NAG Toolbox

nag_nonpar_test_wilcoxon (g08ag)

1 Purpose

nag_nonpar_test_wilcoxon (g08ag) performs the Wilcoxon signed rank test on a single sample of size n.

2 Syntax

```
[w, wnor, p, n1, ifail] = nag_nonpar_test_wilcoxon(x, xme, tail, zer, 'n', n)
[w, wnor, p, n1, ifail] = g08ag(x, xme, tail, zer, 'n', n)
```

3 Description

The Wilcoxon one-sample signed rank test may be used to test whether a particular sample came from a population with a specified median. It is assumed that the population distribution is symmetric. The data consists of a single sample of n observations denoted by $x_1, x_2, \ldots, x_n$. This sample may arise from the difference between pairs of observations from two matched samples of equal size taken from two populations, in which case the test may be used to test whether the median of the first population is the same as that of the second population.

The hypothesis under test, H_0, often called the null hypothesis, is that the median is equal to some given value (X_{med}), and this is to be tested against an alternative hypothesis H_1 which is

H_1: population median $\neq X_{\mathrm{med}}$; or

H_1: population median $> X_{\mathrm{med}}$; or

H_1: population median $< X_{\mathrm{med}}$,

using a two tailed, upper tailed or lower tailed probability respectively. You select the alternative hypothesis by choosing the appropriate tail probability to be computed (see the description of argument **tail** in Section 5).

The Wilcoxon test differs from the Sign test (see nag_nonpar_test_sign (g08aa)) in that the magnitude of the scores is taken into account, rather than simply the direction of such scores.

The test procedure is as follows

(a) For each x_i, for $i = 1, 2, \ldots, n$, the signed difference $d_i = x_i - X_{\mathrm{med}}$ is found, where X_{med} is a given test value for the median of the sample.

(b) The absolute differences $|d_i|$ are ranked with rank r_i and any tied values of $|d_i|$ are assigned the average of the tied ranks. You may choose whether or not to ignore any cases where $d_i = 0$ by removing them before or after ranking (see the description of the argument **zer** in Section 5).

(c) The number of nonzero d_i is found.

(d) To each rank is affixed the sign of the d_i to which it corresponds. Let $s_i = \mathrm{sign}(d_i) r_i$.

(e) The sum of the positive-signed ranks, $W = \sum_{s_i > 0} s_i = \sum_{i=1}^{n} \max(s_i, 0.0)$, is calculated.

nag_nonpar_test_wilcoxon (g08ag) returns

(a) the test statistic W;

(b) the number n_1 of nonzero d_i;

(c) the approximate Normal test statistic z, where

$$z = \frac{\left(W - \frac{n_1(n_1+1)}{4}\right) - \operatorname{sign}\left(W - \frac{n_1(n_1+1)}{4}\right) \times \frac{1}{2}}{\sqrt{\frac{1}{4}\sum_{i=1}^{n} s_i^2}};$$

(d) the tail probability, p, corresponding to W, depending on the choice of the alternative hypothesis, H_1.

If $n_1 \leq 80$, p is computed exactly; otherwise, an approximation to p is returned based on an approximate Normal statistic corrected for continuity according to the tail specified.

The value of p can be used to perform a significance test on the median against the alternative hypothesis. Let α be the size of the significance test (that is, α is the probability of rejecting H_0 when H_0 is true). If $p < \alpha$ then the null hypothesis is rejected. Typically α might be 0.05 or 0.01.

4 References

Conover W J (1980) *Practical Nonparametric Statistics* Wiley

Neumann N (1988) Some procedures for calculating the distributions of elementary nonparametric teststatistics *Statistical Software Newsletter* **14(3)** 120–126

Siegel S (1956) *Non-parametric Statistics for the Behavioral Sciences* McGraw–Hill

5 Parameters

5.1 Compulsory Input Parameters

1: **x**(**n**) – REAL (KIND=nag_wp) array

The sample observations, $x_1, x_2, \ldots, x_n$.

2: **xme** – REAL (KIND=nag_wp)

The median test value, X_{med}.

