NAG Toolbox for MATLAB

g02bx

1 Purpose

g02bx calculates the sample means, the standard deviations, the variance-covariance matrix, and the matrix of Pearson product-moment correlation coefficients for a set of data. Weights may be used.

2 Syntax

```
[xbar, std, v, r, ifail] = g02bx(weight, n, x, wt, 'm', m)
```

3 Description

g02bx uses a one-pass algorithm to compute the (optionally weighted) means and sums of squares and cross-products of deviations about the means. The algorithm uses a single pass updating algorithm as implemented by g02bu. The variance-covariance matrix, the standard deviations and the Pearson product-moment correlation matrix are then computed from these basic results, the latter by means of g02bw.

4 References

Chan T F, Golub G H and Leveque R J 1982 *Updating Formulae and a Pairwise Algorithm for Computing Sample Variances* Compstat, Physica-Verlag

West D H D 1979 Updating mean and variance estimates: An improved method *Comm. ACM* **22** 532–555

5 Parameters

5.1 Compulsory Input Parameters

1: **weight – string**

Indicates whether weights are to be used.

weight = 'U'

Weights are not used and unit weights are assumed.

weight = 'W' or 'V'

Weights are used and must be supplied in **wt**. The only difference between **weight** = 'W' or **weight** = 'V' is in computing the variance. If **weight** = 'W' the divisor for the variance is the sum of the weights minus one and if **weight** = 'V' the divisor is the number of observations with nonzero weights minus one. The former is useful if the weights represent the frequency of the observed values.

Constraint: **weight** = 'U', 'V' or 'W'.

2: **n – int32 scalar**

The number of data observations in the sample.

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

3: **x(ldx,m) – double array**

ldx, the first dimension of the array, must be at least **n**.

$\mathbf{x}(i,j)$ must contain the ith observation for the jth variable, for $i = 1, 2, \ldots, \mathbf{n}$ and $j = 1, 2, \ldots, \mathbf{m}$.

4: **wt**$(*)$ **– double array**

Note: the dimension of the array **wt** must be at least **n** if **weight** = 'W' or 'V', and at least 1 otherwise.

The optional weights.

If **weight** = 'W' or 'V', **wt**(i) must contain the weight for the ith observation, and the effective number of observations in the sum of weights.

If **weight** = 'U', **wt** is not referenced.

Constraint: if **weight** = 'W' or 'V', **wt**$(i) \geq 0.0$ and

$$\sum_{i=1}^{\mathbf{n}} \mathbf{wt}(i) > 1.$$

, for $i = 1, 2, \ldots, \mathbf{n}$.

5.2 Optional Input Parameters

1: **m – int32 scalar**

Default: The first dimension of the arrays **x**, **v**, **r** and the dimension of the array **xbar**. (An error is raised if these dimensions are not equal.)

the number of variables.

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

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, ldv

5.4 Output Parameters

1: **xbar**$(\mathbf{m})$ **– double array**

The sample means. **xbar**(j) contains the mean of the jth variable.

2: **std**$(\mathbf{m})$ **– double array**

The standard deviations. **std**(j) contains the standard deviation for the jth variable.

3: **v**$(\mathbf{ldv},\mathbf{m})$ **– double array**

The variance-covariance matrix. $\mathbf{v}(j,k)$ contains the covariance between variables j and k, for $j, k = 1, 2, \ldots, \mathbf{m}$.

4: **r**$(\mathbf{ldv},\mathbf{m})$ **– double array**

The matrix of Pearson product-moment correlation coefficients. $\mathbf{r}(j,k)$ contains the correlation coefficient between variables j and k.

5: **ifail – int32 scalar**

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

6 Error Indicators and Warnings

Note: g02bx may return useful information for one or more of the following detected errors or warnings.

ifail = 1

On entry, **m** < 1,
or **n** ≤ 1,
or **ldx** < **n**,
or **ldv** < **m**.

ifail = 2

On entry, **weight** ≠ 'U', 'V' or 'W'.

ifail = 3

On entry, **weight** = 'W' or 'V' and a value of **wt** < 0.0.

ifail = 4

weight = 'W' and the sum of weights is not greater than 1.0, or **weight** = 'V' and fewer than 2 observations have nonzero weights.

ifail = 5

A variable has a zero variance. In this case **v** and **std** are returned as calculated but **r** will contain zero for any correlation involving a variable with zero variance.

7 Accuracy

For a discussion of the accuracy of the one pass algorithm see Chan *et al.* 1982 and West 1979.

8 Further Comments

None.

9 Example

```
weight = 'u';
n = int32(20);
x = [11.25, 48.9, 7.43, 2.27, 15.48;
     10.87, 47.7, 7.45, 1.971, 14.97;
     11.18, 48.2, 7.44, 1.979, 14.2;
     10.62, 49, 7.38, 2.026, 15.02;
     11.02, 47.4, 7.43, 1.974, 12.92;
     10.83, 48.3, 7.72, 2.124, 13.58;
     11.18, 49.3, 7.05, 2.064, 14.12;
     11.05, 48.2, 6.95, 2.001, 15.34;
     11.15, 49.1, 7.12, 2.035, 14.52;
     11.23, 48.6, 7.28, 1.97, 15.25;
     10.94, 49.9, 7.45, 1.974, 15.34;
     11.18, 49, 7.34, 1.942, 14.48;
     11.02, 48.2, 7.29, 2.063, 12.92;
     10.99, 47.8, 7.37, 1.973, 13.61;
     11.03, 48.9, 7.45, 1.974, 14.2;
     11.09, 48.8, 7.08, 2.039, 14.51;
     11.46, 51.2, 6.75, 2.008, 16.07;
     11.57, 49.8, 7, 1.944, 16.6;
     11.07, 47.9, 7.04, 1.947, 13.41;
     10.89, 49.6, 7.07, 1.798, 15.84];
wt = [5.495816452771857e+222];
[xbar, std, v, r, ifail] = g02bx(weight, n, x, wt)
```

```
xbar =
   11.0810
   48.7900
```

```
    7.2545
    2.0038
   14.6190
std =
    0.2132
    0.9002
    0.2349
    0.0902
    1.0249
v =
    0.0455    0.0847   -0.0272    0.0013    0.0855
    0.0847    0.8104   -0.1070   -0.0055    0.6511
   -0.0272   -0.1070    0.0552    0.0059   -0.1048
    0.0013   -0.0055    0.0059    0.0081   -0.0138
    0.0855    0.6511   -0.1048   -0.0138    1.0504
r =
    1.0000    0.4416   -0.5427    0.0696    0.3912
    0.4416    1.0000   -0.5058   -0.0678    0.7057
   -0.5427   -0.5058    1.0000    0.2768   -0.4352
    0.0696   -0.0678    0.2768    1.0000   -0.1494
    0.3912    0.7057   -0.4352   -0.1494    1.0000
ifail =
           0
```