signrank — Equality tests on matched data

Description

signrank tests the equality of matched pairs of observations by using the Wilcoxon matched-pairs signed-rank test (Wilcoxon 1945). The null hypothesis is that both distributions are the same.

signtest also tests the equality of matched pairs of observations (Arbuthnott [1710], but better explained by Snedecor and Cochran [1989]) by calculating the differences between \textit{varname} and the expression. The null hypothesis is that the median of the differences is zero; no further assumptions are made about the distributions. This, in turn, is equivalent to the hypothesis that the true proportion of positive (negative) signs is one-half.

For equality tests on unmatched data, see [R] ranksum.

Quick start

Wilcoxon matched-pairs signed-rank test for \textit{v1} and \textit{v2}

\texttt{signrank v1 = v2}

Compute an exact $p$-value for the signed-rank test

\texttt{signrank v1 = v2, exact}

Conduct signed-rank test separately for groups defined by levels of \textit{catvar}

\texttt{by catvar: signrank v1 = v2}

Test that the median of differences between matched pairs \textit{v1} and \textit{v2} is 0

\texttt{signtest v1 = v2}

Menu

\textbf{signrank}

Statistics $>$ Nonparametric analysis $>$ Tests of hypotheses $>$ Wilcoxon matched-pairs signed-rank test

\textbf{signtest}

Statistics $>$ Nonparametric analysis $>$ Tests of hypotheses $>$ Test equality of matched pairs
Syntax

Wilcoxon matched-pairs signed-rank test

signrank varname = exp [if] [in] [, exact]

Sign test of matched pairs

signtest varname = exp [if] [in]

by is allowed with signrank and signtest; see [D] by.

Option for signrank

exact specifies that the exact \( p \)-value be computed in addition to the approximate \( p \)-value. The exact \( p \)-value is based on the actual randomization distribution of the test statistic. The approximate \( p \)-value is based on a normal approximation to the randomization distribution. By default, the exact \( p \)-value is computed for sample sizes \( n \leq 200 \) because the normal approximation may not be precise in small samples. The exact computation can be suppressed by specifying noexact. For sample sizes larger than 200, you must specify exact to compute the exact \( p \)-value. The exact computation is available for sample sizes \( n \leq 2000 \).

Remarks and examples

Example 1: signrank

We are testing the effectiveness of a new fuel additive. We run an experiment with 12 cars. We first run each car without the fuel treatment and measure the mileage. We then add the fuel treatment and repeat the experiment. The results of the experiment are

<table>
<thead>
<tr>
<th>Without treatment</th>
<th>With treatment</th>
<th>Without treatment</th>
<th>With treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>24</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>23</td>
<td>25</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>25</td>
<td>22</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>18</td>
<td>23</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>19</td>
<td>23</td>
</tr>
</tbody>
</table>
We create two variables called \texttt{mpg1} and \texttt{mpg2}, representing mileage without and with the treatment, respectively. We can test the null hypothesis that the treatment had no effect by typing

```
. use https://www.stata-press.com/data/r16/fuel
. signrank mpg1 = mpg2
```

Wilcoxon signed-rank test

<table>
<thead>
<tr>
<th>sign</th>
<th>obs</th>
<th>sum ranks</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>3</td>
<td>13.5</td>
<td>38.5</td>
</tr>
<tr>
<td>negative</td>
<td>8</td>
<td>63.5</td>
<td>38.5</td>
</tr>
<tr>
<td>zero</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

all | 12 | 78 | 78 |

unadjusted variance 162.50
adjustment for ties -1.62
adjustment for zeros -0.25

adjusted variance 160.62

Ho: mpg1 = mpg2

\[
z = -1.973
\]

\[
\text{Prob > } |z| = 0.0485
\]

\[
\text{Exact Prob} = 0.0479
\]

Despite the small sample size, the \textit{p}-value computed using a normal approximation, 0.0485, is similar to the exact \textit{p}-value, 0.0479. These results indicate that we can reject the null hypothesis at a significance level of 0.05.

\section*{Example 2: signtest}

\texttt{signtest} tests that the median of the differences is zero, making no further assumptions, whereas \texttt{signrank} assumed that the distributions are equal as well. Using the data above, we type

```
. signtest mpg1 = mpg2
```

Sign test

<table>
<thead>
<tr>
<th>sign</th>
<th>observed</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>3</td>
<td>5.5</td>
</tr>
<tr>
<td>negative</td>
<td>8</td>
<td>5.5</td>
</tr>
<tr>
<td>zero</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

all | 12 | 12 |

One-sided tests:

- Ho: median of mpg1 - mpg2 = 0 vs. 
- Ha: median of mpg1 - mpg2 > 0
  \[
  \text{Pr}(\#positive >= 3) = \text{Binomial}(n = 11, x >= 3, p = 0.5) = 0.9673
  \]

