NAG Toolbox for MATLAB

g01as

1 Purpose

g01as produces a specified number of box and whisker plots on a character printing device, with a chosen number of character positions in each direction.

2 Syntax

```
[plot, work, ifail] = g01as(prt, n, x, nstepx, nstepy, 'm', m)
```

3 Description

g01as produces a series of box and whisker plots representing m data batches each of size n_i, for $i = 1, 2, \ldots, m$. A box and whisker plot is a diagrammatic representation of the five-point summary of a data batch. The plot consists of a box spanning the hinges with the median indicated by a third line and two whiskers to represent the extreme values. The five-point summary is calculated internally and is returned in the workspace array.

The plot is returned in the character array **plot**. The size of the plot may be controlled using the parameters **nstepx** and **nstepy**. Optionally the plot can be output to an external file, in which case output is directed to the current advisory message unit as defined by x04ab.

An axis corresponding to the y axis is drawn and annotated and data points are plotted to the nearest character position.

4 References

Erickson B H and Nosanchuk T A 1985 *Understanding Data* Open University Press, Milton Keynes

Tukey J W 1977 *Exploratory Data Analysis* Addison–Wesley

5 Parameters

5.1 Compulsory Input Parameters

1: **prt – string**

Indicates whether the box and whisker plot is to be output to an external file.

prt = 'N'

The box and whisker plot is not output to an external file.

prt = 'P'

The box and whisker plot is output to the current advisory message unit as defined by x04ab.

Constraint: **prt** = 'P' or 'N'.

2: **n(m) – int32 array**

n(i) contains the number of observations in the ith batch, n_i, for $i = 1, 2, \ldots, m$.

If $n_i < 5$ the ith batch is omitted from the plot.

Constraint: at least one **n**(i), for $i = 1, 2, \ldots m$, must be greater than or equal to 5.

3: **x(ldx,m) – double array**

ldx, the first dimension of the array, must be at least $\max\{\mathbf{n}(i)\}$.

The ith column of **x** must contain the data for the ith batch, that is $\mathbf{x}(j,i)$ must contain the jth observation of the ith batch, for $j = 1,2,\ldots,n_i$; $i = 1,2,\ldots,m$.

4: **nstepx – int32 scalar**

the number of character positions to be plotted in the x-direction.

Constraint: **nstepx** $\geq \max(19,(15 \times \mathbf{m}/4 + 9))$.

5: **nstepy – int32 scalar**

The number of character positions to be plotted in the y-direction.

Constraint: **nstepy** ≥ 9.

5.2 Optional Input Parameters

1: **m – int32 scalar**

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

m, the number of data batches that are to be represented.

Constraint: **m** > 0.

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, ldplot, iwork

5.4 Output Parameters

1: **plot(ldplot,nstepx) – string array**

Contains the box and whisker plots.

2: **work(5 × m) – double array**

work(j), for $j = (i-1) \times 5 + 1, (i-1) \times 5 + 2, \ldots, (i-1) \times 5 + 5$, contains the five-point summary of the ith batch.

3: **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, $\mathbf{n}(i) < 5$ for some i, for $i = 1,2,\ldots,m$. For each batch where this occurs, 5 crosses are plotted in a vertical line to indicate that insufficient data was provided to produce a five-point summary and box-plot for that particular batch.

ifail $= 2$

On entry, **nstepx** $< \max(19, 15 \times \mathbf{m}/4 + 9)$. This indicates that the data region defined by **nstepx** is too small to produce the required plot.

ifail $= 3$

On entry, **nstepy** < 9.

ifail $= 4$

On entry, **ldplot** $<$ **nstepy**.

ifail $= 5$

On entry, **prt** $\neq$ 'P' or 'N'.

ifail $= 6$

On entry, **ldx** $< \max(\mathbf{n}(i))$, for $i = 1, 2, \ldots, m$.

ifail $= 7$

The number of observations in all batches is less than 5.

ifail $= 8$

On entry, the data values are all identical.

7 Accuracy

If the range of observations in a particular batch is too small to allow each item of the five-point summary to be plotted separately, then a sequence of stars are plotted at the median point of the batch to indicate that the full box-plot could not be plotted.

8 Further Comments

The time taken by g01as increases with m and n_i, for $i = 1, 2, \ldots, m$.

9 Example

```
prt = 'Print';
n = [int32(5);
     int32(6);
     int32(8);
     int32(8);
     int32(7)];
x = [-9, -5.6, -9, 12, -5;
     -7.3, 12, 12, 10, 8;
     -4.9, -9, -6, 10, 1.5;
     -2.4, -3.9, -3, 8, -3.2;
     -0.6, -2.4, 0, 6, -3.2;
     0, -7.3, 3, 4, 6.2;
     0, 0, 6, 4, 6.2;
     0, 0, 9, -9, 0];
nstepx = int32(55);
nstepy = int32(21);
[plot, work, ifail] = g01as(prt, n, x, nstepx, nstepy)
```

```
0.1E+02+           ---      ---      ---
         :           :        :        :
         :           :        :      -----
  0.8E+01+           :        :      :   :
         :           :        -----  :   :     ---
         :           :        :   :  :---:       :
         :           :        :   :  :   :     -----
  0.4E+01+           :        :   :  :   :     :   :
         :           :        :   :  -----     :   :
         :           :        :   :    :       :   :
         :           :        :---:    :       :---:
         :           :        :   :    :       :   :
 -0.6E+00+  ---      :        :   :    :       :   :
         :   :       :        :   :    :       :   :
```

```
         : -----    -----    :   :      :      -----
         : :   :    :   :    :   :      :        :
 -0.5E+01+ :---:    :---:    -----      :       ---
         : :   :    :   :      :        :
         : -----    -----      :        :
         :   :        :        :        :
 -0.9E+01+  ---      ---      ---      ---
plot =
     array elided
work =
   -9.0000
   -7.3000
   -4.9000
   -2.4000
   -0.6000
   -9.0000
   -7.3000
   -4.7500
   -2.4000
   12.0000
   -9.0000
   -4.5000
    1.5000
    7.5000
   12.0000
   -9.0000
    4.0000
    7.0000
   10.0000
   12.0000
   -5.0000
   -3.2000
    1.5000
    6.2000
    8.0000
ifail =
           0
```