

G02BXF – NAG Fortran Library Routine Document

Note. Before using this routine, please read the Users' Note for your implementation to check the interpretation of bold italicised terms and other implementation-dependent details.

1 Purpose

G02BXF 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 Specification

```

SUBROUTINE G02BXF(WEIGHT, N, M, X, LDX, WT, XBAR, STD, V, LDV, R,
1                IFAIL)
  INTEGER        N, M, LDX, LDV, IFAIL
  real         X(LDX,M), WT(*), XBAR(M), STD(M), V(LDV,M),
1                R(LDV,M)
  CHARACTER*1    WEIGHT

```

3 Description

G02BXF 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 G02BUF. 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 G02BWF.

4 References

- [1] Chan T F, Golub G H and Leveque R J (1982) *Updating Formulae and a Pairwise Algorithm for Computing Sample Variances* Compstat, Physica-Verlag
- [2] West D H D (1979) Updating mean and variance estimates: An improved method *Comm. ACM* **22** 532–535

5 Parameters

- 1: WEIGHT — CHARACTER*1 *Input*
On entry: indicates whether weights are to be used.

If WEIGHT = 'U', weights are not used and unit weights are assumed.

If 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 non-zero 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 — INTEGER *Input*
On entry: the number of data observations in the sample.
Constraint: N > 1.

- 3:** M — INTEGER *Input*
On entry: the number of variables.
Constraint: $M \geq 1$.
- 4:** X(LDX,M) — *real* array *Input*
On entry: X(*i*,*j*) must contain the *i*th observation for the *j*th variable, for $i = 1, 2, \dots, N$; $j = 1, 2, \dots, M$.
- 5:** LDX — INTEGER *Input*
On entry: the first dimension of the array X as declared in the (sub)program from which G02BXF is called.
Constraint: $LDX \geq N$.
- 6:** WT(*) — *real* array *Input*
Note: the dimension of the array WT must be at least N if WEIGHT = 'W' or 'V' and 1 otherwise.
On entry: the optional weights.
 If WEIGHT = 'W' or 'V', then WT(*i*) must contain the weight for the *i*th observation, and the effective number of observations in the sum of weights.
 If WEIGHT = 'U', then WT is not referenced.
Constraints: if WEIGHT = 'W' or 'V', $WT(i) \geq 0.0$, for $i = 1, 2, \dots, N$,
- $$\sum_{i=1}^N WT(i) > 1.$$
- 7:** XBAR(M) — *real* array *Output*
On exit: the sample means. XBAR(*j*) contains the mean of the *j*th variable.
- 8:** STD(M) — *real* array *Output*
On exit: the standard deviations. STD(*j*) contains the standard deviation for the *j*th variable.
- 9:** V(LDV,M) — *real* array *Output*
On exit: the variances-covariance matrix. V(*j*,*k*) contains the covariance between variables *j* and *k*, for $j, k = 1, 2, \dots, M$.
- 10:** LDV — INTEGER *Input*
On entry: the first dimension of the array R as declared in the (sub)program from which G02BXF is called.
Constraint: $LDV \geq M$.
- 11:** R(LDV,M) — *real* array *Output*
On exit: the matrix of Pearson product-moment correlation coefficients. R(*j*,*k*) contains the correlation coefficient between variables *j* and *k*.
- 12:** IFAIL — INTEGER *Input/Output*
On entry: IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.
On exit: IFAIL = 0 unless the routine detects an error or gives a warning (see Section 6).

For this routine, because the values of output parameters may be useful even if IFAIL \neq 0 on exit, users are recommended to set IFAIL to -1 before entry. **It is then essential to test the value of IFAIL on exit.**

6 Error Indicators and Warnings

If on entry `IFAIL = 0` or `-1`, explanatory error messages are output on the current error message unit (as defined by `X04AAF`).

Errors or warnings specified by the routine:

`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 less than 2 observations have non-zero 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.* [1] and West [2].

8 Further Comments

None.

9 Example

The data are some of the results from 1988 Olympic Decathlon. They are the times (in seconds) for the 100m and 400m races and the distances (in metres) for the long jump, high jump and shot. Twenty observations are input and the correlation matrix is computed and printed.

