NAG Library Routine Document

G01DHF

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

G01DHF computes the ranks, Normal scores, an approximation to the Normal scores or the exponential scores as requested by you.

2 Specification

```
SUBROUTINE G01DHF (SCORES, TIES, N, X, R, IWRK, IFAIL)

INTEGER             N, IWRK(N), IFAIL
REAL (KIND=nag_wp) X(N), R(N)
CHARACTER(1)        SCORES, TIES
```

3 Description

G01DHF computes one of the following scores for a sample of observations, $x_1, x_2, \ldots, x_n$.

1. **Rank Scores**

   The ranks are assigned to the data in ascending order, that is the ith observation has score $s_i = k$ if it is the kth smallest observation in the sample.

2. **Normal Scores**

   The Normal scores are the expected values of the Normal order statistics from a sample of size n. If x_i is the kth smallest observation in the sample, then the score for that observation, s_i, is $E(Z_k)$ where Z_k is the kth order statistic in a sample of size n from a standard Normal distribution and E is the expectation operator.

3. **Blom, Tukey and van der Waerden Scores**

   These scores are approximations to the Normal scores. The scores are obtained by evaluating the inverse cumulative Normal distribution function, $\Phi^{-1}(\cdot)$, at the values of the ranks scaled into the interval $(0,1)$ using different scaling transformations.

   The Blom scores use the scaling transformation $\frac{r_i - \frac{3}{8}}{n + \frac{1}{4}}$ for the rank r_i, for $i = 1, 2, \ldots, n$. Thus the Blom score corresponding to the observation x_i is

$$s_i = \Phi^{-1}\left(\frac{r_i - \frac{3}{8}}{n + \frac{1}{4}}\right).$$

   The Tukey scores use the scaling transformation $\frac{r_i - \frac{1}{3}}{n + \frac{1}{3}}$; the Tukey score corresponding to the observation x_i is

$$s_i = \Phi^{-1}\left(\frac{r_i - \frac{1}{3}}{n + \frac{1}{3}}\right).$$

   The van der Waerden scores use the scaling transformation $\frac{r_i}{n+1}$; the van der Waerden score corresponding to the observation x_i is

$$s_i = \Phi^{-1}\left(\frac{r_i}{n + 1}\right).$$

The van der Waerden scores may be used to carry out the van der Waerden test for testing for differences between several population distributions, see Conover (1980).

4. **Savage Scores**

The Savage scores are the expected values of the exponential order statistics from a sample of size n. They may be used in a test discussed by Savage (1956) and Lehmann (1975). If x_i is the kth smallest observation in the sample, then the score for that observation is

$$s_i = E(Y_k) = \tfrac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-k+1},$$

where Y_k is the kth order statistic in a sample of size n from a standard exponential distribution and E is the expectation operator.

Ties may be handled in one of five ways. Let $x_{t(i)}$, for $i = 1, 2, \ldots, m$, denote m tied observations, that is $x_{t(1)} = x_{t(2)} = \cdots = x_{t(m)}$ with $t(1) < t(2) < \cdots < t(m)$. If the rank of $x_{t(1)}$ is k, then if ties are ignored the rank of $x_{t(j)}$ will be $k + j - 1$. Let the scores ignoring ties be $s^*_{t(1)}, s^*_{t(2)}, \ldots, s^*_{t(m)}$. Then the scores, $s_{t(i)}$, for $i = 1, 2, \ldots, m$, may be calculated as follows:

–if averages are used, then $s_{t(i)} = \sum_{j=1}^{m} s^*_{t(j)}/m$;

–if the lowest score is used, then $s_{t(i)} = s^*_{t(1)}$;

–if the highest score is used, then $s_{t(i)} = s^*_{t(m)}$;

–if ties are to be broken randomly, then $s_{t(i)} = s^*_{t(I)}$ where $I \in \{\text{random permutation of } 1, 2, \ldots, m\}$;

–if ties are to be ignored, then $s_{t(i)} = s^*_{t(i)}$.

4 References

Blom G (1958) *Statistical Estimates and Transformed Beta-variables* Wiley

Conover W J (1980) *Practical Nonparametric Statistics* Wiley

Lehmann E L (1975) *Nonparametrics: Statistical Methods Based on Ranks* Holden–Day

Savage I R (1956) Contributions to the theory of rank order statistics – the two-sample case *Ann. Math. Statist.* **27** 590–615

Tukey J W (1962) The future of data analysis *Ann. Math. Statist.* **33** 1–67

5 Arguments

1: SCORES – CHARACTER(1) *Input*

On entry: indicates which of the following scores are required.

