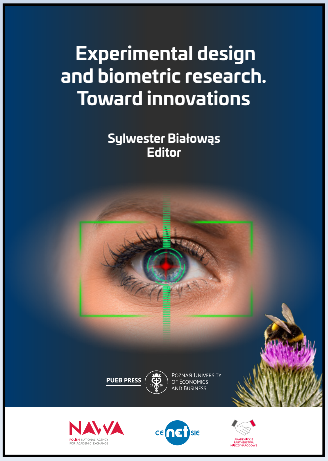

Experimental design and biometric research. Toward innovations

Sylwester Białowąs
Editor

POLISH NATIONAL AGENCY
FOR ACADEMIC EXCHANGE

AKADEMICKIE
PARTNERSTWA
MIĘDZYNARODOWE

eISBN 978-83-8211-079-1

https://doi.org/10.18559/978-83-8211-079-1

PUEB PRESS

Poznań University
of Economics
and Business

Experimental design and biometric research. Toward innovations, pp. 153-163
https://doi.org/10.18559/978-83-8211-079-1/III3

3.

DEPENDENT SAMPLES—SINGLE HYPOTHESIS TESTING

Sylwester Białowąs
Poznań University of Economics and Business

Adrianna Szyszka
Poznań University of Economics and Business

Abstract: This chapter deals with the approach of "within subjects" and focuses on single hypothesis testing. Both parametrical and non-parametrical versions are described. Every test is introduced, and the full step-by-step SPSS guidance is presented. The sections about effect size and about writing the report are included as well.

Keywords: paired sample *t*-test, Wilcoxon test.

3.1. The paired samples *t*-test

General information

Paired *t*-test is used to compare two related means mostly coming from a repeated measures design. In other words, data are collected by two measures from each observation, e.g. before and after a process or a phenomenon. For example, a researcher wants to test if the changes in the weight before and after a diet are significantly different from zero.

Hypotheses
H0: There is no difference between the paired mean scores.
H1: There is a difference between the paired mean scores.

Assumptions

There are the following assumptions associated with the paired samples *t*-test:
- the level of measurement should be interval or ratio (what in SPSS is indicated as scale level of measurement);
- the sample should be randomly selected which means that the data constitute a representative portion of the total population and every individual has the same chance to be selected into the sample (Verma & Abdel-Salam, 2019; Waters, 2011);
- the difference scores (not the raw scores) should follow the normal distribution.

Example

The community managing the apartment blocks has chosen a random group consisting of 58 families living in middle-size flats. The group got the instructions about electricity savings and recommendation to use the tools of controlling the electricity expenses. We have recorded two electricity bills of every family—one from the period of before, and the other one—after the recommendations.

Data info:
- variable 1: pretest—expenses before the recommendation—measurement level: scale (values: recorded electricity expenses per flat per month in EUR);
- variable 2: posttest—expenses after the recommendation—measurement level: scale (values: recorded electricity expenses per flat per month in EUR).

Testing the assumptions

Normality of distribution of differences

The first step in testing the normality of differences between scores is to calculate a new variable that is the difference between pretest and posttest values.

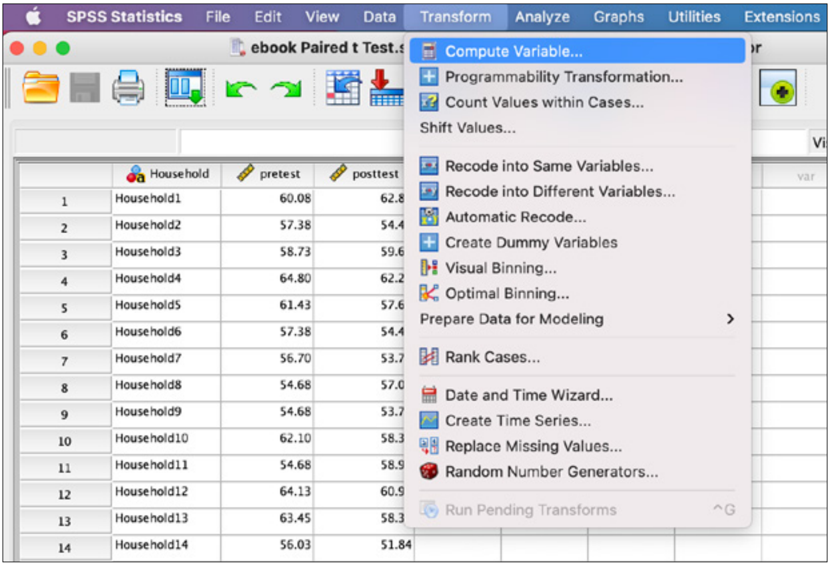

Figure 1. Calculating the difference between pretest and posttest values—path (1)

Source: The authors' own elaboration, IBM SPSS screenshot.

