Title

signrank — Equality tests on matched data

Syntax

Wilcoxon matched-pairs signed-ranks test

signrank *varname* = *exp* [*if*] [*in*]

Sign test of matched pairs

signtest *varname* = *exp* [*if*] [*in*]

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

Menu

signrank

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

signtest

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

Description

signrank tests the equality of matched pairs of observations by using the Wilcoxon matched-pairs signed-ranks 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 *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**.

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

Without treatment	With treatment	Without treatment	With treatment
20	24	18	17
23	25	24	28
21	21	20	24
25	22	24	27
18	23	23	21
17	18	19	23

We create two variables called mpg1 and mpg2, representing mileage without and with the treatment, respectively. We can test the null hypothesis that the treatment had no effect by typing

```
. use http://www.stata-press.com/data/r13/fuel
. signrank mpg1=mpg2
Wilcoxon signed-rank test
```

sign	obs	sum ranks	expected
positive	3	13.5	38.5
negative	8	63.5	38.5
zero	1	1	1
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
    Prob > |z| =    0.0485
```

The output indicates that we can reject the null hypothesis at any level above 4.85%.   ◁

▷ Example 2: signtest

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

```
. signtest mpg1=mpg2
Sign test
```

sign	observed	expected
positive	3	5.5
negative	8	5.5
zero	1	1
all	12	12

```
One-sided tests:
  Ho: median of mpg1 - mpg2 = 0 vs.
  Ha: median of mpg1 - mpg2 > 0
      Pr(#positive >= 3) =
          Binomial(n = 11, x >= 3, p = 0.5) =  0.9673

  Ho: median of mpg1 - mpg2 = 0 vs.
  Ha: median of mpg1 - mpg2 < 0
      Pr(#negative >= 8) =
          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
      Pr(#positive >= 8 or #negative >= 8) =
          min(1, 2*Binomial(n = 11, x >= 8, p = 0.5)) =  0.2266
```

The summary table indicates that there were three comparisons for which mpg1 exceeded mpg2, eight comparisons for which mpg2 exceeded mpg1, and one comparison for which they were the same.

The output below the summary table is based on the binomial distribution. The significance of the one-sided test, where the alternative hypothesis is that the median of mpg2 − mpg1 is greater than zero, is 0.1133. The significance of 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_neg)	number of negative comparisons	r(sum_neg)	sum of the negative ranks
r(N_pos)	number of positive comparisons	r(z)	z statistic
r(N_tie)	number of tied comparisons	r(Var_a)	adjusted variance
r(sum_pos)	sum of the positive ranks		

signtest stores the following in r():

Scalars
r(N_neg)	number of negative comparisons	r(p_2)	two-sided probability
r(N_pos)	number of positive comparisons	r(p_neg)	one-sided probability of negative comparison
r(N_tie)	number of tied comparisons	r(p_pos)	one-sided probability of positive comparison

Methods and formulas

For a practical introduction to these techniques with an emphasis on examples rather than theory, see Bland (2000) 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 $D = varname - exp$ 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,

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

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 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, and thus the randomization datasets are the 2^n possible sign changes for the d_j. Thus the randomization distribution of our test statistic T can be computed by considering all the 2^n possible values of

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

where the r_j are the observed signed ranks (considered fixed) and S_j is either $+1$ or -1.

With this distribution, the mean and variance of T 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

$$E(T_+) = \frac{n(n+1)}{4} \quad \text{and} \quad \text{Var}_{\text{adj}}(T_+) = \frac{1}{4} \sum_{j=1}^{n} r_j^2$$

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, d_j will always be zero in each of the randomization datasets (using $\text{sign}(0) = 0$). When there are ties, you can assign averaged ranks for each group of ties and then treat them the same as the other ranks.

The "unadjusted variance" reported by `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

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

signtest

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

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

greater than zero. Assuming that the probability of a difference being equal to zero is exactly zero, then, under the null hypothesis, $n_+ \sim \text{binomial}(n, p = 1/2)$, where n is the total number of observations.

But what if some differences are zero? This question has a ready answer if you view the test from the perspective of 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.

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 again leads to $n_+ \sim \text{binomial}(n, p = 1/2)$.

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 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

Arbuthnott, J. 1710. An argument for divine providence, taken from the constant regularity observed in the births of both sexes. *Philosophical Transaction of the Royal Society of London* 27: 186–190.

Bland, M. 2000. *An Introduction to Medical Statistics.* 3rd ed. Oxford: Oxford University Press.

Bradley, R. A. 2001. Frank Wilcoxon. In *Statisticians of the Centuries*, ed. C. C. Heyde and E. Seneta, 420–424. New York: Springer.

Fisher, R. A. 1935. *The Design of Experiments.* Edinburgh: Oliver & Boyd.

Hald, A. 2003. *A History of Probability and Statistics and Their Applications before 1750.* New York: Wiley.

Harris, T., and J. W. Hardin. 2013. Exact Wilcoxon signed-rank and Wilcoxon Mann–Whitney ranksum tests. *Stata Journal* 13: 337–343.

Kaiser, J. 2007. An exact and a Monte Carlo proposal to the Fisher–Pitman permutation tests for paired replicates and for independent samples. *Stata Journal* 7: 402–412.

Newson, R. B. 2006. Confidence intervals for rank statistics: Somers' D and extensions. *Stata Journal* 6: 309–334.

Snedecor, G. W., and W. G. Cochran. 1989. *Statistical Methods.* 8th ed. Ames, IA: Iowa State University Press.

Sprent, P., and N. C. Smeeton. 2007. *Applied Nonparametric Statistical Methods.* 4th ed. Boca Raton, FL: Chapman & Hall/CRC.

Sribney, W. M. 1995. crc40: Correcting for ties and zeros in sign and rank tests. *Stata Technical Bulletin* 26: 2–4. Reprinted in *Stata Technical Bulletin Reprints*, vol. 5, pp. 5–8. College Station, TX: Stata Press.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics* 1: 80–83.

Also see

[R] **ranksum** — Equality tests on unmatched data

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