Title

ranksum — Equality tests on unmatched data

Description

`ranksum` tests the hypothesis that two independent samples (that is, *unmatched* data) are from populations with the same distribution by using the Wilcoxon rank-sum test, which is also known as the Mann–Whitney two-sample statistic (Wilcoxon 1945; Mann and Whitney 1947).

`median` performs a nonparametric *k*-sample test on the equality of medians. It tests the null hypothesis that the *k* samples were drawn from populations with the same median. For two samples, the chi-squared test statistic is computed both with and without a continuity correction.

`ranksum` and `median` are for use with *unmatched* data. For equality tests on matched data, see [R] **signrank**.

Quick start

Wilcoxon rank-sum test

Test for equality of distributions of `v` over two groups defined by the levels of `catvar1`

```
ranksum v, by(catvar1)
```

As above, and estimate probability that a case from the first level of `catvar1` has a greater value of `v` than a case from the second level of `catvar1`

```
ranksum v, by(catvar1) porder
```

Nonparametric equality-of-medians test

Equality of medians test for `v` over two or more groups defined by the levels of `catvar2`

```
median v, by(catvar2)
```

Also report Fisher's exact test

```
median v, by(catvar2) exact
```

As above, but split cases at the median evenly between the above and below groups

```
median v, by(catvar2) exact medianties(split)
```

Menu

ranksum

Statistics > Nonparametric analysis > Tests of hypotheses > Wilcoxon rank-sum test

median

Statistics > Nonparametric analysis > Tests of hypotheses > K-sample equality-of-medians test

Syntax

Wilcoxon rank-sum test

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

Nonparametric equality-of-medians test

median *varname* [*if*] [*in*] [*weight*] , by(*groupvar*) [*median_options*]

ranksum_options	Description
Main	
* by(*groupvar*)	grouping variable
porder	probability that variable for first group is larger than variable for second group

median_options	Description
Main	
* by(*groupvar*)	grouping variable
exact	perform Fisher's exact test
medianties(below)	assign values equal to the median to below group
medianties(above)	assign values equal to the median to above group
medianties(drop)	drop values equal to the median from the analysis
medianties(split)	split values equal to the median equally between the two groups

* by(*groupvar*) is required.

by is allowed with ranksum and median; see [D] **by**.

fweights are allowed with median; see [U] **11.1.6 weight**.

Options for ranksum

Main

by(*groupvar*) is required. It specifies the name of the grouping variable.

porder displays an estimate of the probability that a random draw from the first population is larger than a random draw from the second population.

Options for median

Main

by(*groupvar*) is required. It specifies the name of the grouping variable.

exact displays the significance calculated by Fisher's exact test. For two samples, both one- and two-sided probabilities are displayed.

`medianties(below|above|drop|split)` specifies how values equal to the overall median are to be handled. The median test computes the median for *varname* by using all observations and then divides the observations into those falling above the median and those falling below the median. When values for an observation are equal to the sample median, they can be dropped from the analysis by specifying `mediantes(drop)`; added to the group above or below the median by specifying `mediantes(above)` or `mediantes(below)`, respectively; or if there is more than 1 observation with values equal to the median, they can be equally divided into the two groups by specifying `mediantes(split)`. If this option is not specified, `mediantes(below)` is assumed.

Remarks and examples

stata.com

▷ Example 1

We are testing the effectiveness of a new fuel additive. We run an experiment with 24 cars: 12 cars with the fuel treatment and 12 cars without. We input these data by creating a dataset with 24 observations. `mpg` records the mileage rating, and `treat` records 0 if the mileage corresponds to untreated fuel and 1 if it corresponds to treated fuel.

```
. use http://www.stata-press.com/data/r15/fuel2
. ranksum mpg, by(treat)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
        treat |      obs    rank sum    expected
-------------+---------------------------------
    untreated |       12         128         150
      treated |       12         172         150
-------------+---------------------------------
     combined |       24         300         300
unadjusted variance      300.00
adjustment for ties       -4.04
                     ----------
adjusted variance        295.96
Ho: mpg(treat==untreated) = mpg(treat==treated)
             z =  -1.279
    Prob > |z| =   0.2010
```

These results indicate that the medians are not statistically different at any level smaller than 20.1%. Similarly, the median test,

```
. median mpg, by(treat) exact
Median test
  Greater |  whether car received
 than the |     fuel additive
   median | untreated    treated |     Total
----------+----------------------+----------
       no |         7          5 |        12
      yes |         5          7 |        12
----------+----------------------+----------
    Total |        12         12 |        24
          Pearson chi2(1) =   0.6667   Pr = 0.414
           Fisher's exact =                 0.684
   1-sided Fisher's exact =                 0.342
   Continuity corrected:
          Pearson chi2(1) =   0.1667   Pr = 0.683
```

fails to reject the null hypothesis that there is no difference between the fuel with the additive and the fuel without the additive.