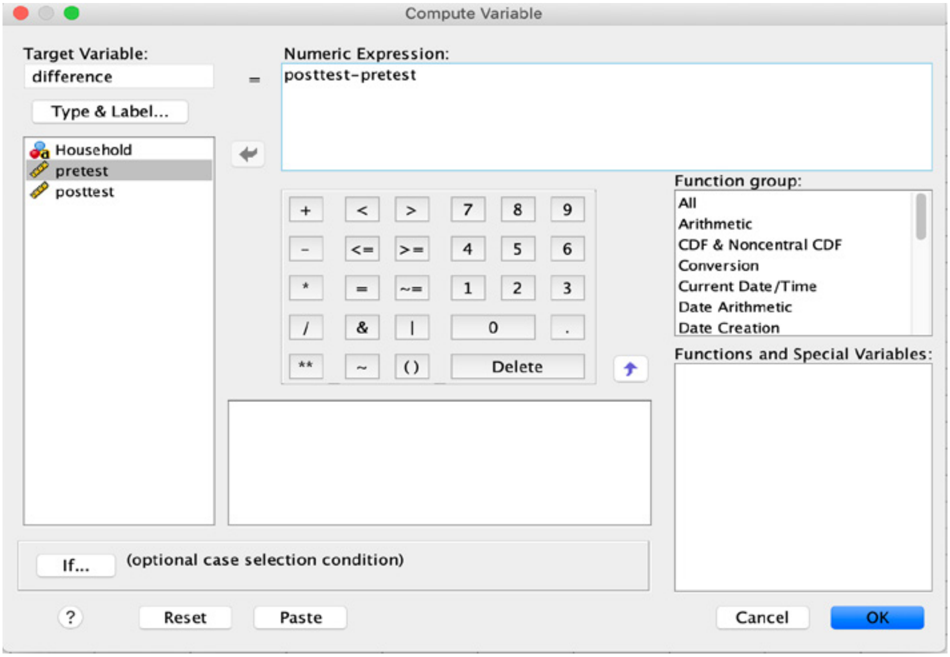

Figure 2. Calculating the difference between pretest and posttest values—path (2)

Source: The authors' own elaboration, IBM SPSS screenshot.

The commonly used test for testing the normality is the Kolmogorov-Smirnov test. This test compares the set of scores obtained in the study to the normally distributed scores.

The procedure of running the Kolmogorov-Smirnov test is shown in part 3, chapter 1. Of course, in the paired samples *t*-test we don't split the file and we measure only one variable—difference.

One-Sample Kolmogorov-Smirnov Test

		difference
N		58
Normal Parameters[a,b]	Mean	-3.5766
	Std. Deviation	3.46691
Most Extreme Differences	Absolute	.096
	Positive	.077
	Negative	-.096
Test Statistic		.096
Asymp. Sig. (2-tailed)		.200[c,d]

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

d. This is a lower bound of the true significance.

Figure 3. Kolmogorov-Smirnov test—results

Source: The authors' own elaboration, IBM SPSS screenshot.

We decide about the hypothesis by interpreting the *p*-value. If the test is significant ($p <. 05$) it means that the data do not follow normal distribution. If the test is non-significant ($p > .05$) the distribution of the obtained scores is normal (Field, 2013; Verma & Abdel-Salam, 2019). In this case, $p \geq .200$ which means that the assumption of normality is fulfilled.