- Ho: median of mpg1 - mpg2 = 0 vs. 
- Ha: median of mpg1 - mpg2 < 0
  \[
  \text{Pr}(\#negative >= 8) = \text{Binomial}(n = 11, x >= 8, p = 0.5) = 0.1133
  \]

Two-sided test:

- Ho: median of mpg1 - mpg2 = 0 vs. 
- Ha: median of mpg1 - mpg2 \(!=\) 0
  \[
  \text{Pr}(\#positive >= 8 \text{ or } \#negative >= 8) = \min(1, 2*\text{Binomial}(n = 11, x >= 8, p = 0.5)) = 0.2266
  \]

The summary table indicates that there were three comparisons for which \texttt{mpg1} exceeded \texttt{mpg2}, eight comparisons for which \texttt{mpg2} exceeded \texttt{mpg1}, and one comparison for which they were the same.
The \( p \)-values displayed below the summary table are based on a binomial\((n, p = 1/2)\) distribution for the test statistic, which is the number of positive or negative signs. Zeros are ignored. The \( n \) for the test is the number of nonzero differences. Because no approximation is used, the \( p \)-values are “exact” \( p \)-values. The \( p \)-value for the one-sided test, where the alternative hypothesis is that the median of \( \text{mpg}_1 - \text{mpg}_2 \) is smaller than zero, is 0.1133. The \( p \)-value for the two-sided test, where the alternative hypothesis is simply that the median of the differences is different from zero, is \( 0.2266 = 2 \times 0.1133 \).

### Stored results

**signrank** stores the following in \( r() \):

Scalars
- \( r(N) \) sample size
- \( r(N_{\text{pos}}) \) number of positive comparisons
- \( r(N_{\text{neg}}) \) number of negative comparisons
- \( r(N_{\text{tie}}) \) number of tied comparisons
- \( r(z) \) \( z \) statistic
- \( r(\text{Var}_a) \) adjusted variance
- \( r(\text{sum}_\text{pos}) \) sum of the positive ranks
- \( r(\text{sum}_\text{neg}) \) sum of the negative ranks
- \( r(p) \) two-sided \( p \)-value from normal approximation
- \( r(p_{\text{l}}) \) lower one-sided \( p \)-value from normal approximation
- \( r(p_{\text{u}}) \) upper one-sided \( p \)-value from normal approximation
- \( r(p_{\text{exact}}) \) two-sided exact \( p \)-value
- \( r(p_{\text{l_exact}}) \) lower one-sided exact \( p \)-value
- \( r(p_{\text{u_exact}}) \) upper one-sided exact \( p \)-value

**signtest** stores the following in \( r() \):

Scalars
- \( r(N) \) sample size
- \( r(N_{\text{pos}}) \) number of positive comparisons
- \( r(N_{\text{neg}}) \) number of negative comparisons
- \( r(N_{\text{tie}}) \) number of tied comparisons
- \( r(p) \) two-sided \( p \)-value
- \( r(p_{\text{l}}) \) lower one-sided \( p \)-value
- \( r(p_{\text{u}}) \) upper one-sided \( p \)-value

### Methods and formulas

For a practical introduction to these techniques with an emphasis on examples rather than theory, see Bland (2015) or Sprent and Smeeton (2007). For a summary of these tests, see Snedecor and Cochran (1989).

Methods and formulas are presented under the following headings:

- **signrank**
- **signtest**
signrank

Both the sign test and Wilcoxon signed-rank tests test the null hypothesis that the distribution of a random variable $D = \text{varname} - \text{exp}$ has median zero. The sign test makes no additional assumptions, but the Wilcoxon signed-rank test makes the additional assumption that the distribution of $D$ is symmetric. If $D = X_1 - X_2$, where $X_1$ and $X_2$ have the same distribution, then it follows that the distribution of $D$ is symmetric about zero. Thus, the Wilcoxon signed-rank test is often described as a test of the hypothesis that two distributions are the same, that is, $X_1 \sim X_2$.

Let $d_j$ denote the difference for any matched pair of observations,

$$d_j = x_{1j} - x_{2j} = \text{varname} - \text{exp}$$

for $j = 1, 2, \ldots, n$.

Rank the absolute values of the differences, $|d_j|$, and assign any tied values the average rank. Consider the signs of $d_j$, and let

$$r_j = \text{sign}(d_j) \text{ rank}(|d_j|)$$

be the signed ranks. The test statistic is

$$T_{\text{obs}} = \sum_{j=1}^{n} r_j = (\text{sum of ranks for } + \text{ signs}) - (\text{sum of ranks for } - \text{ signs})$$

The distribution of the test statistic is based on Fisher’s principle of randomization (Fisher 1935). Fisher’s idea (stated in a modern way) was to look at a family of transformations of the observed data such that the a priori likelihood (under the null hypothesis) of the transformed data is the same as the likelihood of the observed data. The distribution of the test statistic is then produced by calculating its value for each of the transformed “randomization” datasets, assuming that each dataset is equally likely.

