NAG Library Routine Document

G07DDF

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

G07DDF calculates the trimmed and Winsorized means of a sample and estimates of the variances of the two means.

2 Specification

```
SUBROUTINE G07DDF (N, X, ALPHA, TMEAN, WMEAN, TVAR, WVAR, K, SX, IFAIL)
INTEGER            N, K, IFAIL
REAL (KIND=nag_wp) X(N), ALPHA, TMEAN, WMEAN, TVAR, WVAR, SX(N)
```

3 Description

G07DDF calculates the α-trimmed mean and α-Winsorized mean for a given α, as described below.

Let x_i, for $i = 1, 2, \ldots, n$ represent the n sample observations sorted into ascending order. Let $k = [\alpha n]$ where $[y]$ represents the integer nearest to y; if $2k = n$ then k is reduced by 1.

Then the trimmed mean is defined as:

$$\bar{x}_t = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} x_i,$$

and the Winsorized mean is defined as:

$$\bar{x}_w = \frac{1}{n}\left(\sum_{i=k+1}^{n-k} x_i + kx_{k+1} + kx_{n-k}\right).$$

G07DDF then calculates the Winsorized variance about the trimmed and Winsorized means respectively and divides by n to obtain estimates of the variances of the above two means.

Thus we have;

$$\text{Estimate of } \operatorname{var}(\bar{x}_t) = \frac{1}{n^2}\left(\sum_{i=k+1}^{n-k}(x_i - \bar{x}_t)^2 + k(x_{k+1} - \bar{x}_t)^2 + k(x_{n-k} - \bar{x}_t)^2\right)$$

and

$$\text{Estimate of } \operatorname{var}(\bar{x}_w) = \frac{1}{n^2}\left(\sum_{i=k+1}^{n-k}(x_i - \bar{x}_w)^2 + k(x_{k+1} - \bar{x}_w)^2 + k(x_{n-k} - \bar{x}_w)^2\right).$$

4 References

Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W A (1986) *Robust Statistics. The Approach Based on Influence Functions* Wiley

Huber P J (1981) *Robust Statistics* Wiley

5 Arguments

1: N – INTEGER *Input*

On entry: n, the number of observations.

Constraint: N ≥ 2.

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

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

3: ALPHA – REAL (KIND=nag_wp) *Input*

On entry: α, the proportion of observations to be trimmed at each end of the sorted sample.

Constraint: $0.0 \leq$ ALPHA < 0.5.

4: TMEAN – REAL (KIND=nag_wp) *Output*

On exit: the α-trimmed mean, $\bar{x}_t$.

5: WMEAN – REAL (KIND=nag_wp) *Output*

On exit: the α-Winsorized mean, $\bar{x}_w$.

6: TVAR – REAL (KIND=nag_wp) *Output*

On exit: contains an estimate of the variance of the trimmed mean.

7: WVAR – REAL (KIND=nag_wp) *Output*

On exit: contains an estimate of the variance of the Winsorized mean.

8: K – INTEGER *Output*

On exit: contains the number of observations trimmed at each end, k.

9: SX(N) – REAL (KIND=nag_wp) array *Output*

On exit: contains the sample observations sorted into ascending order.

10: 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, N ≤ 1.

IFAIL $= 2$

On entry, ALPHA < 0.0,
or ALPHA ≥ 0.5.

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 results should be accurate to within a small multiple of ***machine precision***.

8 Parallelism and Performance

G07DDF 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

The time taken is proportional to n.

10 Example

The following program finds the α-trimmed mean and α-Winsorized mean for a sample of 16 observations where $\alpha = 0.15$. The estimates of the variances of the above two means are also calculated.

10.1 Program Text

```
    Program g07ddfe

!     G07DDF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g07ddf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: alpha, propn, tmean, tvar, wmean,    &
                                          wvar
      Integer                          :: ifail, k, n
!     .. Local Arrays ..
```

```
      Real (Kind=nag_wp), Allocatable  :: sx(:), x(:)
!     .. Intrinsic Procedures ..
      Intrinsic                        :: real
!     .. Executable Statements ..
      Write (nout,*) 'G07DDF Example Program Results'
      Write (nout,*)

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

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

      Allocate (x(n),sx(n))

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

!     Trim data
      ifail = 0
      Call g07ddf(n,x,alpha,tmean,wmean,tvar,wvar,k,sx,ifail)

!     Calculate proportion of data cut
      propn = real(k,kind=nag_wp)/real(n,kind=nag_wp)
      propn = 100.0E0_nag_wp - 200.0E0_nag_wp*propn

!     Display results
      Write (nout,99999) 'Statistics from middle ', propn, '% of data'
      Write (nout,*)
      Write (nout,99998) '               Trimmed-mean = ', tmean
      Write (nout,99998) '   Variance of Trimmed-mean = ', tvar
      Write (nout,*)
      Write (nout,99998) '            Winsorized-mean = ', wmean
      Write (nout,99998) 'Variance of Winsorized-mean = ', wvar

99999 Format (1X,A,F6.2,A)
99998 Format (1X,A,F11.4)
    End Program g07ddfe
```

10.2 Program Data

```
G07DDF Example Program Data
 16  0.15             :: N,ALPHA
 26.0 12.0  9.0  2.0
  5.0  6.0  8.0 14.0
  7.0  3.0  1.0 11.0
 10.0  4.0 17.0 21.0  :: End of X
```

10.3 Program Results

```
 G07DDF Example Program Results

 Statistics from middle  75.00% of data

               Trimmed-mean =      8.8333
   Variance of Trimmed-mean =      1.5434

            Winsorized-mean =      9.1250
 Variance of Winsorized-mean =      1.5381
```