One-Sample Wilcoxon Signed Rank Test (ts_wilcoxon_os)

Author: P. Stikker
Website: https://peterStatistics.com
YouTube: https://www.youtube.com/stikpet
Version: 0.2 (2023-01-25)

Introduction

The *ts_wilcoxon_os* function performs one-sample Wilcoxon signed rank test.

This document contains the details on how to use the functions, and formulas used in them.

1 About the Function

1.1 Input parameters:

- **data**
  the data as numbers.
  - *Excel* as a fixed range,
  - *Python* as a pandas data series
  - *R* as a vector

- *Optional parameters*
  - **hypMed**
    the hypothesized median to be used. The default will use the mid-range of the data
  - **ties**
    Boolean to indicate the use of the ties correction (i.e. scores equal to each other). Default is True.
  - **appr**
    Method to use for approximation. Either
    - "wilcoxon" (default) : standard z approximation
    - "exact" : exact test, only applicable if there are no ties
    - "imant": Iman's t approximation
    - "imanz": Iman's z approximation
  - **eqMed**
    method to deal with scores equal to hypMed. Either
    - "wilcoxon" (default) : scores equal will be removed
    - "pratt" : using Pratt method
    - "zsplit" : using z-split method
  - **cc**
    Boolean to indicate the use of a continuity correction. Default is False.
  - **out** (default is "pvalue") – only applies to non-array VBA function
    Choice what to show as result. Either:
    - "pvalue": show the p-value (significance)
    - "statistic": show the test-statistic used
    - "w": show the sum of positive ranks

1.2 Output

- **W**
  The sum of the positive ranks (Wilcoxon test statistic)

- **statistic**
  The test statistic used.

- **df**
  The degrees of freedom. Only applicable if Iman's t approximation is used

- **pVal**
  The p-value (significance) of the test, two tailed

- **testUsed**
  A description of the test used

Note for *Excel:*

the array function *ts_wilcoxon_os_arr* will require 2 rows and 5 columns.

1.3 Dependencies

- **Excel**
  The additional functions 'srf' for the signed ranks frequencies, 'wpmf' for the Wilcoxon probability mass function, and 'wcdf' for the Wilcoxon cumulative density function are needed and included to determine the exact distribution.
  You can run the **ts_wilcoxon_os_addHelp** macro so that the function will be available with some help in the 'User Defined' category in the functions overview.

- **Python**
  The following libraries are needed:
  - The *comb* function from Python's *math* library is used
  - The NormalDist function from Python's statistics library is used
  - Pandas is needed for data entry and showing the results, as well as the ranks function
  - The t function from Scipy.stats for the t-distribution

  Additionally the helper functions 'wcdf' and 'srf' are needed (and included) for the exact Wilcoxon distribution.

- **R**
  No other libraries required.

2 Examples

2.1 Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	Teach_Motivate											
2	1											
3	2				**appr**							
4	5				*wilcoxon*	*imant*	*imanz*					
5	1		**eqMed**	*wilcoxon*	0,252647321	0,265803801	0,250177215					
6	1			*pratt*	0,263623438	0,274483089	0,262042691					
7	5			*zsplit*	0,265831697	0,276775358	0,264322581					
8	3											
9	1		E5:G7 =>	=ts_wilcoxon_os($A$2:$A$21;;;E$4;$D5)								
10	5											
11	1		**hyp. Med.**	**ties**	**appr**	**eqMed**	**cc**	**out**				
12	1		3	TRUE	wilcoxon	wilcoxon	FALSE	pvalue				
13	5			FALSE	imant	pratt	TRUE	w				
14	1				imanz	zsplit		statistic				
15	1											
16	3		pvalue	0,252647321	=ts_wilcoxon_os(A2:A21;C12;D12;E12;F12;G12;H12)							
17	3		w	47	=ts_wilcoxon_os(A2:A21;C12;D12;E12;F12;G12;H13)							
18	3		statistic	1,143942892	=ts_wilcoxon_os(A2:A21;C12;D12;E12;F12;G12;H14)							
19	4											
20	2		W	statistic	df	p-value	test					
21	4		47	1,143942892	n.a.	0,252647321	one-sample Wilcoxon signed rank test, ties correction applied					
22												
23			C20:G21 =>	=ts_wilcoxon_os_arr(A2:A21)								
24												

2.2 Python

```
[3]: dataList = [1, 2, 5, 1, 1, 5, 3, 1, 5, 1, 1, 5, 1, 1, 3, 3, 3, 4, 2, 4]
     data = pd.Series(dataList)
```

```
[4]: ts_wilcoxon_os(data)
```

[4]:

	W	statistic	df	p-value	test
0	47.0	1.143943	n.a.	0.252647	one-sample Wilcoxon signed rank test, , with ties correction

```
[5]: ts_wilcoxon_os(data, ties=False, appr="imanz", eqMed = "zsplit", cc = False)
```

[5]:

	W	statistic	df	p-value	test
0	76.0	1.085116	n.a.	0.27787	one-sample Wilcoxon signed rank test, , using Iman (1974) z approximation, , z-split method for equal to hyp. med.

