hypothesis can be stated as two-sided or one-sided:

$H_1: \mu(X_{\text{posttest}} - X_{\text{pretest}}) \neq 0 \ldots$ there is a change in diastolic blood pressure

or (stronger statement)

$H_1: \mu(X_{\text{posttest}} - X_{\text{pretest}}) > 0 \ldots$ there is an increase in diastolic blood pressure

First step in processing this test is calculation of differences. So, we get a new variable D $(D = X_{\text{posttest}} - X_{\text{pretest}})$.

The second step is checking the assumption of normality. As can be seen in Figure 1, the differences have a normal distribution (both, histogram and box-plot have symmetrical pattern without outliers).

The third step is obtaining an average value and standard deviation of differences – both measures are needed for calculating the confidence interval for mean. The mentioned values can be seen in the Table 1:

- sample mean equals to 2.82;
- limits of 95% confidence interval for mean are $\langle 1.59; 4.04 \rangle$ – it means that we are 95% confident that the true population mean lies between these two points. Because this confidence interval does not overlap the value zero, we can expect H_0 to be rejected.

Next step is calculating the test statistic and associated p-value (Table 1):

- test statistic (*t Value*) has the value of 4.64;
- p-value of the test ($Pr>|t|$) is lower than 0.0001.

Due to the p-value that is lower than the significance level 0,05, H_0 has to be rejected. There exists statistically significant difference between diastolic blood pressure before and after specific physical activity. Given the positive average difference, we can say that there was a significant increase in diastolic blood pressure. You would need to consider if this difference in blood pressure is practically important, not just statistically significant.

To complete the result, Cohen´s *d* index ($d = \frac{2.8200}{4.2985} = 0{,}656$) tells that the effect size is moderate. We can say that difference is not trivial, it is statistically significant difference.