SCORES = 'R'
The ranks.

SCORES = 'N'
The Normal scores, that is the expected value of the Normal order statistics.

SCORES = 'B'
The Blom version of the Normal scores.

SCORES = 'T'
The Tukey version of the Normal scores.

SCORES = 'V'
The van der Waerden version of the Normal scores.

SCORES = 'S'
The Savage scores, that is the expected value of the exponential order statistics.

Constraint: SCORES = 'R', 'N', 'B', 'T', 'V' or 'S'.

2: TIES – CHARACTER(1) *Input*

On entry: indicates which of the following methods is to be used to assign scores to tied observations.

TIES = 'A'
The average of the scores for tied observations is used.

TIES = 'L'
The lowest score in the group of ties is used.

TIES = 'H'
The highest score in the group of ties is used.

TIES = 'N'
The nonrepeatable random number generator is used to randomly untie any group of tied observations.

TIES = 'R'
The repeatable random number generator is used to randomly untie any group of tied observations.

TIES = 'I'
Any ties are ignored, that is the scores are assigned to tied observations in the order that they appear in the data.

Constraint: TIES = 'A', 'L', 'H', 'N', 'R' or 'I'.

3: N – INTEGER *Input*

On entry: n, the number of observations.

Constraint: N ≥ 1.

4: X(N) – REAL (KIND=nag_wp) array *Input*

On entry: the sample of observations, x_i, for $i = 1, 2, \ldots, n$.

5: R(N) – REAL (KIND=nag_wp) array *Output*

On exit: contains the scores, s_i, for $i = 1, 2, \ldots, n$, as specified by SCORES.

6: IWRK(N) – INTEGER array *Workspace*

7: 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, SCORES $\neq$ 'R', 'N', 'B', 'T', 'V' or 'S',
or TIES $\neq$ 'A', 'L', 'H', 'N', 'R' or 'I',
or N < 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

For SCORES = 'R', the results should be accurate to ***machine precision***.

For SCORES = 'S', the results should be accurate to a small multiple of ***machine precision***.

For SCORES = 'N', the results should have a relative accuracy of at least $\max\left(100 \times \epsilon, 10^{-8}\right)$ where ϵ is the ***machine precision***.

For SCORES = 'B', 'T' or 'V', the results should have a relative accuracy of at least $\max\left(10 \times \epsilon, 10^{-12}\right)$.

8 Parallelism and Performance

G01DHF is threaded by NAG for parallel execution in multithreaded implementations of the NAG Library.

Please consult the X06 Chapter Introduction for information on how to control and interrogate the OpenMP environment used within this routine. Please also consult the Users' Note for your implementation for any additional implementation-specific information.

9 Further Comments

If more accurate Normal scores are required G01DAF should be used with appropriate settings for the input argument ETOL.

10 Example

This example computes and prints the Savage scores for a sample of five observations. The average of the scores of any tied observations is used.

10.1 Program Text

```
    Program g01dhfe

!     G01DHF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g01dhf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Integer                          :: ifail, n
      Character (20)                   :: scores, ties
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: r(:), x(:)
      Integer, Allocatable             :: iwrk(:)
!     .. Executable Statements ..
      Write (nout,*) 'G01DHF Example Program Results'
      Write (nout,*)

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

!     Read in the problem size
      Read (nin,*) scores, ties, n

      Allocate (r(n),x(n),iwrk(n))

!     Read in data
      Read (nin,*) x(1:n)

!     Compute ranks
      ifail = 0
      Call g01dhf(scores,ties,n,x,r,iwrk,ifail)

!     Display results
      Write (nout,*) 'Scores: ', scores
      Write (nout,*) 'Ties  : ', ties
      Write (nout,*)
      Write (nout,99999) r(1:n)

99999 Format (1X,F10.4)
    End Program g01dhfe
```

10.2 Program Data

```
G01DHF Example Program Data
Savage Average 5  :: SCORES,TIES,N
2 0 2 2 0         :: End of X
```

10.3 Program Results

```
 G01DHF Example Program Results

 Scores: Savage
 Ties  : Average

     1.4500
     0.3250
     1.4500
     1.4500
     0.3250
```