NAG Fortran 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             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 part of y.

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;

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

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;

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

and

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

4   References

Huber P J (1981) *Robust Statistics* Wiley

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

5 Parameters

1:     N – INTEGER   *Input*

      On entry: the number of observations, n.

      Constraint: $N \geq 2$.

2:     X(N) – *real* array   *Input*

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

3:     ALPHA – *real*   *Input*

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

      Constraint: $0.0 \leq \text{ALPHA} < 0.5$.

4:     TMEAN – *real*   *Output*

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

5:     WMEAN – *real*   *Output*

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

6:     TVAR – *real*   *Output*

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

7:     WVAR – *real*   *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* 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. 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, $N \leq 1$.

IFAIL = 2

      On entry, $ALPHA < 0.0$,
      or $ALPHA \geq 0.5$.

7   Accuracy

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

8   Further Comments

The time taken by the routine is proportional to n.

9   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.

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.

```
*      G07DDF Example Program Text
*      Mark 14 Release.  NAG Copyright 1989.
*      .. Parameters ..
      INTEGER         NIN, NOUT
      PARAMETER       (NIN=5,NOUT=6)
      INTEGER         NMAX
      PARAMETER       (NMAX=1000)
*      .. Local Scalars ..
      real            ALPHA, PROPN, TMEAN, TVAR, WMEAN, WVAR
      INTEGER         I, IFAIL, K, N
*      .. Local Arrays ..
      real            SX(NMAX), X(NMAX)
*      .. External Subroutines ..
      EXTERNAL        G07DDF
*      .. Intrinsic Functions ..
      INTRINSIC       real
*      .. Executable Statements ..
      WRITE (NOUT,*) 'G07DDF Example Program Results'
*      Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N, (X(I),I=1,N), ALPHA
      IFAIL = 0
*
      CALL G07DDF(N,X,ALPHA,TMEAN,WMEAN,TVAR,WVAR,K,SX,IFAIL)
*
      PROPN = real(K)/N
      PROPN = 100.0e0 - 200.0e0*PROPN
      WRITE (NOUT,*)
      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
      STOP
*
99999 FORMAT (1X,A,F6.2,A)
99998 FORMAT (1X,A,F11.4)
      END
```

9.2 Program Data

```
G07DDF Example Program Data
 16
 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
 0.15
```

9.3 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
```