

NAG Toolbox

nag_stat_summary_2var (g01ab)

1 Purpose

nag_stat_summary_2var (g01ab) computes the means, standard deviations, corrected sums of squares and products, maximum and minimum values, and the product-moment correlation coefficient for two variables. Unequal weighting may be given.

2 Syntax

```
[res, ifail] = nag_stat_summary_2var(x1, x2, 'n', n, 'wt', wt)
```

```
[res, ifail] = g01ab(x1, x2, 'n', n, 'wt', wt)
```

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

At Mark 23: **wt** is no longer an output parameter.

3 Description

The data consist of two samples of n observations, denoted by x_i , and y_i , for $i = 1, 2, \dots, n$, with corresponding weights w_i , for $i = 1, 2, \dots, n$.

If no specific weighting is given, then each w_i is set to 1.0 in nag_stat_summary_2var (g01ab).

The quantities calculated are:

- (a) The sum of weights,

$$W = \sum_{i=1}^n w_i.$$

- (b) The means,

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{W}, \quad \bar{y} = \frac{\sum_{i=1}^n w_i y_i}{W}.$$

- (c) The corrected sums of squares and products

$$c_{11} = \sum_{i=1}^n w_i (x_i - \bar{x})^2$$

$$c_{21} = c_{12} = \sum_{i=1}^n w_i (x_i - \bar{x})(y_i - \bar{y})$$

$$c_{22} = \sum_{i=1}^n w_i (y_i - \bar{y})^2.$$

- (d) The standard deviations

$$s_j = \sqrt{\frac{c_{jj}}{d}}, \quad \text{where } j = 1, 2 \quad \text{and} \quad d = W - \frac{\sum_{i=1}^n w_i^2}{W}.$$

(e) The product-moment correlation coefficient

$$R = \frac{c_{12}}{\sqrt{c_{11}c_{22}}}.$$

(f) The minimum and maximum elements in each of the two samples.

(g) The number of pairs of observations, m , for which $w_i > 0$, i.e., the number of **valid** observations. The quantities in (d) and (e) above will only be computed if $m \geq 2$. All other items are computed if $m \geq 1$.

4 References

None.

5 Parameters

5.1 Compulsory Input Parameters

1: **x1(n)** – REAL (KIND=nag_wp) array

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

2: **x2(n)** – REAL (KIND=nag_wp) array

The observations from the second sample, y_i , for $i = 1, 2, \dots, n$.

5.2 Optional Input Parameters

1: **n** – INTEGER

Default: the dimension of the arrays **x1**, **x2**, **wt**. (An error is raised if these dimensions are not equal.)

n , the number of pairs of observations.

Constraint: $n \geq 1$.

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

If weights are being supplied then the elements of **wt** must contain the weights associated with the observations, w_i , for $i = 1, 2, \dots, n$.

Constraint: if $iwt = 1$, $\mathbf{wt}(i) \geq 0.0$, for $i = 1, 2, \dots, n$.

5.3 Output Parameters

1: **res(13)** – REAL (KIND=nag_wp) array

The elements of **res** contain the following results:

- res(1)** mean of the first sample, \bar{x} ;
- res(2)** mean of the second sample, \bar{y} ;
- res(3)** standard deviation of the first sample, s_1 ;
- res(4)** standard deviation of the second sample, s_2 ;
- res(5)** corrected sum of squares of the first sample, c_{11} ;
- res(6)** corrected sum of products of the two samples, c_{12} ;
- res(7)** corrected sum of squares of the second sample, c_{22} ;
- res(8)** product-moment correlation coefficient, R ;
- res(9)** minimum of the first sample;
- res(10)** maximum of the first sample;
- res(11)** minimum of the second sample;

res(12) maximum of the second sample;
res(13) sum of weights, $\sum_{i=1}^n w_i$ (= **n**, if **wt** is **null** on entry).

2: **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, **n** < 1.

ifail = 2 (*warning*)

The number of valid cases, *m*, is 1, hence the standard deviation, 3(d), and the product-moment correlation coefficient, 3(e), cannot be calculated.

ifail = 3

The number of valid cases, *m*, is 0, or at least one of the weights is negative.

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_summary_2var` (g01ab) increases linearly with *n*.

9 Example

In the program below, `NPROB` determines the number of datasets to be analysed. For each analysis, a set of observations and, optionally, weights, is read and printed. After calling `nag_stat_summary_2var` (g01ab), all the calculated quantities are printed. In the example, there is one set of data, with 29 (unweighted) pairs of observations.

9.1 Program Text

```
function g01ab_example
  fprintf('g01ab example results\n\n');
  x1 = [350    550    380    510    1270    300    2630    810 ...
        140    450    2280   250    540    720    90    480 ...
        180   3160    220    860    300   1460    400    620 ...
        120    780    230   1070    160];
```

```

x2 = [ 47      95      211      122      530      38      278      309 ...
      75      43      407      142      89      159      35      103 ...
      78     969      120      333      73      147      30      100 ...
      55     145      101      468      86];
n = size(x1,2);

[res, ifail] = g01ab(x1, x2);

fprintf('Number of cases      %7d\n',n);
fprintf('Data as input -\n');
v1 = 'Var 1';
v2 = 'Var 2';
fprintf('%12s%12s%12s%12s%12s%12s\n', v1, v2, v1, v2, v1, v2);
fprintf('%12.1f%12.1f%12.1f%12.1f%12.1f%12.1f\n',[x1; x2])
fprintf('\n\n');
fprintf('No. of valid cases %7d\n\n',n);
fprintf('%30s%30s\n', 'Variable 1', 'Variable 2');
fprintf('Mean                %11.1f%30.1f\n',res(1:2));
fprintf('Minimum             %11.1f%30.1f\n',res(9:2:11));
fprintf('Maximum              %11.1f%30.1f\n',res(10:2:12));
fprintf('Standard deviation %11.1f%30.1f\n',res(3:4));
fprintf('Sum of squares        %11.1f%30.1f\n',res(5:2:7));
fprintf('Sum of weights        %26.4f\n',res(13));
fprintf('Sum of products       %26.4f\n',res(6));
fprintf('Correlation           %26.4f\n',res(8));

```

9.2 Program Results

g01ab example results

Number of cases 29

Data as input -

Var 1	Var 2	Var 1	Var 2	Var 1	Var 2
350.0	47.0	550.0	95.0	380.0	211.0
510.0	122.0	1270.0	530.0	300.0	38.0
2630.0	278.0	810.0	309.0	140.0	75.0
450.0	43.0	2280.0	407.0	250.0	142.0
540.0	89.0	720.0	159.0	90.0	35.0
480.0	103.0	180.0	78.0	3160.0	969.0
220.0	120.0	860.0	333.0	300.0	73.0
1460.0	147.0	400.0	30.0	620.0	100.0
120.0	55.0	780.0	145.0	230.0	101.0
1070.0	468.0	160.0	86.0		

No. of valid cases 29

	Variable 1	Variable 2
Mean	734.8	185.8
Minimum	90.0	30.0
Maximum	3160.0	969.0
Standard deviation	765.2	201.1
Sum of squares	16395924.1	1131860.8
Sum of weights		29.0000
Sum of products		3483048.9655
Correlation		0.8085