The null hypothesis is that the distribution of $d_j$ is symmetric about 0. Hence, the likelihood is unchanged if we flip signs on the $d_j$. The randomization distribution of our test statistic, $T$, is all of its values resulting from the $2^n$ possible sign changes for the $d_j$. Namely, the distribution is all the $2^n$ possible values of

$$T = \sum_{j=1}^{n} S_j r_j$$

where $r_j$ are the observed signed ranks (considered fixed) and $S_j$ is either +1 or −1. When the exact option is specified (or implied for $n \leq 200$), this distribution is computed using a recursive algorithm whose computational time is proportional to $n^3$. (See Fisher [1935] for the principle of randomization; Wilcoxon, Katti, and Wilcox [1970] for the computation with untied ranks; and Baker and Tilbury [1993] for the general recursive algorithm.)

$p$-values can also be computed using a normal approximation to the randomization distribution. For the randomization distribution, the mean and variance are given by

$$E(T) = 0 \quad \text{and} \quad \text{Var}_{\text{adj}}(T) = \sum_{j=1}^{n} r_j^2$$

The test statistic for the Wilcoxon signed-rank test is often expressed (equivalently) as the sum of the positive signed ranks, $T_+$, where
\[ E(T_+) = \frac{n(n + 1)}{4} \text{ and } \text{Var}_{\text{adj}}(T_+) = \frac{1}{4} \sum_{j=1}^{n} r_j^2 \]

Zeros and ties do not affect the theory above, and the exact variance is still given by the above formula for \( \text{Var}_{\text{adj}}(T_+) \). When \( d_j = 0 \) is observed, \( r_j \) will always be zero in each of the randomization datasets, using \( \text{sign}(0) = 0 \). (This method of handling zeros is based on the theoretical arguments made by Pratt [1959].) When there are ties, averaged ranks are assigned for each group of ties and then treated the same as other ranks.

The “unadjusted variance” reported by \texttt{signrank} is the variance that the randomization distribution would have had if there had been no ties or zeros:

\[ \text{Var}_{\text{unadj}}(T_+) = \frac{1}{4} \sum_{j=1}^{n} j^2 = \frac{n(n + 1)(2n + 1)}{24} \]

The adjustment for zeros is the change in the variance when the ranks for the zeros are signed to make \( r_j = 0 \),

\[ \Delta \text{Var}_{\text{zero adj}}(T_+) = -\frac{1}{4} \sum_{j=1}^{n_0} j^2 = -\frac{n_0(n_0 + 1)(2n_0 + 1)}{24} \]

where \( n_0 \) is the number of zeros. The adjustment for ties is the change in the variance when the ranks (for nonzero observations) are replaced by averaged ranks:

\[ \Delta \text{Var}_{\text{ties adj}}(T_+) = \text{Var}_{\text{adj}}(T_+) - \text{Var}_{\text{unadj}}(T_+) - \Delta \text{Var}_{\text{zero adj}}(T_+) \]

A normal approximation is used to calculate

\[ z = \frac{T_+ - E(T_+)}{\sqrt{\text{Var}_{\text{adj}}(T_+)}} \]

\texttt{signtest}

The test statistic for the sign test is the number \( n_+ \) of differences

\[ d_j = x_{1j} - x_{2j} = \text{varname} - \text{exp} \]

greater than zero.

Again, the distribution of the test statistic is based on Fisher’s principle of randomization, which we described above for \texttt{signrank}. For the sign test, the “data” are simply the set of signs of the differences. Under the null hypothesis of the sign test, the probability that \( d_j \) is less than zero is equal to the probability that \( d_j \) is greater than zero. Thus, you can transform the observed signs by flipping any number of them, and the set of signs will have the same likelihood. The \( 2^n \) possible sign changes form the family of randomization datasets. If you have no zeros, this procedure leads to \( n_+ \sim \text{binomial}(n, p = 1/2) \).

But what if some differences are zero? If you do have zeros, changing their signs leaves them as zeros. So, if you observe \( n_0 \) zeros, each of the \( 2^n \) sign-change datasets will also have \( n_0 \) zeros. Hence, the values of \( n_+ \) calculated over the sign-change datasets range from 0 to \( n - n_0 \), and the “randomization” distribution of \( n_+ \) is \( \text{binomial}(n - n_0, p = 1/2) \).
The work of Arbuthnott (1710) and later eighteenth-century contributions is discussed by Hald (2003, chap. 17).

Frank Wilcoxon (1892–1965) was born in Ireland to American parents. After working in various occupations (including merchant seaman, oil-well pump attendant, and tree surgeon), he settled in chemistry, gaining degrees from Rutgers and Cornell and employment from various companies. Working mainly on the development of fungicides and insecticides, Wilcoxon became interested in statistics in 1925 and made several key contributions to nonparametric methods. After retiring from industry, he taught statistics at Florida State until his death.

References


Also see

[R] ranksum — Equality tests on unmatched data

[R] ttest — t tests (mean-comparison tests)