The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*, *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$ 98	1-year subscription	$138
2-year subscription	$165	2-year subscription	$245
3-year subscription	$225	3-year subscription	$345
1-year student subscription	$ 75	1-year student subscription	$ 99
1-year university library subscription	$125	1-year university library subscription	$165
2-year university library subscription	$215	2-year university library subscription	$295
3-year university library subscription	$315	3-year university library subscription	$435
1-year institutional subscription	$245	1-year institutional subscription	$285
2-year institutional subscription	$445	2-year institutional subscription	$525
3-year institutional subscription	$645	3-year institutional subscription	$765
Electronic only		**Electronic only**	
1-year subscription	$ 75	1-year subscription	$ 75
2-year subscription	$125	2-year subscription	$125
3-year subscription	$165	3-year subscription	$165
1-year student subscription	$ 45	1-year student subscription	$ 45

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, STATA, Stata Press, Mata, MATA, and NetCourse are registered trademarks of StataCorp LP.

The Stata Journal (2013)
13, Number 2, pp. 337–343

Exact Wilcoxon signed-rank and Wilcoxon Mann–Whitney ranksum tests

Tammy Harris
Institute for Families in Society
Department of Epidemiology & Biostatistics
University of South Carolina
Columbia, SC
harris68@mailbox.sc.edu

James W. Hardin
Institute for Families in Society
Department of Epidemiology & Biostatistics
University of South Carolina
Columbia, SC
jhardin@sc.edu

Abstract. We present new Stata commands for carrying out exact Wilcoxon one-sample and two-sample comparisons of the median. Nonparametric tests are often used in clinical trials, in which it is not uncommon to have small samples. In such situations, researchers are accustomed to making inferences by using exact statistics. The `ranksum` and `signrank` commands in Stata provide only asymptotic results, which assume normality. Because large-sample results are unacceptable in many clinical trials studies, these researchers must use other software packages. To address this, we have developed new commands for Stata that provide exact statistics in small samples. Additionally, when samples are large, we provide results based on the Student's t distribution that outperform those based on the normal distribution.

Keywords: st0297, ranksumex, signrankex, exact distributions, nonparametric tests, median, Wilcoxon matched-pairs signed-rank test, Wilcoxon ranksum test

1 Introduction

Many statistical analysis methods are derived after making an assumption about the underlying distribution of the data (for example, normality). However, one may also consider nonparametric methods from which to draw statistical inferences where no assumptions are made about an underlying population or distribution. For the nonparametric equivalents to the parametric one-sample and two-sample t tests, the Wilcoxon signed-rank test (one sample) is used to test the hypothesis that the median difference between the absolute values of positive and negative paired differences is 0. The Wilcoxon Mann–Whitney ranksum test is used to test the hypothesis of a zero-median difference between two independently sampled populations.

We present Stata commands to evaluate both of these nonparametric statistical tests. This article is organized as follows. In section 2, we review the test statistics. In section 3, Stata syntax is presented for the new commands, followed by examples in section 4. A final summary is presented in section 5.

2 Nonparametric Wilcoxon tests

2.1 Wilcoxon signed-rank test

Let X_i and Y_i be continuous paired random variables from data consisting of n observations, where observations are denoted as $\mathbf{X} = (X_1, \ldots, X_n)^T$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$. For these paired bivariate data, $(x_1, y_1), \ldots, (x_n, y_n)$, the differences are calculated as $D_i = Y_i - X_i$. We omit consideration of the subset of observations for which the absolute difference is 0. From this one sample of $n_r \leq n$ nonzero differences, ranks (r_i) are applied to the absolute differences $|D_i|$, where rank 1 is the smallest absolute difference and rank n_r is the largest absolute difference. Before assigning ranks, we omit absolute differences of 0, $D_i = 0$.

