NAG Library Routine Document

G08AJF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

1 Purpose

G08AJF calculates the exact tail probability for the Mann–Whitney rank sum test statistic for the case where there are no ties in the two samples pooled together.

2 Specification

```
SUBROUTINE G08AJF (N1, N2, TAIL, U, P, WRK, LWRK, IFAIL)

INTEGER             N1, N2, LWRK, IFAIL
REAL (KIND=nag_wp) U, P, WRK(LWRK)
CHARACTER(1)        TAIL
```

3 Description

G08AJF computes the exact tail probability for the Mann–Whitney U test statistic (calculated by G08AHF and returned through the argument U) using a method based on an algorithm developed by Harding (1983), and presented by Neumann (1988), for the case where there are no 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 $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$, 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 argument 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 $G(y)$.

G08AJF 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

Harding E F (1983) An efficient minimal-storage procedure for calculating the Mann–Whitney U, generalised U and similar distributions *Appl. Statist.* **33** 1–6

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 Arguments

1: N1 – INTEGER *Input*

On entry: the number of non-tied pairs, n_1.

Constraint: N1 ≥ 1.

2: N2 – INTEGER *Input*

On entry: the size of the second sample, n_2.

Constraint: N2 ≥ 1.

3: TAIL – CHARACTER(1) *Input*

On entry: 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: U – REAL (KIND=nag_wp) *Input*

On entry: U, the value of the Mann–Whitney rank sum test statistic. This is the statistic returned through the argument U by G08AHF.

Constraint: U ≥ 0.0.

5: P – REAL (KIND=nag_wp) *Output*

On exit: the exact tail probability, p, as specified by the argument TAIL.

6: WRK(LWRK) – REAL (KIND=nag_wp) array *Workspace*
7: LWRK – INTEGER *Input*

On entry: the dimension of the array WRK as declared in the (sub)program from which G08AJF is called.

Constraint: LWRK $\geq (\text{N1} \times \text{N2})/2 + 1$.

8: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this argument you should refer to Section 3.4 in How to Use the NAG Library and its Documentation for details.

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, if you are not familiar with this argument, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

On exit: IFAIL $= 0$ unless the routine detects an error or a warning has been flagged (see Section 6).

6 Error Indicators and Warnings

If on entry IFAIL $= 0$ or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

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 $< (\text{N1} \times \text{N2})/2 + 1$.

IFAIL $= -99$

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

See Section 3.9 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -399$

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

See Section 3.8 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -999$

Dynamic memory allocation failed.

See Section 3.7 in How to Use the NAG Library and its Documentation for further information.

7 Accuracy

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

8 Parallelism and Performance

G08AJF is not threaded in any implementation.

9 Further Comments

The time taken by G08AJF increases with n_1 and n_2 and the product $n_1 n_2$.

10 Example

This example finds the Mann–Whitney test statistic, using G08AHF for two independent samples of size 16 and 23 respectively. This is used to test the null hypothesis that the distributions of the two populations from which the samples were taken are the same against the alternative hypothesis that the distributions are different. The test statistic, the approximate normal statistic and the approximate two tail probability are printed. G08AJF is then called to obtain the exact two tailed probability. The exact probability is also printed.

10.1 Program Text

```
    Program g08ajfe

!     G08AJF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g08ahf, g08ajf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter              :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)              :: p, pexact, u, unor
      Integer                         :: ifail, lwrk, lwrk1, lwrk2, n1, n2,   &
                                         nsum
      Logical                         :: ties
      Character (1)                   :: tail
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: ranks(:), wrk(:), x(:), y(:)
!     .. Intrinsic Procedures ..
      Intrinsic                       :: int, max
!     .. Executable Statements ..
      Write (nout,*) 'G08AJF Example Program Results'
      Write (nout,*)

!     Skip heading in data file
      Read (nin,*)

!     Read in problem size
      Read (nin,*) n1, n2, tail

!     Calculate sizes of various workspaces
      nsum = n1 + n2

!     Workspace for G08AHF
      lwrk1 = nsum

!     Workspace for G08AJF
      lwrk2 = int(n1*n2/2) + 1

      lwrk = max(lwrk1,lwrk2)
      Allocate (x(n1),y(n2),ranks(nsum),wrk(lwrk))

!     Read in data
      Read (nin,*) x(1:n1)
      Read (nin,*) y(1:n2)

!     Display title
      Write (nout,*) 'Mann-Whitney U test'
      Write (nout,*)

!     Display input data
      Write (nout,99999) 'Sample size of group 1 = ', n1
      Write (nout,99999) 'Sample size of group 2 = ', n2
      Write (nout,*)
      Write (nout,*) 'Data values'
      Write (nout,*)
      Write (nout,99998) '    Group 1  ', x(1:n1)
      Write (nout,*)
      Write (nout,99998) '    Group 2  ', y(1:n2)

!     Perform test
      ifail = 0
      Call g08ahf(n1,x,n2,y,tail,u,unor,p,ties,ranks,wrk,ifail)
```

```
!     Calculate exact probabilities
      If (.Not. ties) Then
        ifail = 0
        Call g08ajf(n1,n2,tail,u,pexact,wrk,lwrk,ifail)
      End If

!     Display results
      Write (nout,*)
      Write (nout,99997) 'Test statistic     = ', u
      Write (nout,99997) 'Normal statistic   = ', unor
      Write (nout,99997) 'Tail probability   = ', p
      Write (nout,*)
      If (.Not. ties) Then
        Write (nout,99997) 'Exact tail probability = ', pexact
      Else
        Write (nout,*)                                                 &
          'There are ties in the pooled sample so G08AJF was not called.'
      End If

99999 Format (1X,A,I5)
99998 Format (1X,A,8F5.1,2(/,14X,8F5.1))
99997 Format (1X,A,F10.4)
    End Program g08ajfe
```

10.2 Program Data

```
G08AJF Example Program Data
 16  23  'L'                              :: N1,N2,TAIL
 13.0  5.8 11.7  6.5 12.3  6.7  9.2  6.9
 10.0  7.3 16.0  7.0 10.5  8.5  9.0  7.5 :: End of X
 17.0  6.2 10.1  8.0 15.3  8.2 15.0  9.6
 14.9 10.4 14.2  9.8 13.8 11.0 14.0 11.1
 12.9 11.6 12.8 12.0 13.1 12.4 11.9      :: End of Y
```

10.3 Program Results

```
 G08AJF Example Program Results

 Mann-Whitney U test

 Sample size of group 1 =    16
 Sample size of group 2 =    23

 Data values

     Group 1   13.0  5.8 11.7  6.5 12.3  6.7  9.2  6.9
               10.0  7.3 16.0  7.0 10.5  8.5  9.0  7.5


     Group 2   17.0  6.2 10.1  8.0 15.3  8.2 15.0  9.6
               14.9 10.4 14.2  9.8 13.8 11.0 14.0 11.1
               12.9 11.6 12.8 12.0 13.1 12.4 11.9

 Test statistic     =    86.0000
 Normal statistic   =    -2.7838
 Tail probability   =     0.0027

 Exact tail probability =     0.0022
```