2.3 R

```
> data <- data <- c(1, 2, 5, 1, 1, 5, 3, 1, 5, 1, 1, 5, 1, 1, 3, 3, 3, 4, 2, 4)
> ts_wilcoxon_os(data)
   W statistic   df      pVal                                         testUsed
1 47  1.143943 n.a. 0.2526473 one-sample Wilcoxon signed rank test, with ties correction
> ts_wilcoxon_os(data, ties=FALSE, appr="imanz", eqMed = "zsplit", cc = FALSE)
   W statistic   df      pVal
1 76  1.085116 n.a. 0.2778702
                                                                                               testUsed
1 one-sample Wilcoxon signed rank test, using Iman (1974) z approximation, z-split method for equal to hyp. med.
> 
```

3 Details of Calculations

3.1 The Test Statistic W

$$W_+ = \sum_{i=1}^{n_r^+} r_i^+$$

With

$$r_i = rank(|d_i|)$$

$$d_i = y_i - \theta$$

Symbols

- n_r^+ is the number of ranks with a positive deviation from the hypothesized median
- r_i^+ the i-th rank of the ranks with a positive deviation from the hypothesized median
- θ is the median tested (the hypothesized median).
- y_i is the i-th score of the variable after removing scores that were equal to θ.

3.2 Testing

3.2.1 Exact

$$sig. = 2 \times W(w_{min}, n)$$

With:

$$W(w_{min}, n) = \frac{\sum_{i=0}^{W} srf(i, n)}{2^n}$$

$$w_{min} = MIN(W_-, W_+)$$

$$srf(x, y) = \begin{cases} 0, & x < 0 \\ 0, & x > \binom{y+1}{2} \\ 1, & y = 1\ AND(x = 0\ OR\ x = 1) \\ srf^*(x, y), & y \geq 0 \end{cases}$$

$$srf^*(x, y) = srf(x - y, y - 1) + srf(x, y - 1)$$

$$\binom{a}{b} = \frac{a!}{b!\,(a - b)!}$$

Symbols:

- W_- is the same as W_+ but the sum of the ranks with a negative deviation.
- $srf(x, y)$ is a recursive function determining the permutation distribution.
- n is the number of cases in the analysis
- $\binom{n}{k}$ is the binomial coefficient, the number of possible combinations to choose k out of n, without replacement, and the order is not important.

3.2.2 Approximations

If the sample size is large enough, we can use a normal approximation. What is large enough varies quite per author. A few examples: n > 8 (slideplayer, 2015), n > 15 (SigMaxl, n.d.), n > 20 (Wikipedia, n.d.), n > 25 (Harris & Hardin, 2013), n > 30 (Winthrop, n.d.) .

3.2.2.1 Wilcoxon

The 'standard' method is from Wilcoxon, where the z-statistic is given by:

$$z = \frac{W - \mu_W}{\sigma_W}$$

$$sig. = 2 \times \left(1 - Z(z)\right)$$

With:

$$\mu_W = \frac{n_r(n_r + 1)}{4}$$

$$\sigma_W = \sqrt{\sigma_W^2}$$

$$\sigma_W^2 = \frac{n_r(n_r + 1)(2n_r + 1)}{24}$$

Symbols:

- n_r is the number of ranks used
- k is the number of unique ranks
- $Z(\dots)$ the standard normal cumulative density function.

3.2.2.2 Using t- distribution

Iman (1974, p. 799) also gives an approximation using a t-distribution with:

$$t = \frac{W - \mu_W}{\sqrt{\frac{\sigma_W^2 \times n_r - (W - \mu_W)^2}{n_r - 1}}}$$

And also for when applying a continuity correction:

$$t_c = \frac{W - \mu_W}{\sqrt{\frac{\sigma_W^2 \times n_r - (|W - \mu_W| - 0.5)^2}{n_r - 1}}}$$

The degrees of freedom (df) for this is given by:

$$df = n_r - 1$$

The p-value is given by:

$$sig. = 2 \times \left(1 - T(t, df)\right)$$

Symbol:

- $T(\dots, \dots)$ the Student t distribution cumulative density function.

3.2.2.3 Adjusting z-value

Iman (Iman, 1974, p. 803) also provides a combination of Z and T values resulting in:

$$J = \frac{Z}{2}\left(1 + \sqrt{\frac{n_r - 1}{n_r - Z^2}}\right)$$

The p-value is given by:

$$sig. = 2 \times \left(1 - Z(J)\right)$$

3.3 Ties Correction

In case there are ties (scores that are equal) an adjustment is needed for σ_W^2

$$\sigma_W^{*2} = \sigma_W^2 - A$$

With

$$A = \frac{\sum_{i=1}^{k}\left(t_i^3 - t_i\right)}{48}$$

Symbols

- t_i is the frequency of the i-th unique rank.

3.4 Keep scores equal to hypothesized median

But that's not all. We can also not remove the scores that equal the hypothesized median. In that case there are two options. One by Pratt (1959), and another called the z-split method.