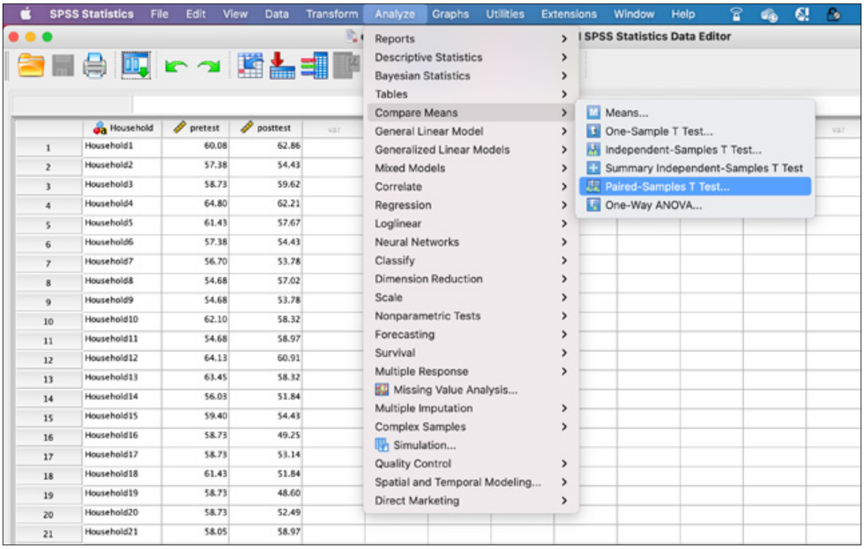

Figure 4. Paired samples *t*-test—path

Source: The authors' own elaboration, IBM SPSS screenshot.

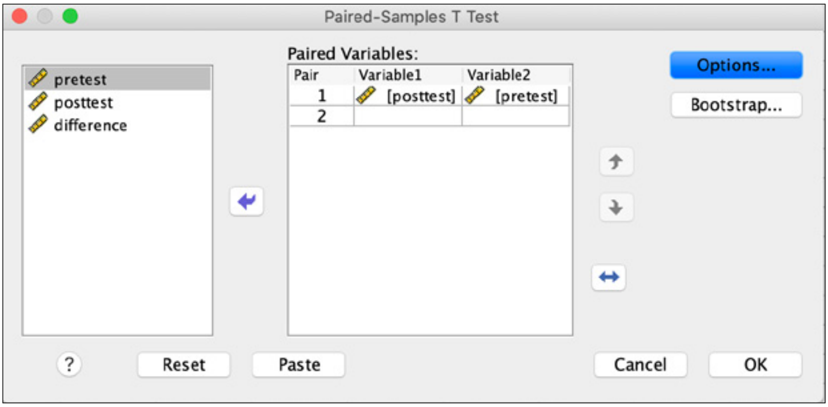

Figure 5. Paired samples *t*-test—dialog box

Source: The authors' own elaboration, IBM SPSS screenshot.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	posttest	55.9514	58	4.23716	.55637
	pretest	59.5280	58	3.45449	.45360

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	posttest & pretest	58	.610	.000

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	posttest - pretest	-3.57657	3.46691	.45523	-4.48815	-2.66499	-7.857	57	.000

Figure 6. Paired samples *t*-test—results

Source: The authors' own elaboration, IBM SPSS screenshot.

Results

In the upper table of the outcome (Paired Samples Statistics) we can read that the mean for the pretest is 59.53 and for the posttest is 55.95. It means that the average electricity bill declined by 3.58 EUR.

In the lowest table we can check if the difference is statistically significant by interpreting the p-value from the last column (Sig. 2-tailed). This value equals $p < .001$ which is lower than the critical value $p = .05$. It means that we can reject the null hypothesis and interpret the results as the statistically significant difference between pretest and posttest.

Paired samples t-test hypotheses resolution:

$p < .05$—there is a significant difference between pretest and posttest; reject H0;

$p > .05$—there is no significant difference between pretest and posttest; do not reject H0.

Effect size

In order to examine whether the observed difference is important, we can calculate effect size. For paired samples t-test a popular measure is Cohen's d:

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{pre}}$$

x_1, x_2 – means of both groups;

s_{pre} – standard deviation of the pretest group.

The Cohen's d has the following interpretation:
Below 0.2—no effect,
< 0.2 – 0.5)—small effect,
< 0.5 – 0.8)—medium effect,
0.8 and more—large effect.

$$d = \frac{|55.95 - 59.53|}{3.45} = 1.04$$

In our case, we can observe the large effect ($d = 1.04$).

Summary

The community managing the apartment blocks has chosen a random group consisting of 58 families living in middle-size flats. The group got the instructions about electricity savings and recommendation to use the tools of controlling the electricity expenses. We have recorded two electricity bills of every family—one from the period of before, and the other one—after the recommendations.

Data info:

- variable 1: pretest—expenses before the recommendation—measurement level: scale (values: recorded electricity expenses per flat per month in EUR);
- variable 2: posttest—expenses after the recommendation—measurement level: scale (values: recorded electricity expenses per flat per month in EUR).