9.1 Program Text

Note. The listing of the example program presented below uses bold italicised terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*      G02BXF Example Program Text
*      Mark 17 Revised.  NAG Copyright 1995.
*      .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          MMAX, NMAX
      PARAMETER        (NMAX=20,MMAX=5)
*      .. Local Scalars ..
      INTEGER          I, IFAIL, J, LDX, LDV, M, N
```

```

CHARACTER          WEIGHT
*   .. Local Arrays ..
  real              R(MMAX,MMAX), STD(MMAX), V(MMAX,MMAX), WT(NMAX),
+                  X(NMAX,MMAX), XBAR(MMAX)
*   .. External Subroutines ..
  EXTERNAL          GO2BXF, X04CAF
*   .. Executable Statements ..
  WRITE (NOUT,*) 'G02BXF Example Program Results'
*   Skip heading in data file
  READ (NIN,*)
  READ (NIN,*) WEIGHT, N, M
  IF (M.LE.MMAX .AND. N.LE.NMAX) THEN
    IF (WEIGHT.EQ.'W' .OR. WEIGHT.EQ.'w') THEN
      DO 20 I = 1, N
        READ (NIN,*) (X(I,J),J=1,M), WT(I)
20      CONTINUE
      ELSE
        DO 40 I = 1, N
          READ (NIN,*) (X(I,J),J=1,M)
40      CONTINUE
      END IF
      LDX = NMAX
      LDV = MMAX
      IFAIL = -1
*
*   CALL GO2BXF(WEIGHT,N,M,X,LDX,WT,XBAR,STD,V,LDV,R,IFAIL)
*
  IF (IFAIL.EQ.0 .OR. IFAIL.EQ.5) THEN
    WRITE (NOUT,*)
    WRITE (NOUT,*) '      Means'
    WRITE (NOUT,*)
    WRITE (NOUT,99999) (XBAR(I),I=1,M)
    WRITE (NOUT,*)
    WRITE (NOUT,*) '      Standard deviations'
    WRITE (NOUT,*)
    WRITE (NOUT,99999) (STD(I),I=1,M)
*
*   Print the correlation matrix
*
  IF (IFAIL.EQ.5) THEN
    WRITE (NOUT,*)
    WRITE (NOUT,*) ' NOTE: some variances are zero'
  END IF
  WRITE (NOUT,*)
  CALL X04CAF('Upper', 'Non-unit', M,M,R,LDV,
+           '      Correlation matrix', IFAIL)
  WRITE (NOUT,*)
  END IF
  ELSE
    WRITE (NOUT,99998) 'M or N is too large. M =', M, ', N =', N
  END IF
  STOP
*
99999 FORMAT (1X,10F13.4)
99998 FORMAT (1X,A,I6,A,I6)
END

```

9.2 Program Data

G02BXF Example Program Data

'u'	20	5			
11.25	48.9	7.43	2.270	15.48	
10.87	47.7	7.45	1.971	14.97	
11.18	48.2	7.44	1.979	14.20	
10.62	49.0	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	1.970	15.25	
10.94	49.9	7.45	1.974	15.34	
11.18	49.0	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	14.20	
11.09	48.8	7.08	2.039	14.51	
11.46	51.2	6.75	2.008	16.07	
11.57	49.8	7.00	1.944	16.60	
11.07	47.9	7.04	1.947	13.41	
10.89	49.6	7.07	1.798	15.84	

9.3 Program Results

G02BXF Example Program Results

Means

11.0810	48.7900	7.2545	2.0038	14.6190
---------	---------	--------	--------	---------

Standard deviations

0.2132	0.9002	0.2349	0.0902	1.0249
--------	--------	--------	--------	--------

Correlation matrix

	1	2	3	4	5
1	1.0000	0.4416	-0.5427	0.0696	0.3912
2		1.0000	-0.5058	-0.0678	0.7057
3			1.0000	0.2768	-0.4352
4				1.0000	-0.1494
5					1.0000