We then test the hypothesis that X_i and Y_i are distributed interchangeably by using the signed-rank test statistic,

$$S = \sum_{i=1}^{n_r} r_i I(D_i > 0) - \frac{n_r(n_r+1)}{4}$$

where $I(D_i > 0)$ is an indicator function that the ith difference is positive. Ranks of tied absolute differences are averaged for the relevant set of observations. The variance of S is given by

$$V = \frac{1}{24} n_r(n_r+1)(2n_r+1) - \frac{1}{48} \sum_{j}^{m} t_j(t_j+1)(t_j-1)$$

where t_j is the number of values tied in absolute value for the jth rank (Lehmann 1975) out of the m unique assigned ranks; $m = n_r$ and $t_j = 1 \; \forall j$ if there are no ties. The significance of S is then computed one of two ways, contingent on sample size (n_r). If $n_r > 25$, the significance of S can be based on the normal approximation (as is done in Stata's `signrank` command) or on Student's t distribution,

$$S\sqrt{\frac{n_r - 1}{n_r V - S^2}}$$

with $n_r - 1$ degrees of freedom (Iman 1974). When $n_r \leq 25$, the significance of S is computed from the exact distribution.

An algorithm for calculation of associated probabilities is the network algorithm of Mehta and Patel (1986). Many new improvements and modifications of that algorithm have been implemented in various applications to compute the exact p-value. Some include polynomial time algorithms for permutation distributions (Pagano and Tritchler

1983), Mann–Whitney-shifted fast Fourier transform (FFT) (Nagarajan and Keich 2009), and decreased computation time for the network algorithm described in Requena and Martín Ciudad (2006). Comprehensive summaries for exact inference methods are published in Agresti (1992) and Waller, Turnbull, and Hardin (1995).

2.2 Wilcoxon Mann–Whitney ranksum test

Let X be a binary variable (group 1 and group 2) and Y_n be a continuous random variable from data consisting of n observations where $\mathbf{Y} = (Y_1, \dots, Y_n)^T$. Ranks are assigned to the data, 1 to n, smallest to largest, where tied ranks are given the average of the ranks. If $n > 25$, the (asymptotically normal) test statistic Z is given by

$$Z = \frac{R_1 - n_1(n+1)/2}{\sqrt{n_1 n_2 V_R / n}}$$

where R_1 is the sum of the ranks from group 1, n_1 is the sample size of group 1, n_2 is the sample size of group 2, and V_R is the variance of the ranks. In Stata, group 1 is lesser in numeric value than group 2. However, if $n \leq 25$, the normal approximation is not appropriate. In this situation, we calculate the exact test by using the approach outlined in the following section.

2.3 An exact method based on the characteristic function

Pagano and Tritchler (1983) present the basic methodology for computing distribution functions through Fourier analysis of the characteristic function. Superficially, this approach appears as complicated as the complete enumeration of results for the distributions of the Wilcoxon test statistics, but Fourier analysis via the FFT in the approach based on the characteristic function is calculated much faster.

Basically, if X is a discrete random variable with a distribution function given by $P(X = x) = p_j$ for $j = 0, \dots, U$, then the complex valued characteristic function is given by

$$\phi(\theta) = \sum_{j=0}^{U} p_j \exp(ij\theta)$$

where $i = \sqrt{-1}$ and $\theta \in [0, 2\pi)$. Because X is defined on a finite integer lattice, the basic theorem in Fourier series is used to obtain the probabilities p_j. For any integer $Q > U$ and $j = 0, \dots, U$,

$$p_j = \frac{1}{Q} \sum_{k=0}^{Q-1} \phi\left(\frac{2\pi k}{Q}\right) \exp\left(-\frac{2\pi ijk}{Q}\right) \tag{1}$$

Thus knowing the characteristic function at Q equidispersed points on the interval $[0, 2\pi)$ is equivalent to knowing it everywhere. Furthermore, the probabilities of the distribution are easily obtained from the characteristic function. We emphasize that the imaginary part of (1) is 0.