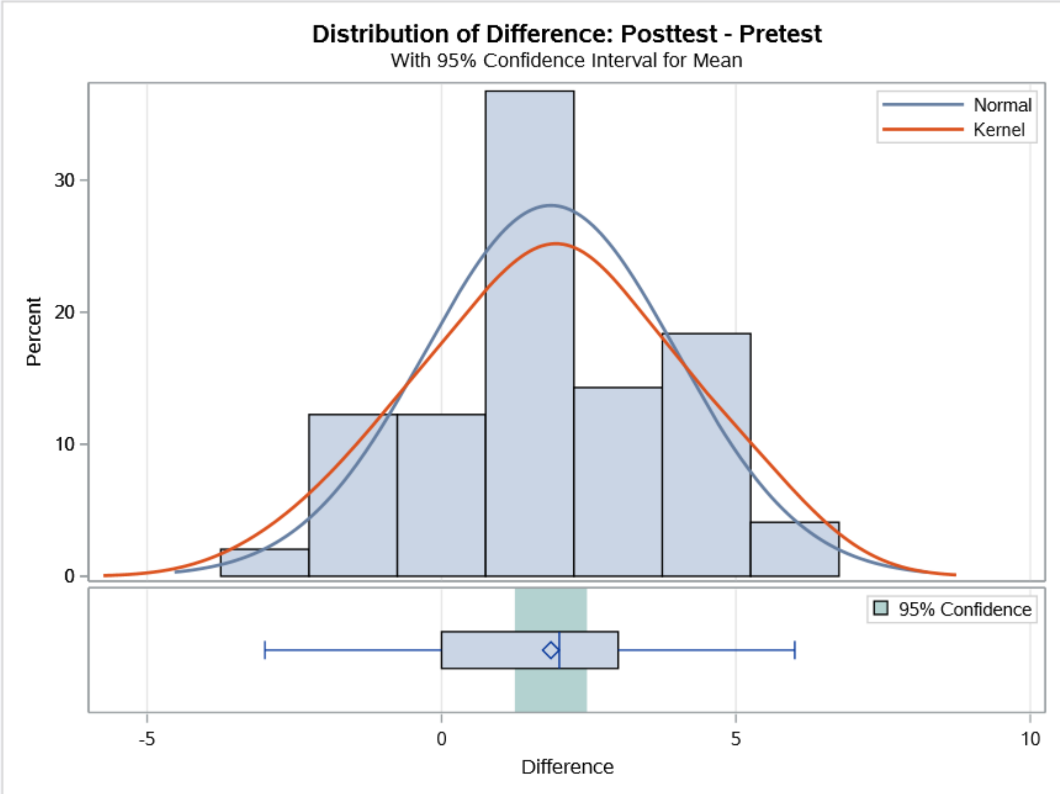

Figure 1 Histogram (with normality curve), box plot and 95% confidence interval of differences

Source: own processing; output from SAS® Enterprise Guide analytical software

Table 1 Mean, standard deviation and 95% confidence interval of differences. Test statistic (t Value) and p-value (Pr >ItI) of paired t-test

N	Mean	95% CL Mean		Std Dev	t Value	Pr > |t|
50	2.8200	1.5984	4.0416	4.2985	4.64	<.0001

Source: own processing

2.2 T-test for two independent samples – wrong choice

In the following example of a t-test calculation we do not take into account the dependency of the pre-test and post-test samples, so, they are considered to be independent.

As Figure 2 shows, both variables, pre-test and post-test, are normally distributed with approximately equal variances. The assumptions of the two-sample t-test are satisfied. Tested hypothesis is:

$H_0: \mu_{posttest} = \mu_{pretest}$ … there is no difference between average values.

Alternative hypothesis can be stated as two-sided or one-sided:

$H_1: \mu_{posttest} \neq \mu_{pretest}$ … there is a difference between average values.

or (stronger statement)

$H_1: \mu_{posttest} > \mu_{pretest}$ ... an average value of post-test is greater than that of pre-test.

