

NAG Toolbox

nag_univar_outlier_peirce_2var (g07gb)

1 Purpose

nag_univar_outlier_peirce_2var (g07gb) returns a flag indicating whether a single data point is an outlier as defined by Peirce's criterion.

2 Syntax

```
[result, x, lx, ux, ifail] = nag_univar_outlier_peirce_2var(n, e, var1, var2)
[result, x, lx, ux, ifail] = g07gb(n, e, var1, var2)
```

3 Description

nag_univar_outlier_peirce_2var (g07gb) 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

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

where

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

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

Unlike nag_univar_outlier_peirce_1var (g07ga), 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 function uses an algorithm described in nag_opt_one_var_func (e04ab) 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 Parameters

5.1 Compulsory Input Parameters

- 1: **n** – INTEGER
 n , the number of observations.
Constraint: $\mathbf{n} \geq 3$.
- 2: **e** – REAL (KIND=nag_wp)
 \tilde{e} , the value being tested.
- 3: **var1** – REAL (KIND=nag_wp)
 σ^2 , the residual variance on fitting model M to y .
Constraint: $\mathbf{var1} > 0.0$.
- 4: **var2** – REAL (KIND=nag_wp)
 $\tilde{\sigma}^2$, the residual variance on fitting model M to \tilde{y} .
Constraints:
 $\mathbf{var2} > 0.0$;
 $\mathbf{var2} < \mathbf{var1}$.

5.2 Optional Input Parameters

None.

5.3 Output Parameters

- 1: **result**
The result of the function.
- 2: **x** – REAL (KIND=nag_wp)
An estimated value of x , the cutoff that indicates an outlier.
- 3: **lx** – REAL (KIND=nag_wp)
 l , the lower limit for x .
- 4: **ux** – REAL (KIND=nag_wp)
 u , the upper limit for x .
- 5: **ifail** – INTEGER
ifail = 0 unless the function detects an error (see Section 5).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail = 1

Constraint: $\mathbf{n} \geq 3$.

ifail = 3

Constraint: $\mathbf{var1} > 0.0$.

ifail = 4

Constraint: **var2** < **var1**.

Constraint: **var2** > 0.0.

ifail = -99

An unexpected error has been triggered by this routine. Please contact NAG.

ifail = -399

Your licence key may have expired or may not have been installed correctly.

ifail = -999

Dynamic memory allocation failed.

7 Accuracy

Not applicable.

8 Further Comments

None.

9 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.

9.1 Program Text

```
function g07gb_example

fprintf('g07gb example results\n\n');

ns = [nag_int(15); 14; 13];
es = [-1.4; 1.01; 0.63];
var1s = [0.303; 0.161; 0.103];
var2s = [0.161; 0.103; 0.08];

for i = 1:numel(ns)
    % Check whether es(i) is a potential outlier
    [outlier, x, lx, ux, ifail] = ...
    g07gb( ...
        ns(i), es(i), var1s(i), var2s(i));

    % Display results
    fprintf('\nSample size                : %10d\n', ns(i));
    fprintf('Largest absolute residual (E)    : %10.3f\n', es(i));
    fprintf('Variance for whole sample              : %10.3f\n', var1s(i));
    fprintf('Variance excluding E                   : %10.3f\n', var2s(i));
    fprintf('Estimate for cutoff (X)                 : %10.3f\n', x);
    fprintf('Lower limit for cutoff (LX)             : %10.3f\n', lx);
    fprintf('Upper limit for cutoff (UX)             : %10.3f\n', ux);
    if outlier
        fprintf('E is a potential outlier\n');
    else
        fprintf('E does not appear to be an outlier\n');
    end
end
end
```

9.2 Program Results

g07gb example 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		