Compare these results from these two tests with those obtained from the `signrank` and `signtest` where we found significant differences; see [R] **signrank**. An experiment run on 24 different cars is not as powerful as a before-and-after comparison using the same 12 cars.

◁

Stored results

`ranksum` stores the following in `r()`:

Scalars

`r(N_1)`	sample size n_1
`r(N_2)`	sample size n_2
`r(z)`	z statistic
`r(Var_a)`	adjusted variance
`r(group1)`	value of variable for first group
`r(sum_obs)`	actual sum of ranks for first group
`r(sum_exp)`	expected sum of ranks for first group
`r(porder)`	probability that draw from first population is larger than draw from second population

`median` stores the following in `r()`:

Scalars

`r(N)`	sample size
`r(chi2)`	Pearson's χ^2
`r(p)`	significance of Pearson's χ^2
`r(p_exact)`	Fisher's exact p
`r(groups)`	number of groups compared
`r(chi2_cc)`	continuity-corrected Pearson's χ^2
`r(p_cc)`	continuity-corrected significance
`r(p1_exact)`	one-sided Fisher's exact p

Methods and formulas

For a practical introduction to these techniques with an emphasis on examples rather than theory, see Acock (2016), 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:

ranksum
median

ranksum

For the Wilcoxon rank-sum test, there are two independent random variables, X_1 and X_2, and we test the null hypothesis that $X_1 \sim X_2$. We have a sample of size n_1 from X_1 and another of size n_2 from X_2.

The data are then ranked without regard to the sample to which they belong. If the data are tied, averaged ranks are used. Wilcoxon's test statistic (1945) is the sum of the ranks for the observations in the first sample:

$$T = \sum_{i=1}^{n_1} R_{1i}$$

Mann and Whitney's U statistic (1947) is the number of pairs (X_{1i}, X_{2j}) such that $X_{1i} > X_{2j}$. These statistics differ only by a constant:

$$U = T - \frac{n_1(n_1+1)}{2}$$

Again Fisher's principle of randomization provides a method for calculating the distribution of the test statistic, ties or not. The randomization distribution consists of the $\binom{n}{n_1}$ ways to choose n_1 ranks from the set of all $n = n_1 + n_2$ ranks and assign them to the first sample.

It is a straightforward exercise to verify that

$$E(T) = \frac{n_1(n+1)}{2} \qquad \text{and} \qquad \text{Var}(T) = \frac{n_1 n_2 s^2}{n}$$

where s is the standard deviation of the combined ranks, r_i, for both groups:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(r_i - \overline{r})^2$$

This formula for the variance is exact and holds both when there are no ties and when there are ties and we use averaged ranks. (Indeed, the variance formula holds for the randomization distribution of choosing n_1 numbers from any set of n numbers.)

Using a normal approximation, we calculate

$$z = \frac{T - E(T)}{\sqrt{\text{Var}(T)}}$$

When the `porder` option is specified, the probability

$$p = \frac{U}{n_1 n_2}$$

is computed.

❑ Technical note

We follow the great majority of the literature in naming these tests for Wilcoxon, Mann, and Whitney. However, they were independently developed by several other researchers in the late 1940s and early 1950s. In addition to Wilcoxon, Mann, and Whitney, credit is due to Festinger (1946), Whitfield (1947), Haldane and Smith (1947), and Van der Reyden (1952). Leon Festinger (1919–1989), John Burdon Sanderson Haldane (1892–1964), and Cedric Austen Bardell Smith (1917–2002) are well known for other work, but little seems to be known about Whitfield or van der Reyden. For a detailed study, including information on these researchers, see Berry, Mielke, and Johnston (2012).
❑

median

The median test examines whether it is likely that two or more samples came from populations with the same median. The null hypothesis is that the samples were drawn from populations with the same median. The alternative hypothesis is that at least one sample was drawn from a population with a different median. The test should be used only with ordinal or interval data.