3: **tail** – CHARACTER(1)

Indicates the choice of tail probability, and hence the alternative hypothesis.

tail = 'T'
A two tailed probability is calculated and the alternative hypothesis is H_1: population median $\neq X_{\text{med}}$.

tail = 'U'
An upper tailed probability is calculated and the alternative hypothesis is H_1: population median $> X_{\text{med}}$.

tail = 'L'
A lower tailed probability is calculated and the alternative hypothesis is H_1: population median $< X_{\text{med}}$.

Constraint: **tail** = 'T', 'U' or 'L'.

4: **zer** – CHARACTER(1)

Indicates whether or not to include the cases where $d_i = 0.0$ in the ranking of the d_i's.

zer = 'Y'
All $d_i = 0.0$ are included when ranking.

zer = 'N'
All $d_i = 0.0$, are ignored, that is all cases where $d_i = 0.0$ are removed before ranking.

Constraint: **zer** = 'Y' or 'N'.

5.2 Optional Input Parameters

1: **n** – INTEGER

Default: the dimension of the array **x**.

n, the size of the sample.

Constraint: **n** ≥ 1.

5.3 Output Parameters

1: **w** – REAL (KIND=nag_wp)

The Wilcoxon rank sum statistic, W, being the sum of the positive ranks.

2: **wnor** – REAL (KIND=nag_wp)

The approximate Normal test statistic, z, as described in Section 3.

3: **p** – REAL (KIND=nag_wp)

The tail probability, p, as specified by the argument **tail**.

4: **n1** – INTEGER

The number of nonzero d_i's, n_1.

5: **ifail** – INTEGER

ifail $= 0$ unless the function detects an error (see Section 5).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail $= 1$

On entry, **tail** $\neq$ 'T', 'U' or 'L'.
or **zer** $\neq$ 'Y' or 'N'.

ifail $= 2$

On entry, **n** < 1.

ifail $= 3$ (*warning*)

The whole sample is identical to the given median test value.

ifail $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

ifail $= -399$

Your licence key may have expired or may not have been installed correctly.

ifail $= -999$

Dynamic memory allocation failed.

7 Accuracy

The approximation used to calculate p when $n_1 > 80$ will return a value with a relative error of less than 10% for most cases. The error may increase for cases where there are a large number of ties in the sample.

8 Further Comments

The time taken by nag_nonpar_test_wilcoxon (g08ag) increases with n_1, until $n_1 > 80$, from which point on the approximation is used. The time decreases significantly at this point and increases again modestly with n_1 for $n_1 > 80$.

9 Example

This example performs the Wilcoxon signed rank test on two matched samples of size 8, taken from two populations. The distribution of the differences between pairs of observations from the two populations is assumed to be symmetric. The test is used to test whether the medians of the two distributions of the populations are equal or not. The test statistic, the approximate Normal statistic and the two tailed probability are computed and printed.

9.1 Program Text

```
function g08ag_example

fprintf('g08ag example results\n\n');

% Data
x = [82 69 73 43 58 56 76 65];
y = [63 42 74 37 51 43 80 62];

fprintf('Wilcoxon one sample signed ranks test\n\n');
fprintf('Data values\n    ');
fprintf('%5.1f', x);
fprintf('\n    ');
fprintf('%5.1f', y);
fprintf('\n\n');

% Control parameters
xme = 0;
tail = 'Two-tail';
zer = 'Nozeros';

% z = differences
z = x - y;

[w, wnor, p, n1, ifail] = g08ag( ...
                                 z, xme, tail, zer);

fprintf('Test statistic             = %8.4f\n', w);
fprintf('Normalized test statistic = %8.4f\n', wnor);
fprintf('Degrees of freedom         = %3d\n', n1);
fprintf('Two tail probability       = %8.4f\n', p);
```

9.2 Program Results

```
g08ag example results

Wilcoxon one sample signed ranks test

Data values
    82.0 69.0 73.0 43.0 58.0 56.0 76.0 65.0
    63.0 42.0 74.0 37.0 51.0 43.0 80.0 62.0
```

```
Test statistic            =  32.0000
Normalized test statistic =   1.8904
Degrees of freedom        =   8
Two tail probability      =   0.0547
```