We re-define:

$$d_i = y_i - \theta$$

Where y_i is now simply the i-th score.

3.4.1 z-split method

For the z-split method we only need to re-define:

$$W = \frac{\sum_{i=1}^{n_{d0}} r_i^0}{2} + \sum_{i=1}^{n_r^+} r_i^+$$

Symbols:

- r_i^0 is the rank for a score with zero difference
- n_{d0} is the number of scores equal to the hypothesized median.

Note:

- for the ties correction, the z-split method will include the scores equal to the hypothesized median
- now $n_r = n$, where n is the number of scores

3.4.2 Pratt method

For Pratt (1959) we then get:

$$\mu_W = \frac{n_r \times (n_r - 1) - n_{d0} \times (n_{d0} + 1)}{4}$$

$$\sigma_W^2 = \frac{n_r \times (n_r + 1) \times (2n_r + 1) - n_{d0} \times (n_{d0} + 1)(2 \times n_{d0} + 1)}{24}$$

Symbols

- n_{d0} is the number of scores equal to the hypothesized median.

Notes:

- the ties correction still excludes the ties for the scores that equal the hypothesized median.
- now $n_r = n$, where n is the number of scores

4 Sources

The Wilcoxon test is described in *Individual comparisons by ranking methods* (Wilcoxon, 1945).

The exact test is taken from Algorithm AS 62 (Dinneen & Blakesley, 1973), which is also used by Zaiontz (Zaiontz, n.d.)

Iman's t-approximation:

$$T = \frac{R^+ - n(n+1)/4}{\left[\frac{n^2(n+1)(2n+1)}{24(n-1)} - \frac{[R^+ - n(n+1)/4]^2}{n-1}\right]^{\frac{1}{2}}} . \qquad (2.4)$$

With a correction for continuity the t-statistic becomes

$$T_c = \frac{|R^+ - n(n+1)/4| - .5}{\left[\frac{n^2(n+1)(2n+1)}{24(n-1)} - \frac{[|R^+ - n(n+1)/4| - .5]^2}{n-1}\right]^{\frac{1}{2}}} . \qquad (2.5)$$

The Student's t-distribution with n-1 degrees of freedom is used to obtain the approximate critical values of T_c. The form of the

Iman (1974, p. 799)

Iman's z-approximation:

$$J = \frac{Z}{2}\left[1 + \sqrt{\frac{n-1}{n-Z^2}}\right] . \qquad (3.3)$$

(Iman, 1974, p. 803)

The Pratt (1959) method and z-split method were found in Python's documentation for scipy's Wilcoxon function (scipy, n.d.). They also refer to Cureton (1967) for the Pratt method.

References

Cureton, E. E. (1967). The normal approximation to the signed-rank sampling distribution when zero differences are present. *Journal of the American Statistical Association*, *62*(319), 1068–1069. https://doi.org/10.1080/01621459.1967.10500917

Dinneen, L. C., & Blakesley, B. C. (1973). Algorithm AS 62: A generator for the sampling distribution of the Mann- Whitney U statistic. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *22*(2), 269–273. https://doi.org/10.2307/2346934

Harris, T., & Hardin, J. W. (2013). Exact Wilcoxon Signed-Rank and Wilcoxon Mann–Whitney Ranksum Tests. *The Stata Journal*, *13*(2), 337–343. https://doi.org/10.1177/1536867X1301300208

Iman, R. L. (1974). Use of a t-statistic as an approximation to the exact distribution of the wildcoxon signed ranks test statistic. *Communications in Statistics*, *3*(8), 795–806. https://doi.org/10.1080/03610927408827178

Pratt, J. W. (1959). Remarks on zeros and ties in the Wilcoxon signed rank procedures. *Journal of the American Statistical Association*, *54*(287), 655–667. https://doi.org/10.1080/01621459.1959.10501526

scipy. (n.d.). *Scipy.stats.wilcoxon*. Scipy. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html

SigMaxl. (n.d.). *One Sample Wilcoxon Sign Test Exact*. Retrieved August 30, 2020, from https://www.sigmaxl.com/OneSampleSignWilcoxonExact.shtml

slideplayer. (2015, June 13). *Using statistics to make inferences 6*.

Wikipedia. (n.d.). Wilcoxon signed-rank test. In *Wikipedia*. Retrieved August 30, 2020, from https://en.wikipedia.org/w/index.php?title=Wilcoxon_signed-rank_test&oldid=974561084

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80. https://doi.org/10.2307/3001968

Winthrop. (n.d.). *The Wilcoxon signed rank test for one sample*. Winthrop Univerisy Hospital. https://nyuwinthrop.org/wp-content/uploads/2019/08/wilcoxon-sign-rank-test-one-sample.pdf

Zaiontz, C. (n.d.). *Wilcoxon signed ranks exact test*. Real Statistics Using Excel. Retrieved January 25, 2023, from https://real-statistics.com/non-parametric-tests/wilcoxon-signed-ranks-test/wilcoxon-signed-ranks-exact-test/