NAG Toolbox

nag_stat_frequency_table (g01ae)

1 Purpose

nag_stat_frequency_table (g01ae) constructs a frequency distribution of a variable, according to either user-supplied, or function-calculated class boundary values.

2 Syntax

```
[cb, ifreq, xmin, xmax, ifail] = nag_stat_frequency_table(k, x, 'n', n, 'cb', cb)
[cb, ifreq, xmin, xmax, ifail] = g01ae(k, x, 'n', n, 'cb', cb)
```

Note: the interface to this routine has changed since earlier releases of the toolbox:

At Mark 23: **iclass** is no longer an input parameter; **cb** was made optional; **k** was made a compulsory input parameter.

3 Description

The data consists of a sample of n observations of a continuous variable, denoted by x_i, for $i = 1, 2, \ldots, n$. Let $a = \min(x_1, \ldots, x_n)$ and $b = \max(x_1, \ldots, x_n)$.

nag_stat_frequency_table (g01ae) constructs a frequency distribution with $k(> 1)$ classes denoted by f_i, for $i = 1, 2, \ldots, k$.

The boundary values may be either user-supplied, or function-calculated, and are denoted by y_j, for $j = 1, 2, \ldots, k-1$.

If the boundary values of the classes are to be function-calculated, then they are determined in one of the following ways:

(a) if $k > 2$, the range of x values is divided into $k-2$ intervals of equal length, and two extreme intervals, defined by the class boundary values $y_1, y_2, \ldots, y_{k-1}$;

(b) if $k = 2$, $y_1 = \frac{1}{2}(a+b)$.

However formed, the values $y_1, \ldots, y_{k-1}$ are assumed to be in ascending order. The class frequencies are formed with

$f_1 =$ the number of x values in the interval $(-\infty, y_1)$

$f_i =$ the number of x values in the interval $[y_{i-1}, y_i)$, $\quad i = 2, \ldots, k-1$

$f_k =$ the number of x values in the interval $[y_{k-1}, \infty)$,

where [means inclusive, and) means exclusive. If the class boundary values are function-calculated and $k > 2$, then $f_1 = f_k = 0$, and y_1 and y_{k-1} are chosen so that $y_1 < a$ and $y_{k-1} > b$.

If a frequency distribution is required for a discrete variable, then it is suggested that you supply the class boundary values; function-calculated boundary values may be slightly imprecise (due to the adjustment of y_1 and y_{k-1} outlined above) and cause values very close to a class boundary to be assigned to the wrong class.

4 References

None.

5 Parameters

5.1 Compulsory Input Parameters

1: **k** – INTEGER

k, the number of classes desired in the frequency distribution. Whether or not class boundary values are user-supplied, **k** must include the two extreme classes which stretch to $\pm\infty$.

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

2: **x**(**n**) – REAL (KIND=nag_wp) array

The sample of observations of the variable for which the frequency distribution is required, x_i, for $i = 1, 2, \ldots, n$. The values may be in any order.

5.2 Optional Input Parameters

1: **n** – INTEGER

Default: the dimension of the array **x**.

n, the number of observations.

Constraint: $\mathbf{n} \geq 1$.

2: **cb**(**k**) – REAL (KIND=nag_wp) array

If **cb** is not supplied, nag_stat_frequency_table (g01ae) calculates $k - 1$ class boundary values.

If **cb** is supplied, the first $k - 1$ elements of **cb** must contain the class boundary values you supplied, in ascending order.

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

5.3 Output Parameters

1: **cb**(**k**) – REAL (KIND=nag_wp) array

The first $k - 1$ elements of **cb** contain the class boundary values in ascending order.

2: **ifreq**(**k**) – INTEGER array

The elements of **ifreq** contain the frequencies in each class, f_i, for $i = 1, 2, \ldots, k$. In particular $\mathbf{ifreq}(1)$ contains the frequency of the class up to $\mathbf{cb}(1)$, f_1, and $\mathbf{ifreq}(k)$ contains the frequency of the class greater than $\mathbf{cb}(k-1)$, f_k.

3: **xmin** – REAL (KIND=nag_wp)

The smallest value in the sample, a.

4: **xmax** – REAL (KIND=nag_wp)

The largest value in the sample, b.

5: **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, $\mathbf{k} < 2$.

ifail $= 2$

On entry, **n** < 1.

ifail $= 3$

On entry, the user-supplied class boundary values are not in ascending order.

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 method used is believed to be stable.

8 Further Comments

The time taken by nag_stat_frequency_table (g01ae) increases with **k** and **n**. It also depends on the distribution of the sample observations.

9 Example

This example summarises a number of datasets. For each dataset the sample observations and optionally class boundary values are read. nag_stat_frequency_table (g01ae) is then called and the frequency distribution and largest and smallest observations printed.

9.1 Program Text

```
function g01ae_example

fprintf('g01ae example results\n\n');

x = [22.3; 21.6; 22.6; 22.4; 22.4; 22.4; 22.1; 21.9; 23.1; 23.4; 23.4;
     22.6; 22.5; 22.5; 22.1; 22.6; 22.3; 22.4; 21.8; 22.3; 22.1; 23.6;
     20.8; 22.2; 23.1; 21.1; 21.7; 21.4; 21.6; 22.5; 21.2; 22.6; 22.2;
     22.2; 21.4; 21.7; 23.2; 23.1; 22.3; 22.3; 21.1; 21.4; 21.5; 21.8;
     22.8; 21.4; 20.7; 21.6; 23.2; 23.6; 22.7; 21.7; 23.0; 21.9; 22.6;
     22.1; 22.2; 23.4; 21.5; 23.0; 22.8; 21.4; 23.2; 21.8; 21.2; 22.0;
     22.4; 22.8; 23.2; 23.6];

k = nag_int(7);
[cb, ifreq, xmin, xmax, ifail] = g01ae(k, x);

fprintf('Number of cases     %3d\n',size(x,1));
fprintf('Number of classes   %3d\n\n',k);
fprintf('Routine-supplied class boundaries\n\n');
fprintf('        Class            Frequency\n');
fprintf('%9s to%7.2f%14d\n', 'Up', cb(1), ifreq(1));
for i=2:k-1
  fprintf('%7.2f   to%7.2f%14d\n', cb(i-1), cb(i), ifreq(i));
end
fprintf('%7.2f  and%7s%14d\n\n', cb(k-1), 'over', ifreq(k));
fprintf('Total frequency = %5d\n',sum(ifreq));
fprintf('Minimum         = %8.2f\n',xmin);
fprintf('Maximum         = %8.2f\n',xmax);
```

9.2 Program Results

```
     g01ae example results

Number of cases      70
Number of classes     7

Routine-supplied class boundaries

        Class            Frequency
       Up to  20.70              0
  20.70   to  21.28              6
  21.28   to  21.86             16
  21.86   to  22.44             21
  22.44   to  23.02             14
  23.02   to  23.60             13
  23.60  and   over              0

Total frequency =    70
Minimum         =    20.70
Maximum         =    23.60
```