The electricity expenses of the households changed significantly after recommendations $t(58) = -7.857$, $p < .001$, $d = 1.04$. The bills decreased on average from 59.53 EUR ($SD = 3.45$) to 55.95 EUR ($SD = 4.24$). A t-test revealed that the difference of 3.58 EUR is statistically significant ($p < .001$), suggesting that the informed groups spent less on electricity than the control group. Cohen's d statistic indicates the large effect.

More info about the paired samples *t*-test

In order to estimate the effect size, we used pretest standard deviation as a baseline. The proposed formula of calculating the denominator is used especially when standard deviation is expected to be increased remarkably by the treatment. Nevertheless, the formula of standard deviation in the denominator may be calculated in other ways. The highly recommended estimate of the baseline is s_{av} given by the following formula:

$$s_{av} = \sqrt{\frac{s_{pre}^2 + s_{post}^2}{2}}$$

It enables us to compare the results with effect size *d* estimations for one group or two independent groups. However, sometimes in the literature the effect size is calculated using standard deviation of differences between scores. This approach is not very advisable since it may give notably different estimations in comparison with different methods (e.g. when the standard deviation of differences is small, the *d* estimation is larger than the calculation with s_{av}) (Cumming, 2012).

References

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta--analysis*. Routledge Taylor & Francis Group.
Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (5^{th} ed.). Sage edge.
Verma, J. P., & Abdel-Salam, G. A.-S. (2019). *Testing statistical assumptions in research*. John Wiley & Sons, Inc.
Waters, D. (2011). *Quantitative methods for business* (5th ed.). Pearson Education Limited.

3.2. Wilcoxon signed-rank test

General information
The Wilcoxon signed rank test is a commonly used nonparametric alternative to the paired samples *t*-test (when the assumptions are violated). It applies to the related samples when we compare the scores in two different points or under two different conditions (e.g. before and after the treatment). It is also used when the dependent variable is measured at ordinal scale. Since the Wilcoxon signed rank test does not require the normality of distribution of the data, it does not compare means but ranks ranks (Pallant, 2011; Verma & Abdel-Salam, 2019).

Hypotheses:
H0: There is no difference between the scores.
H1: There is a difference between the scores.

Assumptions
There are the following assumptions associated with the Wilcoxon signed-rank test:
– the level of measurement of dependent variable must be at least ordinal;
– the score of both groups should be related.

Example
Dataset: The company managing sharing bicycles decided to check the impact of the station location on use of the bicycles. The station was set 200 m from the entrance of the high school. Random sample of the students has been selected. Students were asked about the frequency of using the bicycles. In the middle of the semester the

company set the station closer to the entrance. After one month, the same group of students were asked about the frequency of using bicycles again.

Data info:

- variable 1: pretest—ordinal (declared frequency of using the shared bicycles; 1—more than once a day; 2—every day; 3—2–4 times a week; 4—once a week; 5—once a month; 6—less than once a month; 7—never);
- variable 2: posttest—ordinal (declared frequency of using the shared bicycles; 1—more than once a day; 2—every day; 3—2–4 times a week; 4—once a week; 5—once a month; 6—less than once a month; 7—never).

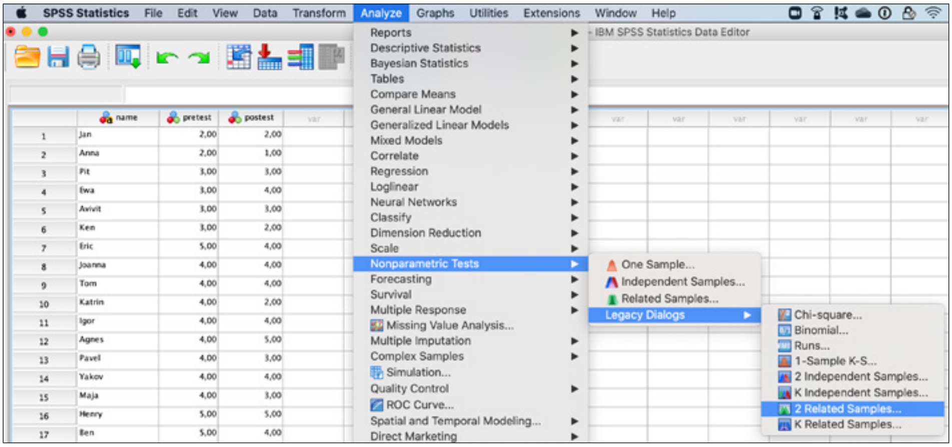

Figure 7. Wilcoxon signed-rank test—path

Source: The authors' own elaboration, IBM SPSS screenshot.

