

Wilcoxon test in SPSS (Quiz)



Wilcoxon Sign Rank Test practical questions

In this example you will use a Wilcoxon signed rank test to test whether students feel more informed about greenhouse gas issues (INFGAS) and nuclear waste issues (INFNUCL). Both variables were rated on the same four-point scale by all participants, making the Wilcoxon an appropriate test.

A first step to perform a Wilcoxon test in SPSS is to confirm that one cannot use a paired t test. To do this we construct a variable that is the difference between the two variables to be compared, in this case **How informed about greenhouse gases[INFGAS]** and **How informed about nuclear waste[INFNUCL]**. Do this now and then use SPSS to test for normality of this difference variable.

- Question: Do the histogram and normality tests suggest that the difference variable follows a normal distribution?

Next perform the Wilcoxon test in SPSS to compare differences between **How informed about greenhouse gases[INFGAS]** and **How informed about nuclear waste[INFNUCL]**.

- Question: Looking at the first SPSS table which variable has the higher mean rank?
- Question: Looking at the second SPSS table does this suggest there are significant differences between the variables?

Solutions to Wilcoxon Sign Rank Test practical questions

To construct the difference variable and test it for normality:

Before we can perform this test we need to check whether the differences between **INFGGAS** and **INFNUCL** are normally distributed. First we need to create a difference variable as follows:

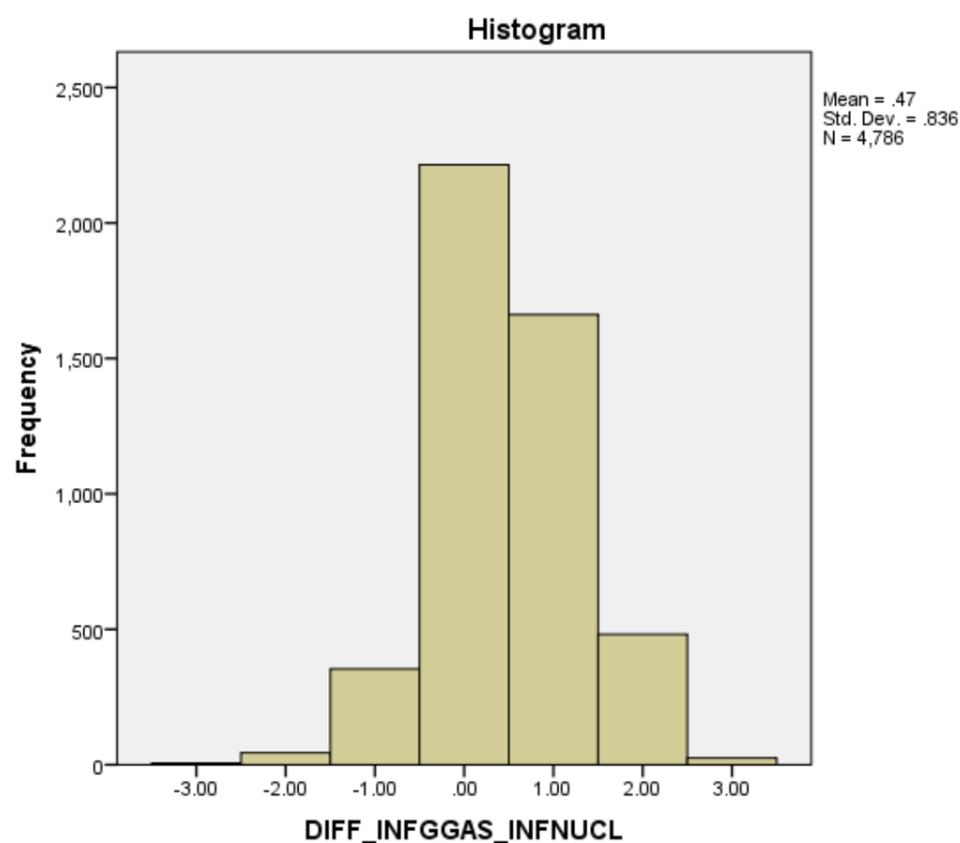
- Select **Compute** from the **Transform** menu.
- Type **DIFF_INFGGAS_INFNUCL** into the **Target Variable** box.
- Type **INFGGAS - INFNUCL** into the **Numeric Expression** box.
- Click on the **OK** button.

