NAG Toolbox for MATLAB

g01ad

1 Purpose

g01ad calculates the mean, standard deviation and coefficients of skewness and kurtosis for data grouped in a frequency distribution.

2 Syntax

```
[xmean, s2, s3, s4, n, ifail] = g01ad(x, ifreq, 'k', k)
```

3 Description

The input data consist of a univariate frequency distribution, denoted by f_i, for $i = 1, 2, \ldots, k-1$, and the boundary-values of the classes x_i, for $i = 1, 2, \ldots, k$. Thus the frequency associated with the interval (x_i, x_{i+1}) is f_i, and g01ad assumes that all the values in this interval are concentrated at the point

$$y_i = (x_{i+1} + x_i)/2, \qquad i = 1, 2, \ldots, k-1.$$

The following quantities are calculated:

(a) total frequency,

$$n = \sum_{i=1}^{k-1} f_i.$$

(b) mean,

$$\bar{y} = \frac{\sum_{i=1}^{k-1} f_i y_i}{n}.$$

(c) standard deviation,

$$s_2 = \sqrt{\frac{\sum_{i=1}^{k-1} f_i (y_i - \bar{y})^2}{(n-1)}}, \qquad n \geq 2.$$

(d) coefficient of skewness,

$$s_3 = \frac{\sum_{i=1}^{k-1} f_i (y_i - \bar{y})^3}{(n-1) \times s_2^3}, \qquad n \geq 2.$$

(e) coefficient of kurtosis,

$$s_4 = \frac{\sum_{i=1}^{k-1} f_i (y_i - \bar{y})^4}{(n-1) \times s_2^4} - 3, \qquad n \geq 2.$$

The function has been developed primarily for groupings of a continuous variable. If, however, the function is to be used on the frequency distribution of a discrete variable, taking the values $y_1, \ldots, y_{k-1}$, then the boundary-values for the classes may be defined as follows:

(i) for $k > 2$,

$$x_1 = (3y_1 - y_2)/2$$
$$x_j = \left(y_{j-1} + y_j\right)/2, \qquad j = 2, \ldots, k-1$$
$$x_k = (3y_{k-1} - y_{k-2})/2$$

(ii) for $k = 2$,

$$x_1 = y_1 - a \qquad \text{and} \qquad x_2 = y_1 + a \qquad \text{for any } a > 0.$$

4 References

None.

5 Parameters

5.1 Compulsory Input Parameters

1: **x(k) – double array**

The elements of **x** must contain the boundary-values of the classes in ascending order, so that class i is bounded by the values in $\mathbf{x}(i)$ and $\mathbf{x}(i+1)$, for $i = 1, 2, \ldots, k-1$.

Constraint: $\mathbf{x}(i) < \mathbf{x}(i+1)$, for $i = 1, 2, \ldots, k-1$.

2: **ifreq(k) – int32 array**

The ith element of **ifreq** must contain the frequency associated with the ith class, for $i = 1, 2, \ldots, k-1$. $\mathbf{ifreq}(k)$ is not used by the function.

Constraint: $\mathbf{ifreq}(i) \geq 0$, for $i = 1, 2, \ldots, k-1$ and $\sum_{i=1}^{k-1} \mathbf{ifreq}(i) > 0$.

5.2 Optional Input Parameters

1: **k – int32 scalar**

Default: The dimension of the arrays **x**, **ifreq**. (An error is raised if these dimensions are not equal.)

k, the number of class boundaries, which is one more than the number of classes of the frequency distribution.

Constraint: $\mathbf{k} > 1$.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **xmean – double scalar**

The mean value, $\bar{y}$.

2: **s2 – double scalar**

The standard deviation, s_2.

3: **s3 – double scalar**

The coefficient of skewness, s_3.

4: **s4 – double scalar**

The coefficient of kurtosis, s_4.

5: **n – int32 scalar**

The total frequency, n.

6: **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, $\mathbf{k} \leq 1$.

ifail $= 2$

On entry, the boundary-values of the classes in **x** are not in ascending order.

ifail $= 3$

On entry, $\sum_{i=1}^{k-1} \mathbf{ifreq}(i) = 0$ or $\mathbf{ifreq}(i) < 0$ for some i, for $i = 1, 2, \ldots, k-1$.

ifail $= 4$

The total frequency, n, is less than 2, hence the quantities s_2, s_3 and s_4 cannot be calculated.

7 Accuracy

The method used is believed to be stable.

8 Further Comments

The time taken by g01ad increases linearly with k.

9 Example

```
x = [9.300000000000001;
     12;
     14;
     16;
     18;
     20;
     22;
     24;
     26;
     28;
     30;
     32;
     34;
     36;
     39.7];
ifreq = [int32(3);
     int32(19);
     int32(52);
```

```
     int32(96);
     int32(121);
     int32(115);
     int32(86);
     int32(70);
     int32(49);
     int32(31);
     int32(16);
     int32(6);
     int32(8);
     int32(7);
     int32(0)];
[xmean, s2, s3, s4, n, ifail] = g01ad(x, ifreq)
```

```
xmean =
   21.4932
s2 =
    4.9325
s3 =
    0.7072
s4 =
    0.5738
n =
         679
ifail =
           0
```