To allow tied ranks in the commands, we multiply all ranks by L to ensure that the ranks and sums of ranks will be integers. This can be accomplished for our two statistics by setting $L = 2$. The ranges of the values of the two statistics are easily calculated so that we may choose $Q \geq U$. Defining U as the largest possible value of our statistic (formed from the largest possible ranks), we can choose $\log_2 Q = \text{ceiling}\{\log_2(U)\}$. We choose Q to be a power of 2 because of the requirements of the FFT algorithm in Stata (Fourier analysis is carried out by using the Mata `fft` command).

Using r_k to denote the rank of the kth observation, the characteristic function for the one-sample statistic S_1 is given by

$$\phi_1\left(-2\pi ij/Q\right) = \left\{\exp(-2\pi ij/Q)\prod_{k=1}^{N}\cos(-2\pi jLr_k/Q)\right\}$$

while the characteristic function for S_2 is calculated by using the difference equation

$$\phi_2(j,k) = \exp(-2\pi ijLr_k/Q)\phi_2(j-1,k-1) + \phi_2(j,k-1)$$

3 Stata syntax

Software accompanying this article includes the command files as well as supporting files for dialogs and help. Equivalent to the `signrank` command, the basic syntax for the new Wilcoxon signed-rank test command is

`signrankex` *varname* `=` *exp* [*if*] [*in*]

Equivalent to the `ranksum` command, the basic syntax for the new Wilcoxon Mann–Whitney ranksum test command is

`ranksumex` *varname* [*if*] [*in*], `by(`*groupvar*`)` [`porder`]

4 Example

In this section, we present real-world examples with the new nonparametric Wilcoxon test commands. In clinical trials, talinolol is used as a β blocker and is controlled by P–glycoprotein, which protects xenobiotic compounds. Eight healthy men between the ages of 22 and 26 were evaluated based on their serum-concentration time profiles of talinolol with kinetic profile differences. These differences were two enantiomers, S(−) talinolol and R(+) talinolol. The trial examined single intravenous (`iv`) and repeated oral talinolol profiles before and after rifampicin comedication. Area under the serum concentration time curves (AUC) was collected for each subject (see Zschiesche et al. [2002]). We compare AUC values of S(−) `iv` talinolol before and after comedication of rifampicin by using the Wilcoxon signed-rank test. The results are given below, where S is the Wilcoxon signed-rank test statistic.

```
. use signrank, clear
. signrankex iv_s_before = iv_s_after
Wilcoxon signed-rank test
        sign │      obs   sum ranks    expected
─────────────┼─────────────────────────────────
    positive │        8          36          18
    negative │        0           0          18
        zero │        0
─────────────┼─────────────────────────────────
         all │        8          36          36
Ho: iv_s_before = iv_s_after
             S =  18.000
    Prob >= |S| =   0.0078
```

The results show there was a statistically significant difference (p-value = 0.0078) between `iv` S(−) talinolol before and after comedication of rifampicin. There were greater S(−) talinolol AUC values shown before rifampicin administration than after.

For the Wilcoxon Mann–Whitney ranksum test example, we will use performance data (table 1) collected on rats' rotarod endurance (in seconds) from two treatment groups. The rats were randomly selected to be in the control group (received saline solvent) or the treatment group (received centrally acting muscle relaxant) (Bergmann, Ludbrook, and Spooren 2000).

Table 1. Rotarod endurance

Treatment group		Control group	
Endurance time (sec)	Rank	Endurance time (sec)	Rank
22	2	300	15
300	15	300	15
75	3	300	15
271	5	300	15
300	15	300	15
18	1	300	15
300	15	300	15
300	15	300	15
163	4	300	15
300	15	300	15
300	15	300	15
300	15	300	15

The results are given below.

```
. use ranksum, clear
. ranksumex edrce, by(trt)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
         trt │      obs    rank sum    expected
─────────────┼──────────────────────────────────
           0 │       12         180         150
           1 │       12         120         150
─────────────┼──────────────────────────────────
    combined │       24         300         300
Exact statistics
Ho: edrce(trt==0) = edrce(trt==1)
    Prob <=       120 = 0.0186
    Prob >=       180 = 0.0186
    Two-sided p-value = 0.0373
```

