

NAG Toolbox for MATLAB

g07dd

1 Purpose

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] = g07dd(x, alpha, 'n', n)
```

3 Description

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:

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

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)** – double array

The sample observations, x_i , for $i = 1, 2, \dots, n$.

2: **alpha** – double scalar

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** – int32 scalar

n , the number of observations.

Constraint: $\mathbf{n} \geq 2$.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **tmean** – double scalar

The α -trimmed mean, $\bar{x}_t$.

2: **wmean** – double scalar

The α -Winsorized mean, $\bar{x}_w$.

3: **tvar** – double scalar

Contains an estimate of the variance of the trimmed mean.

4: **wvar** – double scalar

Contains an estimate of the variance of the Winsorized mean.

5: **k** – int32 scalar

Contains the number of observations trimmed at each end, k .

6: **sx(n)** – double array

Contains the sample observations sorted into ascending order.

7: **ifail** – int32 scalar

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail = 1

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

ifail = 2

On entry, $\mathbf{alpha} < 0.0$,
or $\mathbf{alpha} \geq 0.5$.

7 Accuracy

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

8 Further Comments

The time taken is proportional to n .

9 Example

```
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)  
  
tmean =  
      8.8333  
wmean =  
      9.1250  
tvar =  
      1.5434  
wvar =  
      1.5381  
k =  
      2  
sx =  
      1  
      2  
      3  
      4  
      5  
      6  
      7  
      8  
      9  
     10  
     11  
     12  
     14  
     17  
     21  
     26  
ifail =  
      0
```