NAG Library Routine Document

G07GBF

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

G07GBF returns a flag indicating whether a single data point is an outlier as defined by Peirce's criterion.

2 Specification

```
FUNCTION G07GBF (N, E, VAR1, VAR2, X, LX, UX, IFAIL)
LOGICAL G07GBF
INTEGER            N, IFAIL
REAL (KIND=nag_wp) E, VAR1, VAR2, X, LX, UX
```

3 Description

G07GBF tests a potential outlying value using Peirce's criterion. Let

e denote a vector of n residuals with mean zero and variance σ^2 obtained from fitting some model M to a series of data y,

$\tilde{e}$ denote the largest absolute residual in e, i.e., $|\tilde{e}| \geq |e_i|$ for all i, and let $\tilde{y}$ denote the data series y with the observation corresponding to $\tilde{e}$ having been omitted,

$\tilde{\sigma}^2$ denote the residual variance on fitting model M to $\tilde{y}$,

λ denote the ratio of $\tilde{\sigma}$ and σ with $\lambda = \frac{\tilde{\sigma}}{\sigma}$.

Peirce's method flags $\tilde{e}$ as a potential outlier if $|\tilde{e}| \geq x$, where $x = \sigma^2 z$ and z is obtained from the solution of

$$R = \lambda^{1-n}\frac{(n-1)^{n-1}}{n^n} \tag{1}$$

where

$$R = 2\exp\left(\left(\frac{z^2-1}{2}\right)(1-\Phi(z))\right) \tag{2}$$

and Φ is the cumulative distribution function for the standard Normal distribution.

Unlike G07GAF, both σ^2 and $\tilde{\sigma}^2$ must be supplied and therefore no assumptions are made about the nature of the relationship between these two quantities. Only a single potential outlier is tested for at a time.

This routine uses an algorithm described in E04ABF/E04ABA to refine a lower, l, and upper, u, limit for x. This refinement stops when $|\tilde{e}| < l$ or $|\tilde{e}| > u$.

4 References

Gould B A (1855) On Peirce's criterion for the rejection of doubtful observations, with tables for facilitating its application *The Astronomical Journal* **45**

Peirce B (1852) Criterion for the rejection of doubtful observations *The Astronomical Journal* **45**

5 Arguments

1: N – INTEGER *Input*

On entry: n, the number of observations.

Constraint: N ≥ 3.

2: E – REAL (KIND=nag_wp) *Input*

On entry: $\tilde{e}$, the value being tested.

3: VAR1 – REAL (KIND=nag_wp) *Input*

On entry: σ^2, the residual variance on fitting model M to y.

Constraint: VAR1 > 0.0.

4: VAR2 – REAL (KIND=nag_wp) *Input*

On entry: $\tilde{\sigma}^2$, the residual variance on fitting model M to $\tilde{y}$.

Constraints:

VAR2 > 0.0;
VAR2 $<$ VAR1.

5: X – REAL (KIND=nag_wp) *Output*

On exit: an estimated value of x, the cutoff that indicates an outlier.

6: LX – REAL (KIND=nag_wp) *Output*

On exit: l, the lower limit for x.

7: UX – REAL (KIND=nag_wp) *Output*

On exit: u, the upper limit for x.

8: 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 $= \langle value \rangle$.
Constraint: N ≥ 3.

IFAIL $= 3$

On entry, VAR1 $= \langle value \rangle$.
Constraint: VAR1 > 0.0.

IFAIL $= 4$

On entry, VAR1 $= \langle value \rangle$, VAR2 $= \langle value \rangle$.
Constraint: VAR2 $<$ VAR1.

On entry, VAR2 $= \langle value \rangle$.
Constraint: VAR2 > 0.0.

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

Not applicable.

8 Parallelism and Performance

G07GBF is not threaded in any implementation.

9 Further Comments

None.

10 Example

This example reads in a series of values and variances and checks whether each is a potential outlier.

The dataset used is from Peirce's original paper and consists of fifteen observations on the vertical semidiameter of Venus. Each subsequent line in the dataset, after the first, is the result of dropping the observation with the highest absolute value from the previous data and recalculating the variance.

10.1 Program Text

```
    Program g07gbfe

!     G07GBF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g07gbf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
```

```
!     .. Local Scalars ..
      Real (Kind=nag_wp)              :: e, lx, ux, var1, var2, x
      Integer                         :: ifail, n
      Logical                         :: outlier
!     .. Executable Statements ..
      Write (nout,*) 'G07GBF Example Program Results'
      Write (nout,*)

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

d_lp: Do
!       Read in the sample size, variances and value to test
        Read (nin,*,Iostat=ifail) n, e, var1, var2
        If (ifail/=0) Then
          Exit d_lp
        End If

!       Check whether E is a potential outlier
        ifail = 0
        outlier = g07gbf(n,e,var1,var2,x,lx,ux,ifail)

!       Display results
        Write (nout,99999) 'Sample size                              :', n
        Write (nout,99998) 'Largest absolute residual (E)            :', e
        Write (nout,99998) 'Variance for whole sample                :', var1
        Write (nout,99998) 'Variance excluding E                     :', var2
        Write (nout,99998) 'Estimate for cutoff (X)                  :', x
        Write (nout,99998) 'Lower limit for cutoff (LX)              :', lx
        Write (nout,99998) 'Upper limit for cutoff (UX)              :', ux
        If (outlier) Then
          Write (nout,*) 'E is a potential outlier'
        Else
          Write (nout,*) 'E does not appear to be an outlier'
        End If
        Write (nout,*)
      End Do d_lp

99999 Format (1X,A,1X,I10)
99998 Format (1X,A,1X,F10.3)
    End Program g07gbfe
```

10.2 Program Data

```
G07GBF Example Program Data
15 -1.40 0.303 0.161   :: N, E, VAR1, VAR2
14  1.01 0.161 0.103   :: N, E, VAR1, VAR2
13  0.63 0.103 0.080   :: N, E, VAR1, VAR2
```

10.3 Program Results

```
 G07GBF Example Program Results

 Sample size                          :         15
 Largest absolute residual (E)        :     -1.400
 Variance for whole sample            :      0.303
 Variance excluding E                 :      0.161
 Estimate for cutoff (X)              :      0.000
 Lower limit for cutoff (LX)          :      0.000
 Upper limit for cutoff (UX)          :      0.000
 E is a potential outlier

 Sample size                          :         14
 Largest absolute residual (E)        :      1.010
 Variance for whole sample            :      0.161
 Variance excluding E                 :      0.103
 Estimate for cutoff (X)              :      0.105
 Lower limit for cutoff (LX)          :      0.100
 Upper limit for cutoff (UX)          :      0.110
 E is a potential outlier
```

```
Sample size                          :         13
Largest absolute residual (E)        :      0.630
Variance for whole sample            :      0.103
Variance excluding E                 :      0.080
Estimate for cutoff (X)              :      1.059
Lower limit for cutoff (LX)          :      1.011
Upper limit for cutoff (UX)          :      1.155
E does not appear to be an outlier
```