We can now use this new generated variable to perform normality checks. Do this as follows:

- Select **Descriptive Statistics** from the **Analyze** menu.
- Select **Explore** from the **Descriptive Statistics** sub-menu.
- Click on the **Reset** button.
- Copy the **DIFF_INFGGAS_INFNUCL** variable into the **Dependent List:** box.
- Click on the **Plots...** button.
- On the screen that appears select the **Histogram** tick box.
- Unselect the **Stem and leaf** button.
- Select the **Normality plots with tests** button.
- Click on the **Continue** button.
- Click on the **OK** button.

Question: Do the histogram and normality tests suggest that the difference variable follows a normal distribution?

Solution: The output from SPSS is as follows:



Ideally for a normal distribution this histogram should look symmetric around the mean of the distribution, in this case .4680. This distribution appears to be reasonably symmetric.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
DIFF_INFGGAS_INFNUCL	.259	4786	.000	.871	4786	.000

a. Lilliefors Significance Correction

The Kolmogorov Smirnov test produces test statistics that are used (along with a degrees of freedom parameter) to test for normality. Here we see that the Kolmogorov Smirnov statistic takes value .259. This has degrees of freedom which equals the number of data points, namely 4786.

Here we see that the p value (quoted under Sig. for Kolmogorov Smirnov) is .000 (reported as $p < .001$) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the variable follows a normal distribution.

Although the Kolmogorov Smirnov statistic tells the researcher whether the distribution followed by a variable is statistically significantly different from a normal distribution one should take care in not overinterpreting such findings. Significance will be strongly effected by the number of observations and so only a small discrepancy from normality will be deemed significant for very large sample sizes whilst very large discrepancies will be required to reject the null hypothesis for small sample sizes.

Here are the instructions to perform the Wilcoxon test in SPSS:

- Select **Non Parametric Tests** from the **Analyze** menu.
- Select **Legacy Dialogs** from the **Non Parametric Tests** sub-menu.
- Select **2 Related Samples...** from the **Legacy Dialogs** sub-menu.
- Click on the **Reset** button.
- Copy the **How informed about greenhouse gases[INFGGAS]** variable into the **Test Pairs:** box.
- Copy the **How informed about nuclear waste[INFNUCL]** variable into the **Test Pairs:** box.
- Click on the **Exact...** button.
- On the screen that appears select the **Exact** button.
- Click on the **Continue** button.
- Click on the **OK** button.

- Question: Looking at the first SPSS table which variable has the higher mean rank?

Solution: The output from SPSS is as follows:

		Ranks		
		N	Mean Rank	Sum of Ranks
How informed about nuclear waste - How informed about greenhouse gases	Negative Ranks	2168 ^a	1307.62	2834914.00
	Positive Ranks	403 ^b	1169.71	471392.00
	Ties	2215 ^c		
	Total	4786		

- a. How informed about nuclear waste < How informed about greenhouse gases
- b. How informed about nuclear waste > How informed about greenhouse gases
- c. How informed about nuclear waste = How informed about greenhouse gases

Here we have worked on **INFNUCL - INFGGAS** so that positive ranks are when **INFNUCL > INFGGAS**. Here we see that there are 2168 negative ranks, 403 positive ranks and 2215 ties.

Having worked out which observed pairs result in which sign for their difference, the magnitude (excluding the sign) of these differences is calculated and these are then ranked in order (excluding ties). We now see that the total of the ranks for the negative differences is 2834914.00 resulting in a mean rank of 1307.62 while the total of the ranks for the positive differences is 471392.00 resulting in a mean rank of 1169.71. Here the mean of the negative ranks is larger than that for positive ranks suggesting that values for **INFGGAS** are generally larger than for **INFNUCL**.

- Question: Looking at the second SPSS table does this suggest there are significant differences between the variables?

Solution: The output from SPSS is as follows:

Test Statistics	
How informed about nuclear waste - How informed about greenhouse gases	
Z	-33.512 ^b
Asymp. Sig. (2-tailed)	.000
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

- b. Based on positive ranks.

Although SPSS quotes the p value (quoted next to Asymp. Sig. (2-tailed)) as .000 it is not exactly zero and is in fact simply smaller than 0.001 as SPSS is quoting the first 3 decimal places. We therefore have significant evidence to reject the null hypothesis that the two groups have the same medians. This is given in the next row and we see that the exact p value is .000 whilst the asymptotic p value is .000. The exact p value agrees with the asymptotic p value that the null hypothesis can be rejected.

[It is clear from these results that students feel significantly better informed about greenhouse gases than they do about nuclear waste.](#)