NAG Fortran Library Routine Document

G08ACF

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

G08ACF performs the Median test on two independent samples of possibly unequal size.

2 Specification

```
SUBROUTINE G08ACF(X, N, N1, W, I1, I2, P, IFAIL)
INTEGER           N, N1, I1, I2, IFAIL
real              X(N), W(N), P
```

3 Description

The Median test investigates the difference between the medians of two independent samples of sizes n_1 and n_2, denoted by:

$$x_1, x_2, \ldots, x_{n_1}$$

and

$$x_{n_1+1}, x_{n_1+2}, \ldots, x_n,$$

where $n = n_1 + n_2$.

The hypothesis under test, H_0, often called the null hypothesis, is that the medians are the same, and this is to be tested against the alternative hypothesis H_1 that they are different.

The test proceeds by forming a 2×2 frequency table, giving the number of scores in each sample above and below the median of the pooled sample:

	Sample 1	Sample 2	Total
Scores $\leq$ pooled median	i_1	i_2	$i_1 + i_2$
Scores $\geq$ pooled median	$n_1 - i_1$	$n_2 - i_2$	$n - (i_1 + i_2)$
Total	n_1	n_2	n

Under the null hypothesis, H_0, we would expect about half of each group's scores to be above the pooled median and about half below, that is, we would expect i_1, to be about $n_1/2$ and i_2 to be about $n_2/2$.

G08ACF returns:

(a) the frequencies i_1 and i_2;

(b) the probability, p, of observing a table at least as 'extreme' as that actually observed, given that H_0 is true. If $n < 40$, p is computed directly ('Fisher's exact test'); otherwise a χ^2_1 approximation is used (see G01AFF).

H_0 is rejected by a test of chosen size α if $p < \alpha$.

4 References

Siegel S (1956) *Nonparametric Statistics for the Behavioral Sciences* McGraw-Hill

5 Parameters

1: X(N) – ***real*** array *Input*

On entry: the first n_1 elements of X must be set to the data values in the first sample, and the next n_2 $(= \mathrm{N} - n_1)$ elements to the data values in the second sample.

2: N – INTEGER *Input*

On entry: the total of the two sample sizes, $n(= n_1 + n_2)$.

Constraint: $\mathrm{N} \geq 2$.

3: N1 – INTEGER *Input*

On entry: the size of the first sample n_1.

Constraint: $1 \leq \mathrm{N1} < \mathrm{N}$.

4: W(N) – ***real*** array *Workspace*

5: I1 – INTEGER *Output*

On exit: the number of scores in the first sample which lie below the pooled median, i_1.

6: I2 – INTEGER *Output*

On exit: the number of scores in the second sample which lie below the pooled median, i_2.

7: P – ***real*** *Output*

On exit: the tail probability p corresponding to the observed dichotomy of the two samples.

8: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 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, for users 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, $\mathrm{N} < 2$.

IFAIL $= 2$

On entry, $\mathrm{N1} < 1$,
or $\mathrm{N1} \geq \mathrm{N}$.

7 Accuracy

The probability returned should be accurate enough for practical use.

8 Further Comments

The time taken by the routine is small, and increases with n.

9 Example

This example is taken from page 112 of Siegel (1956). The data relate to scores of 'oral socialisation anxiety' in 39 societies, which can be separated into groups of size 16 and 23 on the basis of their attitudes to illness.

9.1 Program Text

Note: the listing of the example program presented below uses ***bold italicised*** terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     G08ACF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          N
      PARAMETER        (N=39)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*     .. Local Scalars ..
      real             P
      INTEGER          I, I1, I2, IFAIL, N1
*     .. Local Arrays ..
      real             W1(N), X(N)
*     .. External Subroutines ..
      EXTERNAL         G08ACF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G08ACF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) (X(I),I=1,N)
      N1 = 16
      WRITE (NOUT,*)
      WRITE (NOUT,*) 'Median test'
      WRITE (NOUT,*)
      WRITE (NOUT,*) 'Data values'
      WRITE (NOUT,*)
      WRITE (NOUT,99999) '    Group 1  ', (X(I),I=1,N1)
      WRITE (NOUT,*)
      WRITE (NOUT,99999) '    Group 2  ', (X(I),I=N1+1,N)
      IFAIL = 0
*
      CALL G08ACF(X,N,N1,W1,I1,I2,P,IFAIL)
*
      WRITE (NOUT,*)
      WRITE (NOUT,99998) I1, ' scores below median in group 1'
      WRITE (NOUT,99998) I2, ' scores below median in group 2'
      WRITE (NOUT,*)
      WRITE (NOUT,99997) '     Significance  ', P
      STOP
*
99999 FORMAT (1X,A,8F4.0,/(14X,8F4.0))
99998 FORMAT (1X,I6,A)
99997 FORMAT (1X,A,F8.5)
      END
```

9.2 Program Data

```
G08ACF Example Program Data
 13  6 12  7 12  7 10  7 10  7 10  7 10  8  9  8
 17  6 16  8 15  8 15 10 15 10 14 10 14 11 14 11
 13 12 13 12 13 12 12
```

9.3 Program Results

```
 G08ACF Example Program Results

 Median test

 Data values

     Group 1   13.  6. 12.  7. 12.  7. 10.  7.
               10.  7. 10.  7. 10.  8.  9.  8.

     Group 2   17.  6. 16.  8. 15.  8. 15. 10.
               15. 10. 14. 10. 14. 11. 14. 11.
               13. 12. 13. 12. 13. 12. 12.

     13 scores below median in group 1
      6 scores below median in group 2

      Significance   0.00088
```