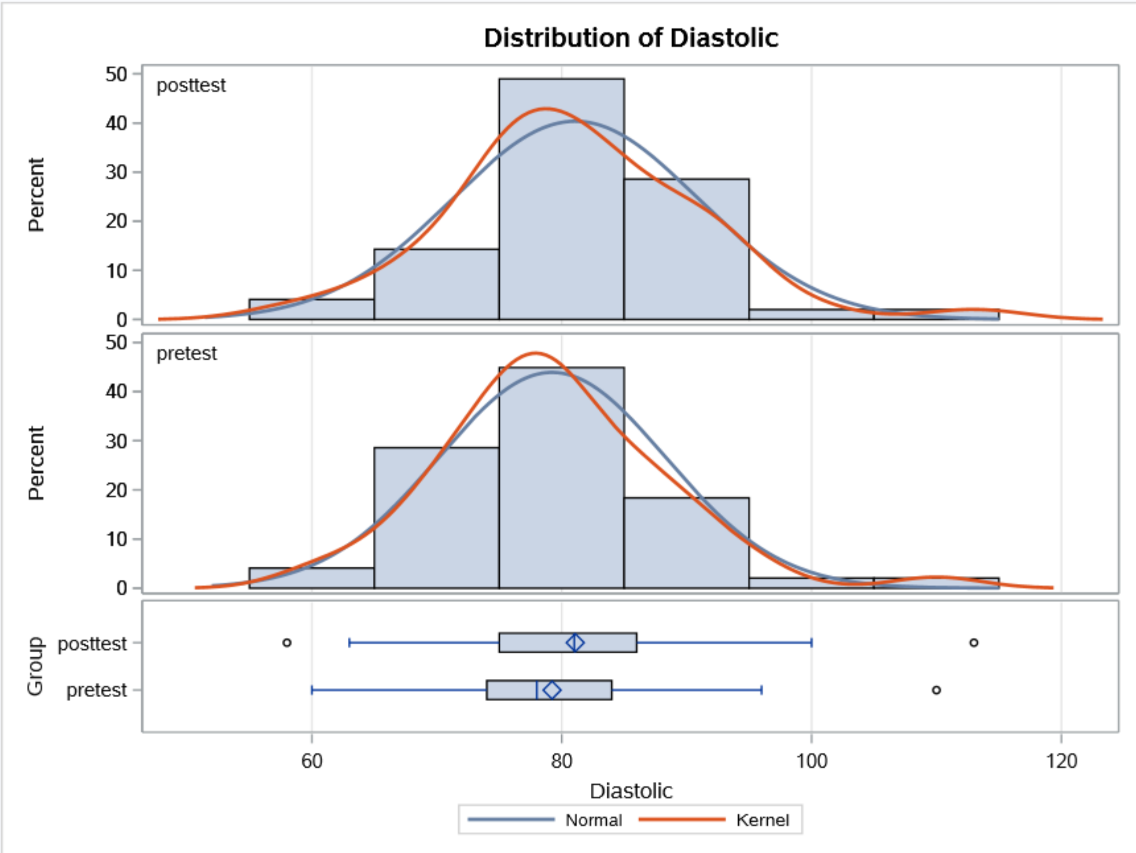

Figure 2 Distribution of pre-test and post-test samples – histograms and box-plots

Source: own processing; output from SAS® Enterprise Guide analytical software

First, this test works with both variables alone, therefore, average values, standard deviations and confidence intervals are calculated separately for both variables. The information is then combined into one test statistic.

Table 2 provides the following information about variables pre-test and post-test:

- Post-test: the average value is 81.08 with 95% confidence interval $\langle 78.24, 83.92 \rangle$
- Pre-test: the average value is 79.22 with 95% confidence interval $\langle 76.61, 81.83 \rangle$

As the confidence intervals for both variables overlap, we can expect H_0 not to be rejected.

Table 2 Means and confidence intervals for mean separately for both variables. Test statistic and p-value of two sample t-test.

Group	Mean	95% CL Mean		DF	t Value	Pr > |t|
posttest	81.0816	78.2425	83.9208	96	0.97	0.3352
pretest	79.2245	76.6160	81.8329			

Source: own processing

Table 2 also shows the test statistic (t Value = 0.97) and p-value (0.3352) of the performed statistical test. Since the p-value is higher than the significance level (0.3352 > 0.05), H_0 cannot be rejected which is a different conclusion than the one in the paired t-test.

Such a conclusion could be expected because the difference between the averages of the samples (81.0816-79.2245) is only 1.85. Given the diastolic blood pressure values, this difference is negligible.

In this example, the wrong test was used because the relationship between the two variables was ignored, and this led to the incorrect conclusion.

Conclusion

The main aim of that paper was to recall some important issues regarding the paired t-test and to point out the importance of choosing the right statistical test in the case of comparison of averages of two dependent samples.

The paper can be useful for beginners in the field of statistical data processing or those who are unsure of the correct choice of statistical test when comparing the mean values of two continuous variables.

References

1. SPSS tutorials: Paired samples t-test. Kent State University. [cited 2021 June 4] Available from: https://libguides.library.kent.edu/SPSS/PairedSamplestTest
2. Sheskin, D. J. Handbook of Parametric and Nonparametric Statistical Procedures. London: Taylor & Francis Group; 2007. pp. 1776. ISBN 1-58488-814-8
3. Lyócsa, Š., Baumöhl, E., Výrost, T. Kvantitatívne metódy v ekonómii II. Košice, 2013. pp. 460. ISBN 978-80-8086-210-7
4. Shier, R. Statistics: Paired t-test. Mathematics learning support centre. 2004. [cited 2021 June 4] Available from: https://www.statstutor.ac.uk/resources/uploaded/paired-t-test.pdf
5. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. Second ed. Hillsdale, NJ: Lawrence Erlbaum Associates,1998. pp. 560. ISBN 0-8058-0283-5