The two-sided exact p-value of 0.0373 exhibits a statistically significant difference in average rotarod endurance between the groups of rats. We can also illustrate how to calculate this exact p-value manually by using the rat rotarod endurance data (table 1). In Conover (1999), the Wilcoxon Mann–Whitney ranksum test exact p-value is illustrated in terms of combinations (arrangements) of ranks. In this example, the number of arrangements of 12 of the ranks in the table having a sum less than or equal to 120 is the number of arrangements of choosing all 5 of the ranks less than 15 and 7 of the 19 tied ranks of 15; this is given by $\binom{5}{5}\binom{19}{7}$. The total number of ways to choose 12 of 24 ranks is given by $\binom{24}{12}$. Thus the p-value is

$$p\text{-value} = \frac{\binom{5}{5}\binom{19}{7}}{\binom{24}{12}} = \frac{50{,}388}{2{,}704{,}156} = 0.0186$$

where each of the new commands returns the p-value as well as the numerator and denominator of the exact fraction (see the return values in the previous example).

5 Summary

In this article, we introduced two supporting Stata commands for the exact nonparametric Wilcoxon signed-rank test and the Wilcoxon Mann–Whitney ranksum test. These one-sample and two-sample test statistics can be used to assess the difference in location (median difference) for small samples (exact distribution) and larger samples (Student's t distribution).

6 References

Agresti, A. 1992. A survey of exact inference for contingency tables. *Statistical Science* 7: 131–153.

Bergmann, R., J. Ludbrook, and W. P. J. M. Spooren. 2000. Different outcomes of the Wilcoxon–Mann–Whitney test from different statistics packages. *American Statistician* 54: 72–77.

Conover, W. J. 1999. *Practical Nonparametric Statistics*. 3rd ed. New York: Wiley.

Iman, R. L. 1974. Use of a t-statistic as an approximation to the exact distribution of the Wilcoxon signed ranks test statistic. *Communications in Statistics* 3: 795–806.

Lehmann, E. L. 1975. *Nonparametrics: Statistical Methods Based on Ranks*. Upper Saddle River, NJ: Springer.

Mehta, C. R., and N. R. Patel. 1986. Algorithm 643: FEXACT: A FORTRAN subroutine for Fisher's exact test on unordered $r \times c$ contingency tables. ACM *Transactions on Mathematical Software* 12: 154–161.

Nagarajan, N., and U. Keich. 2009. Reliability and efficiency of algorithms for computing the significance of the Mann–Whitney test. *Computational Statistics* 24: 605–622.

Pagano, M., and D. Tritchler. 1983. On obtaining permutation distributions in polynomial time. *Journal of the American Statistical Association* 78: 435–440.

Requena, F., and N. Martín Ciudad. 2006. A major improvement to the network algorithm for Fisher's exact test in $2 \times c$ contingency tables. *Computational Statistics and Data Analysis* 51: 490–498.

Waller, L. A., B. W. Turnbull, and J. M. Hardin. 1995. Obtaining distribution functions by numerical inversion of characteristic functions with applications. *American Statistician* 49: 346–350.

Zschiesche, M., G. L. Lemma, K.-J. Klebingat, G. Franke, B. Terhaag, A. Hoffmann, T. Gramatté, H. K. Kroemer, and W. Siegmund. 2002. Stereoselective disposition of talinolol in man. *Journal of Pharmaceutical Sciences* 91: 303–311.

About the authors

Tammy Harris is a PhD candidate in the Department of Epidemiology and Biostatistics and an affiliated researcher in the Institute for Families in Society at the University of South Carolina in Columbia, SC.

James W. Hardin is an associate professor in the Department of Epidemiology and Biostatistics and an affiliated faculty member in the Institute for Families in Society at the University of South Carolina in Columbia, SC.