NAG Toolbox for MATLAB

g08ag

1 Purpose

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

2 Syntax

```
[w, wnor, p, n1, ifail] = g08ag(x, xme, tail, zeros, '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 parameter **tail** in Section 5).

The Wilcoxon test differs from the Sign test (see 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 parameter **zeros** 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.

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) – double array**

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

2: **xme – double scalar**

The median test value, X_{med}.

3: **tail – string**

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 $\neq X_{\mathrm{med}}$.

tail = 'U'

A upper-tailed probability is calculated and the alternative hypothesis is H_1: population median $> X_{\mathrm{med}}$.

tail = 'L'

A lower-tailed probability is calculated and the alternative hypothesis is H_1: population median $< X_{\mathrm{med}}$.

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

4: **zeros – string**

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

zeros = 'Y'

All $d_i = 0.0$ are included when ranking.

zeros = 'N'

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

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

5.2 Optional Input Parameters

1: **n – int32 scalar**

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

n, the size of the sample.

Constraint: **n** ≥ 1.

5.3 Input Parameters Omitted from the MATLAB Interface

wrk

5.4 Output Parameters

1: **w – double scalar**

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

2: **wnor – double scalar**

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

3: **p – double scalar**

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

4: **n1 – int32 scalar**

The number of nonzero d_i's, n_1.

5: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail = 1

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

ifail = 2

On entry, **n** < 1.

ifail = 3

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

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 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

```
x = [19;
     27;
     -1;
     6;
     7;
     13;
     -4;
     3];
xme = 0;
tail = 'Two-tail';
zeros = 'Nozeros';
[w, wnor, p, n1, ifail] = g08ag(x, xme, tail, zeros)
```

```
w =
    32
wnor =
    1.8904
p =
    0.0547
n1 =
           8
ifail =
           0
```