

NAG Library Routine Document

G08AGF

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

G08AGF performs the Wilcoxon signed rank test on a single sample of size n .

2 Specification

```
SUBROUTINE G08AGF(N, X, XME, TAIL, ZEROS, W, WNOR, P, N1, WRK, IFAIL)
INTEGER          N, N1, IFAIL
double precision X(N), XME, W, WNOR, P, WRK(3*N)
CHARACTER*1      TAIL, ZEROS
```

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 $x_1, x_2, \dots, x_n$. 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 $\neq X_{\text{med}}$; or

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

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

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 G08AAF) 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

- For each x_i , for $i = 1, 2, \dots, n$, the signed difference $d_i = x_i - X_{\text{med}}$ is found, where X_{med} is a given test value for the median of the sample.
- 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).
- The number of nonzero d_i is found.
- To each rank is affixed the sign of the d_i to which it corresponds. Let $s_i = \text{sign}(d_i)r_i$.
- 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.

G08AGF returns

- (a) the test statistic W ;
- (b) the number n_1 of nonzero d_i ;
- (c) the approximate Normal test statistic z , where

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

- (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

- 1: N – INTEGER Input
On entry: n , the size of the sample.
Constraint: $N \geq 1$.
- 2: X(N) – **double precision** array Input
On entry: the sample observations, $x_1, x_2, \dots, x_n$.
- 3: XME – **double precision** Input
On entry: the median test value, X_{med} .
- 4: 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 : population median $\neq X_{\text{med}}$.
 TAIL = 'U'
 A 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'.

- 5: ZEROS – CHARACTER*1 *Input*
On entry: 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'.
- 6: W – *double precision* *Output*
On exit: the Wilcoxon rank sum statistic, W , being the sum of the positive ranks.
- 7: WNOR – *double precision* *Output*
On exit: the approximate Normal test statistic, z , as described in Section 3.
- 8: P – *double precision* *Output*
On exit: the tail probability, p , as specified by the parameter TAIL.
- 9: N1 – INTEGER *Output*
On exit: the number of nonzero d_i 's, n_1 .
- 10: WRK($3 \times N$) – *double precision* array *Workspace*
- 11: IFAIL – INTEGER *Input/Output*
On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction for details.
On exit: IFAIL = 0 unless the routine detects an error (see Section 6).
 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 parameter, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

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, 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 G08AGF 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

```
*      G08AGF Example Program Text
*      Mark 14 Release. NAG Copyright 1989.
*      .. Parameters ..
      INTEGER          MAXN
      PARAMETER        (MAXN=10)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      DOUBLE PRECISION XME
      PARAMETER        (XME=0.0D0)
*      .. Local Scalars ..
      DOUBLE PRECISION P, W, WNOR
      INTEGER          I, IFAIL, N, N1
*      .. Local Arrays ..
      DOUBLE PRECISION WRK(3*MAXN), X(MAXN), Y(MAXN), Z(MAXN)
*      .. External Subroutines ..
      EXTERNAL         G08AGF
*      .. Executable Statements ..
      WRITE (NOUT,*) 'G08AGF Example Program Results'
*      Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N
      IF (N.LE.MAXN) THEN
        READ (NIN,*) (X(I),I=1,N), (Y(I),I=1,N)
        WRITE (NOUT,*)
        WRITE (NOUT,*) 'Wilcoxon one sample signed ranks test'
        WRITE (NOUT,*)
        WRITE (NOUT,*) 'Data values'
        WRITE (NOUT,99999) (X(I),I=1,N), (Y(I),I=1,N)
        DO 20 I = 1, N
          Z(I) = X(I) - Y(I)
20      CONTINUE
        IFAIL = 0
*
        CALL G08AGF(N,Z,XME,'Two-tail','Nozeros',W,WNOR,P,N1,WRK,IFAIL)
*
        WRITE (NOUT,*)
        WRITE (NOUT,99998) 'Test statistic           = ', W
        WRITE (NOUT,99998) 'Normalized test statistic = ', WNOR
        WRITE (NOUT,99997) 'Degrees of freedom      = ', N1
        WRITE (NOUT,99998) 'Two tail probability      = ', P
      ELSE
        WRITE (NOUT,99996) 'N is too large : N = ', N
      END IF
*
```

```
99999 FORMAT (4X,8F5.1)
99998 FORMAT (1X,A,F8.4)
99997 FORMAT (1X,A,I8)
99996 FORMAT (1X,A,I16)
      END
```

9.2 Program Data

```
G08AGF Example Program Data
      8
      82 69 73 43 58 56 76 65
      63 42 74 37 51 43 80 62
```

9.3 Program Results

G08AGF Example Program 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
