

NAG Toolbox for MATLAB

g08ak

1 Purpose

g08ak calculates the exact tail probability for the Mann–Whitney rank sum test statistic for the case where there are ties in the two samples pooled together.

2 Syntax

```
[p, ifail] = g08ak(n1, n2, tail, ranks, u)
```

3 Description

g08ak computes the exact tail probability for the Mann–Whitney U test statistic (calculated by g08ah and returned through the parameter **u**) using a method based on an algorithm developed by Neumann 1988, for the case where there are ties in the pooled sample.

The Mann–Whitney U test investigates the difference between two populations defined by the distribution functions $F(x)$ and $G(y)$ respectively. The data consist of two independent samples of size n_1 and n_2 , denoted by $x_1, x_2, \dots, x_{n_1}$ and $y_1, y_2, \dots, y_{n_2}$, taken from the two populations.

The hypothesis under test, H_0 , often called the null hypothesis, is that the two distributions are the same, that is $F(x) = G(x)$, and this is to be tested against an alternative hypothesis H_1 which is

$H_1 : F(x) \neq G(y)$; or

$H_1 : F(x) < G(y)$, i.e., the x 's tend to be greater than the y 's; or

$H_1 : F(x) > G(y)$, i.e., the x 's tend to be less than the y 's,

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).

Note that when using this test to test for differences in the distributions one is primarily detecting differences in the location of the two distributions. That is to say, if we reject the null hypothesis H_0 in favour of the alternative hypothesis $H_1 : F(x) > G(y)$ we have evidence to suggest that the location, of the distribution defined by $F(x)$, is less than the location of the distribution defined by $G(y)$.

g08ak returns the exact tail probability, p , corresponding to U , depending on the choice of alternative hypothesis, H_1 .

The value of p can be used to perform a significance test on the null hypothesis H_0 against the alternative hypothesis H_1 . 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 test statistics *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: **n1 – int32 scalar**

The number of non-tied pairs, n_1 .

Constraint: **n1** ≥ 1 .

2: **n2 – int32 scalar**

the size of the second sample, n_2 .

Constraint: **n2** ≥ 1 .

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 $H_1 : F(x) \neq G(y)$.

tail = 'U'

An upper-tailed probability is calculated and the alternative hypothesis $H_1 : F(x) < G(y)$, i.e., the x 's tend to be greater than the y 's.

tail = 'L'

A lower-tailed probability is calculated and the alternative hypothesis $H_1 : F(x) > G(y)$, i.e., the x 's tend to be less than the y 's.

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

4: **ranks(n1 + n2) – double array**

The ranks of the pooled sample. These ranks are output in the array **ranks** by g08ah and should not be altered in any way if you are using the same n_1 , n_2 and **u** as used in g08ah.

5: **u – double scalar**

U , the value of the Mann–Whitney rank sum test statistic. This is the statistic returned through the parameter **u** by g08ah.

5.2 Optional Input Parameters

None.

5.3 Input Parameters Omitted from the MATLAB Interface

wrk, lwrk, iwrk

5.4 Output Parameters

1: **p – double scalar**

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

2: **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, **n1** < 1,
or **n2** < 1.

ifail = 2

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

ifail = 3

On entry, **u** < 0.0.

ifail = 4

On entry, **lwrk** is too small.

7 Accuracy

The exact tail probability, p , is computed to an accuracy of at least 4 significant figures.

8 Further Comments

The time taken by g08ak increases with n_1 and n_2 and the product $n_1 n_2$. Note that the amount of workspace required becomes very large for even moderate sizes of n_1 and n_2 .

9 Example

```
n1 = int32(16);
n2 = int32(23);
tail = 'Lower-tail';
ranks = [29.5;
        1.5;
        24.5;
        5;
        24.5;
        5;
        16;
        5;
        16;
        5;
        38;
        5;
        16;
        9.5;
        12;
        9.5;
        39;
        1.5;
        16;
        9.5;
        36;
        9.5;
        36;
        16;
        36;
        16;
        33;
        16;
```

```
    33;  
    20.5;  
    33;  
    20.5;  
    29.5;  
    24.5;  
    29.5;  
    24.5;  
    29.5;  
    24.5;  
    24.5];  
u = 86;  
[p, ifail] = g08ak(n1, n2, tail, ranks, u)  
  
p =  
    0.0020  
ifail =  
        0
```