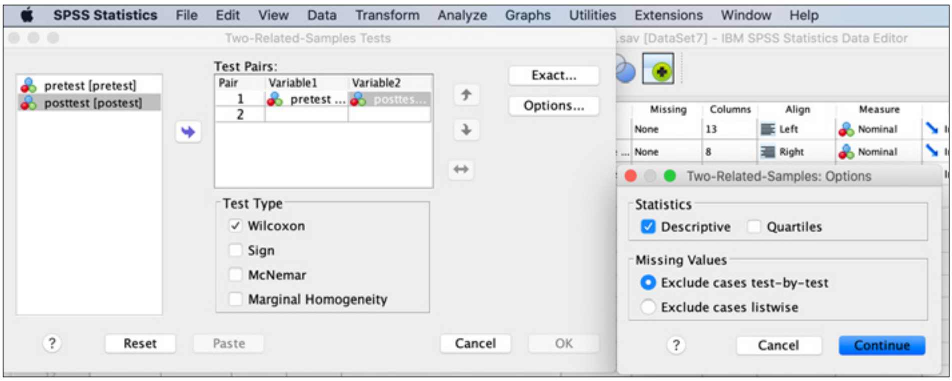

Figure 8. Wilcoxon signed-rank test—dialog box

Source: The authors' own elaboration, IBM SPSS screenshot.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
pretest	35	4.9429	1.49397	2.00	7.00
posttest	35	4.6000	1.66627	1.00	7.00

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
posttest – pretest	Negative Ranks	14[a]	9.79	137.00
	Positive Ranks	4[b]	8.50	34.00
	Ties	17[c]		
	Total	35		

a. posttest < pretest
b. posttest > pretest
c. posttest = pretest

Test Statistics[a]

	posttest – pretest
Z	-2.449[b]
Asymp. Sig. (2-tailed)	.014

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

Figure 9. Wilcoxon signed-rank test—results

Source: The authors' own elaboration, IBM SPSS screenshot.

Results

In the lowest table we can check if the difference is statistically significant by interpreting the p-value from the last row (Asymp. Sig. (2-tailed)). This value equals $p = .014$ which is lower than the critical value $p = .05$. It means that we can reject the null hypothesis and interpret the results as the statistically significant difference between pretest and posttest.

Wilcoxon signed ranked test hypotheses resolution:
$p < .05$—there is a significant difference between pretest and posttest; reject H0;
$p > .05$—there is no significant difference between pretest and posttest; do not reject H0.

Effect size

The effect size measure for Wilcoxon signed ranked test is r that is calculated using the statistic Z value and N which is total number of observations in both groups (the sum of observations in two groups):

$$r = \frac{|Z|}{\sqrt{N}}$$

The *r* has the following interpretation:
Below .1—no effect,
< .1 – .3)—small effect,
< .3 – .5)—medium effect,
.5 and more—large effect (Field, 2013; Pallant, 2011).

$$r = \frac{|-2.449|}{\sqrt{70}} = .29$$

In our example, $r = .29$ which may be considered as a small effect.

Summary
Dataset: The company managing sharing bicycles decided to check the impact of the station location on use of the bicycles. The station was set 200 m from the entrance of the high school. Random sample of the students has been selected. Students were asked about the frequency of using the bicycles. In the middle of the semester the company set the station closer to the entrance. After one month, the same group of students were asked about the frequency of using bicycles again.

Data info:
- variable 1: pretest—ordinal (declared frequency of using the shared bicycles; 1—more than once a day; 2—every day; 3—2–4 times a week; 4—once a week; 5—once a month; 6—less than once a month; 7—never);
- variable 2: posttest—ordinal (declared frequency of using the shared bicycles; 1—more than once a day; 2—every day; 3—2–4 times a week; 4—once a week; 5—once a month; 6—less than once a month; 7—never).

After relocation of the station, the frequency of using the shared bicycles changed significantly $Z(35) = -2.45$, $p = .014$. The students used the shared bicycles more frequent ($Mdn = 4$) compared to the initial location ($Mdn = 5$). However, effect size is rather small ($r = .29$).

References

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage edge.
Pallant, J. (2011). *SPSS survival manual: a step by step guide to data analysis using SPSS* (4th ed.). Allen & Unwin.
Verma, J. P., & Abdel-Salam, G. A.-S. (2019). *Testing statistical assumptions in research*. John Wiley & Sons, Inc.