

NAG Library Function Document

nag_outlier_peirce_two_var (g07gbc)

1 Purpose

nag_outlier_peirce_two_var (g07gbc) returns a flag indicating whether a single data point is an outlier as defined by Peirce's criterion.

2 Specification

```
#include <nag.h>
#include <nagg07.h>
Nag_Boolean nag_outlier_peirce_two_var (Integer n, double e, double var1,
                                         double var2, double *x, double *lx, double *ux, NagError *fail)
```

3 Description

nag_outlier_peirce_two_var (g07gbc) 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_outlier_peirce (g07gac), 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_no_deriv (e04abc) 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: $n \geq 3$.
- 2: **e** – double *Input*
On entry: \tilde{e} , the value being tested.
- 3: **var1** – double *Input*
On entry: σ^2 , the residual variance on fitting model M to y .
Constraint: **var1** > 0.0.
- 4: **var2** – double *Input*
On entry: $\tilde{\sigma}^2$, the residual variance on fitting model M to \tilde{y} .
Constraints:
var2 > 0.0;
var2 < **var1**.
- 5: **x** – double * *Output*
On exit: an estimated value of x , the cutoff that indicates an outlier.
- 6: **lx** – double * *Output*
On exit: l , the lower limit for x .
- 7: **ux** – double * *Output*
On exit: u , the upper limit for x .
- 8: **fail** – NagError * *Input/Output*
The NAG error argument (see Section 3.6 in the Essential Introduction).

6 Error Indicators and Warnings

NE_BAD_PARAM

On entry, argument $\langle value \rangle$ had an illegal value.

NE_INT

On entry, **n** = $\langle value \rangle$.
Constraint: **n** ≥ 3 .

NE_INTERNAL_ERROR

An internal error has occurred in this function. Check the function call and any array sizes. If the call is correct then please contact NAG for assistance.

NE_REAL

On entry, **var1** = $\langle value \rangle$.
Constraint: **var1** > 0.0.
On entry, **var2** = $\langle value \rangle$.
Constraint: **var2** > 0.0.

NE_REAL_2

On entry, **var1** = $\langle value \rangle$, **var2** = $\langle value \rangle$.
 Constraint: **var2** < **var1**.

7 Accuracy

Not applicable.

8 Parallelism and Performance

Not applicable.

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

```

/* nag_outlier_peirce_two_var (g07gbc) Example Program.
 *
 * Copyright 2009, Numerical Algorithms Group.
 *
 * Mark 23, 2011.
 */
/* Pre-processor includes */
#include <stdio.h>
#include <math.h>
#include <nag.h>
#include <nag_stdlib.h>
#include <nagg07.h>

int main(void)
{
    /* Integer scalar and array declarations */
    Integer n, exit_status;

    /* NAG structures and types */
    Nag_Boolean outlier;
    NagError fail;

    /* Double scalar and array declarations */
    double e, var1, var2, x, lx, ux;

    /* Initialise the error structure */
    INIT_FAIL(fail);

    exit_status = 0;

    printf(
        "nag_outlier_peirce_two_var (g07gbc) Example Program Results\n");

    /* Skip headings in data file */
    scanf("%*[^\\n] ");

```

```

/* Loop through all the lines in the input file, reading in the sample size,
   variances and value to test */
while (scanf("%ld %lf %lf %lf%*[\n] ", &n, &e, &var1, &var2)
      != EOF) {

    /* Use nag_outlier_peirce_two_var (g07gbc) to check whether e is a
       potential outlier */
    outlier = nag_outlier_peirce_two_var(n, e, var1, var2, &x, &lx, &ux, &fail);
    if (fail.code != NE_NOERROR) {
        printf("Error from nag_outlier_peirce_two_var (g07gbc).\n%s\n",
              fail.message);
        exit_status = 1;
        goto END;
    }

    /* Display the results */
    printf(" Sample size                : %10ld\n",
           n);
    printf(" Largest absolute residual (E)      : %10.3f\n", e);
    printf(" Variance for whole sample            : %10.3f\n", var1);
    printf(" Variance excluding E                  : %10.3f\n", var2);
    printf(" Estimate for cutoff (X)              : %10.3f\n", x);
    printf(" Lower limit for cutoff (LX)           : %10.3f\n", lx);
    printf(" Upper limit for cutoff (UX)           : %10.3f\n", ux);
    if (outlier)
        printf(" E is a potential outlier\n");
    else
        printf(" E does not appear to be an outlier\n");
    printf("\n");
}

END:

return exit_status;
}

```

10.2 Program Data

```

nag_outlier_peirce_two_var (g07gbc) 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

```

nag_outlier_peirce_two_var (g07gbc) 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
