

NAG Toolbox

nag_univar_robust_1var_trimmed (g07dd)

1 Purpose

nag_univar_robust_1var_trimmed (g07dd) calculates the trimmed and Winsorized means of a sample and estimates of the variances of the two means.

2 Syntax

```
[tmean, wmean, tvar, wvar, k, sx, ifail] = nag_univar_robust_1var_trimmed(x,
alpha, 'n', n)
```

```
[tmean, wmean, tvar, wvar, k, sx, ifail] = g07dd(x, alpha, 'n', n)
```

3 Description

nag_univar_robust_1var_trimmed (g07dd) calculates the α -trimmed mean and α -Winsorized mean for a given α , as described below.

Let x_i , for $i = 1, 2, \dots, n$ 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:

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

nag_univar_robust_1var_trimmed (g07dd) 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;

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

and

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

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 Parameters

5.1 Compulsory Input Parameters

- 1: **x(n)** – REAL (KIND=nag_wp) array
The sample observations, x_i , for $i = 1, 2, \dots, n$.
- 2: **alpha** – REAL (KIND=nag_wp)
 α , the proportion of observations to be trimmed at each end of the sorted sample.
Constraint: $0.0 \leq \mathbf{alpha} < 0.5$.

5.2 Optional Input Parameters

- 1: **n** – INTEGER
Default: the dimension of the array **x**.
 n , the number of observations.
Constraint: $\mathbf{n} \geq 2$.

5.3 Output Parameters

- 1: **tmean** – REAL (KIND=nag_wp)
The α -trimmed mean, \bar{x}_t .
- 2: **wmean** – REAL (KIND=nag_wp)
The α -Winsorized mean, \bar{x}_w .
- 3: **tvar** – REAL (KIND=nag_wp)
Contains an estimate of the variance of the trimmed mean.
- 4: **wvar** – REAL (KIND=nag_wp)
Contains an estimate of the variance of the Winsorized mean.
- 5: **k** – INTEGER
Contains the number of observations trimmed at each end, k .
- 6: **sx(n)** – REAL (KIND=nag_wp) array
Contains the sample observations sorted into ascending order.
- 7: **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

On entry, $\mathbf{n} \leq 1$.

ifail = 2

On entry, **alpha** < 0.0,
or **alpha** \geq 0.5.

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

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

8 Further Comments

The time taken 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

```
function g07dd_example

fprintf('g07dd example results\n\n');

x = [26; 12; 9; 2; 5; 6; 8; 14;
     7; 3; 1; 11; 10; 4; 17; 21];
alpha = 0.15;

[tmean, wmean, tvar, wvar, k, sx, ifail] = ...
    g07dd(x, alpha);

% Calculate proportion of data cut
propn = 100*(1-2*double(k)/numel(x));

fprintf('Statistics from middle %6.2f%% of data\n\n',propn);
fprintf('          Trimmed-mean = %11.4f\n', tmean);
fprintf('    Variance of Trimmed-mean = %11.4f\n\n', tvar);
fprintf('          Winsorized-mean = %11.4f\n', wmean);
fprintf('    Variance of Winsorized-mean = %11.4f\n', wvar);
```

9.2 Program Results

```
g07dd example 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
```