Assume that there are score values for k independent samples to be compared. The median test is performed by first computing the median score for all observations combined, regardless of the sample group. Each score is compared with this computed grand median and is classified as being above the grand median, below the grand median, or equal to the grand median. Observations with scores equal to the grand median can be dropped, added to the "above" group, added to the "below" group, or split between the two groups.

Once all observations are classified, the data are cast into a $2 \times k$ contingency table, and a Pearson's chi-squared test or Fisher's exact test is performed.

Henry Berthold Mann (1905–2000) was born in Vienna, Austria, where he completed a doctorate in algebraic number theory. He moved to the United States in 1938 and for several years made his livelihood by tutoring in New York. During this time, he proved a celebrated conjecture in number theory and studied statistics at Columbia with Abraham Wald, with whom he wrote three papers. After the war, he taught at Ohio State and the Universities of Wisconsin and Arizona. In addition to his work in number theory and statistics, he made major contributions to algebra and combinatorics.

Donald Ransom Whitney (1915–2007) studied at Oberlin, Princeton, and Ohio State Universities and worked at the latter throughout his career. His PhD thesis under Henry Mann was on nonparametric statistics. It was this work that produced the test that bears their names.

References

Acock, A. C. 2016. *A Gentle Introduction to Stata*. 5th ed. College Station, TX: Stata Press.

Berry, K. J., P. W. Mielke, Jr., and J. E. Johnston. 2012. The two-sample rank-sum test: Early development. *Electronic Journal for History of Probability and Statistics* 8: 1–26. http://www.jehps.net/decembre2012/BerryMielkeJohnston.pdf.

Bland, M. 2015. *An Introduction to Medical Statistics*. 4th ed. Oxford: Oxford University Press.

Conroy, R. M. 2012. What hypotheses do nonparametric two-group tests actually test? *Stata Journal* 12: 182–190.

Feiveson, A. H. 2002. Power by simulation. *Stata Journal* 2: 107–124.

Festinger, L. 1946. The significance of difference between means without reference to the frequency distribution function. *Psychometrika* 11: 97–105.

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

Goldstein, R. 1997. sg69: Immediate Mann–Whitney and binomial effect-size display. *Stata Technical Bulletin* 36: 29–31. Reprinted in *Stata Technical Bulletin Reprints*, vol. 6, pp. 187–189. College Station, TX: Stata Press.

Haldane, J. B. S., and C. A. B. Smith. 1947. A simple exact test for birth-order effect. *Annals of Human Genetics* 14: 117–124.

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

Kruskal, W. H. 1957. Historical notes on the Wilcoxon unpaired two-sample test. *Journal of the American Statistical Association* 52: 356–360.

Mann, H. B., and D. R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18: 50–60.

Newson, R. B. 2000a. snp15: somersd—Confidence intervals for nonparametric statistics and their differences. *Stata Technical Bulletin* 55: 47–55. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 312–322. College Station, TX: Stata Press.

———. 2000b. snp15.1: Update to somersd. *Stata Technical Bulletin* 57: 35. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 322–323. College Station, TX: Stata Press.

———. 2000c. snp15.2: Update to somersd. *Stata Technical Bulletin* 58: 30. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, p. 323. College Station, TX: Stata Press.

———. 2001. snp15.3: Update to somersd. *Stata Technical Bulletin* 61: 22. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, p. 324. College Station, TX: Stata Press.

———. 2003. snp15_4: Software update for somersd. *Stata Journal* 3: 325.

———. 2005. snp15_5: Software update for somersd. *Stata Journal* 5: 470.

Perkins, A. M. 1998. snp14: A two-sample multivariate nonparametric test. *Stata Technical Bulletin* 42: 47–49. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, pp. 243–245. College Station, TX: Stata Press.

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.

Van der Reyden, D. 1952. A simple statistical significance test. *Rhodesia Agricultural Journal* 49: 96–104.

Whitfield, J. W. 1947. Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika* 34: 292–296.

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

Also see

[R] **signrank** — Equality tests on matched data

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