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)`

Compute an exact *p*-value for the Wilcoxon rank-sum test

`ranksum v, by(catvar1) exact`

Estimate the 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) [exact porder]
```

Nonparametric equality-of-medians test

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

<table>
<thead>
<tr>
<th>ranksum_options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>* by(groupvar)</td>
<td>grouping variable</td>
</tr>
<tr>
<td>exact</td>
<td>report exact ( p )-value for rank-sum test; by default, exact ( p )-value is computed when total sample size ( \leq 200 )</td>
</tr>
<tr>
<td>porder</td>
<td>probability that variable for first group is larger than variable for second group</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>median_options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>* by(groupvar)</td>
<td>grouping variable</td>
</tr>
<tr>
<td>exact</td>
<td>report ( p )-value from Fisher’s exact test</td>
</tr>
<tr>
<td>medianties(below)</td>
<td>assign values equal to the median to below group</td>
</tr>
<tr>
<td>medianties(above)</td>
<td>assign values equal to the median to above group</td>
</tr>
<tr>
<td>medianties(drop)</td>
<td>drop values equal to the median from the analysis</td>
</tr>
<tr>
<td>medianties(split)</td>
<td>split values equal to the median equally between the two groups</td>
</tr>
</tbody>
</table>

*by(groupvar) is required.

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

`fweight`s are allowed with `median`; see [U] 11.1.6 weight.
Options for ranksum

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

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 $n = n_1 + n_2 \leq 200$ 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 $n \leq 1000$. As the sample size approaches 1,000, the computation takes significantly longer.

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

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

exact displays the $p$-value calculated by Fisher’s exact test. For two samples, both one- and two-sided $p$-values 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 medianties (drop); added to the group above or below the median by specifying medianties (above) or medianties (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 medianties (split). If this option is not specified, medianties (below) is assumed.

Remarks and examples

Examples 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 https://www.stata-press.com/data/r16/fuel2
ranksum mpg, by(treat)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

<table>
<thead>
<tr>
<th>treat</th>
<th>obs</th>
<th>rank sum</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>untreated</td>
<td>12</td>
<td>128</td>
<td>150</td>
</tr>
<tr>
<td>treated</td>
<td>12</td>
<td>172</td>
<td>150</td>
</tr>
<tr>
<td>combined</td>
<td>24</td>
<td>300</td>
<td>300</td>
</tr>
</tbody>
</table>

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
Exact Prob = 0.2117

Because the total sample is only 24 cars, the exact p-value is computed by default. If the sample size were greater than 200, we would have to specify the exact option if we wanted the exact p-value computed.

Despite the small sample size, the p-value computed using a normal approximation, 0.2010, is similar to the exact p-value, 0.2117. These results indicate that the distributions are not statistically different at a 0.05 significance level.

Similarly, the median test,

median mpg, by(treat) exact

Median test

<table>
<thead>
<tr>
<th>Greater than the median</th>
<th>whether car received fuel additive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>untreated</td>
<td>treated</td>
</tr>
<tr>
<td>no</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>yes</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

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)` sample size
- `r(N_1)` sample size of first group
- `r(N_2)` sample size of second group
- `r(z)` z statistic
- `r(Var_a)` adjusted variance
- `r(group1)` value of variable for first group
- `r(sum_obs)` observed sum of ranks for first group
- `r(sum_exp)` expected sum of ranks for first group
- `r(p)` two-sided p-value from normal approximation
- `r(p_l)` lower one-sided p-value from normal approximation
- `r(p_u)` upper one-sided p-value from normal approximation
- `r(p_exact)` two-sided exact p-value
- `r(p_l_exact)` lower one-sided exact p-value
- `r(p_u_exact)` upper one-sided exact p-value
- `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 χ²
- `r(chi2_cc)` continuity-corrected Pearson’s χ²
- `r(groups)` number of groups compared
- `r(p)` p-value for Pearson’s χ² test
- `r(p_cc)` continuity-corrected p-value
- `r(p_exact)` Fisher’s exact p-value
- `r(p1_exact)` one-sided Fisher’s exact p-value

Methods and formulas

For a practical introduction to these techniques with an emphasis on examples rather than theory, see Acock (2018), 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:

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

Fisher’s principle of randomization provides a method for calculating the distribution of the test statistic. The randomization distribution consists of all the possible values of $T$ resulting from the $\binom{n}{n_1}$ ways to choose $n_1$ ranks from the set of all $n = n_1 + n_2$ observed ranks (untied or tied) and assign them to the first sample. 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 $n^4$. (See Fisher [1935] for the principle of randomization; Wilcoxon, Katti, and Wilcox [1970] for the computation with untied ranks; and Hill and Peto [1971] for the general recursive algorithm.)

$p$-values can also be computed using a normal approximation to the randomization distribution. It is a straightforward exercise to verify that

$$E(T) = \frac{n_1(n + 1)}{2} \quad \text{and} \quad \text{Var}(T) = \frac{n_1n_2s^2}{n}$$

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

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

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.)

For the normal approximation, we calculate

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

When the `porder` option is specified, the probability

$$p = \frac{U}{n_1n_2}$$

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).
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


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


**Also see**

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

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