G02BQF – 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

G02BQF computes Kendall and/or Spearman non-parametric rank correlation coefficients for a set of data; the data array is preserved, and the ranks of the observations are not available on exit from the routine.

2 Specification

```
 SUBROUTINE G02BQF(N, M, X, IX, ITYPE, RR, IRR, KWORKA, KWORKB,
1                  WORK1, WORK2, IFAIL)
 INTEGER           N, M, IX, ITYPE, IRR, KWORKA(N), KWORKB(N), IFAIL
 real              X(IX,M), RR(IRR,M), WORK1(N), WORK2(N)
```

3 Description

The input data consists of n observations for each of m variables, given as an array

$$[x_{ij}], \quad i = 1, 2, \ldots, n \ \ (n \geq 2)$$

$$j = 1, 2, \ldots, m \ \ (m \geq 2),$$

where x_{ij} is the ith observation on the jth variable.

The observations are first ranked, as follows:

For a given variable, j say, each of the n observations, $x_{1j}, x_{2j}, \ldots, x_{nj}$, has associated with it an additional number, the 'rank' of the observation, which indicates the magnitude of that observation relative to the magnitude of the other $n - 1$ observations on that same variable.

The smallest observation for variable j is assigned the rank 1, the second smallest observation for variable j the rank 2, the third smallest the rank 3, and so on until the largest observation for variable j is given the rank n.

If a number of cases all have the same value for the given variable, j, then they are each given an 'average' rank – e.g., if in attempting to assign the rank $h + 1$, k observations were found to have the same value, then instead of giving them the ranks

$$h + 1, h + 2, \ldots, h + k,$$

all k observations would be assigned the rank

$$\frac{2h + k + 1}{2}$$

and the next value in ascending order would be assigned the rank

$$h + k + 1.$$

The process is repeated for each of the m variables.

Let y_{ij} be the rank assigned to the observation x_{ij} when the jth variable is being ranked.

The quantities calculated are:

(a) Kendall's tau rank correlation coefficients:

$$R_{jk} = \frac{\sum_{h=1}^{n} \sum_{i=1}^{n} \text{sign}(y_{hj} - y_{ij}) \, \text{sign}(y_{hk} - y_{ik})}{\sqrt{[n(n-1) - T_j][n(n-1) - T_k]}}, \qquad j, k = 1, 2, \ldots, m;$$

and $\operatorname{sign} u = 1$ if $u > 0$
$\operatorname{sign} u = 0$ if $u = 0$
$\operatorname{sign} u = -1$ if $u < 0$

and $T_j = \sum t_j(t_j - 1)$, t_j being the number of ties of a particular value of variable j, and the summation being over all tied values of variable j.

(b) Spearman's rank correlation coefficients:

$$R^*_{jk} = \frac{n(n^2-1) - 6\sum_{i=1}^{n}(y_{ij} - y_{ik})^2 - \frac{1}{2}(T^*_j + T^*_k)}{\sqrt{[n(n^2-1) - T^*_j][n(n^2-1) - T^*_k]}}, \quad j,k = 1,2,\ldots,m;$$

where $T^*_j = \sum t_j(t_j^2 - 1)$ where t_j is the number of ties of a particular value of variable j, and the summation is over all tied values of variable j.

4 References

[1] Siegel S (1956) *Non-parametric Statistics for the Behavioral Sciences* McGraw–Hill

5 Parameters

1: N — INTEGER *Input*

On entry: the number n, of observations or cases.

Constraint: N ≥ 2.

2: M — INTEGER *Input*

On entry: the number m, of variables.

Constraint: M ≥ 2.

3: X(IX,M) — ***real*** array *Input*

On entry: X(i,j) must be set to data value x_{ij}, the value of the ith observation on the jth variable, for $i = 1,2,\ldots,n$; $j = 1,2,\ldots,m$.

4: IX — INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G02BQF is called.

Constraint: IX $\geq$ N.

5: ITYPE — INTEGER *Input*

On entry: the type of correlation coefficients which are to be calculated. If ITYPE $= -1$, only Kendall's tau coefficients are calculated; if ITYPE $= 0$, both Kendall's tau and Spearman's coefficients are calculated; if ITYPE $= 1$, only Spearman's coefficients are calculated.

6: RR(IRR,M) — ***real*** array *Output*

On exit: the requested correlation coefficients. If only Kendall's tau coefficients are requested (ITYPE $= -1$), then RR(j,k) contains Kendall's tau for the jth and kth variables; if only Spearman's coefficients are requested (ITYPE $= 1$), then RR(j,k) contains Spearman's rank correlation coefficient for the jth and kth variables. If both Kendall's tau and Spearman's coefficients are requested (ITYPE $= 0$), then the upper triangle of RR contains the Spearman coefficients and the lower triangle the Kendall coefficients. That is, for the jth and kth variables, where j is less than k, RR(j,k) contains the Spearman rank correlation coefficient, and RR(k,j) contains Kendall's tau, for $j,k = 1,2,\ldots,m$.

(Diagonal terms, RR(j,j), are unity for all three values of ITYPE).

7: IRR — INTEGER *Input*

On entry: the first dimension of the array RR as declared in the (sub)program from which G02BQF is called.

Constraint: IRR $\geq$ M.

8: KWORKA(N) — INTEGER array *Workspace*

9: KWORKB(N) — INTEGER array *Workspace*

10: WORK1(N) — ***real*** array *Workspace*

11: WORK2(N) — ***real*** array *Workspace*

12: IFAIL — INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. For users not familiar with this parameter (described in Chapter P01) the recommended value is 0.

On exit: IFAIL = 0 unless the routine detects an error (see Section 6).

6 Error Indicators and Warnings

Errors detected by the routine:

IFAIL = 1

On entry, N < 2.

IFAIL = 2

On entry, M < 2.

IFAIL = 3

On entry, IX $<$ N,
or IRR $<$ M.

IFAIL = 4

On entry, ITYPE < -1,
or ITYPE > 1.

7 Accuracy

The method used is believed to be stable.

8 Further Comments

The time taken by the routine depends on n and m.

9 Example

The example program reads in a set of data consisting of nine observations on each of three variables. The program then calculates and prints both Kendall's tau and Spearman's rank correlation coefficients for all three variables.

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.

```
*     G02BQF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          M, N, IA, ICORR
      PARAMETER        (M=3,N=9,IA=N,ICORR=M)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, ITYPE, J
*     .. Local Arrays ..
      real             A(IA,M), CORR(ICORR,M), WA(N), WB(N)
      INTEGER          IW(N), JW(N)
*     .. External Subroutines ..
      EXTERNAL         G02BQF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G02BQF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) ((A(I,J),J=1,M),I=1,N)
      WRITE (NOUT,*)
      WRITE (NOUT,99999) 'Number of variables (columns) =', M
      WRITE (NOUT,99999) 'Number of cases     (rows)    =', N
      WRITE (NOUT,*)
      WRITE (NOUT,*) 'Data matrix is:-'
      WRITE (NOUT,*)
      WRITE (NOUT,99998) (J,J=1,M)
      WRITE (NOUT,99997) (I,(A(I,J),J=1,M),I=1,N)
      WRITE (NOUT,*)
      IFAIL = 1
      ITYPE = 0
*
      CALL G02BQF(N,M,A,IA,ITYPE,CORR,ICORR,IW,JW,WA,WB,IFAIL)
*
      IF (IFAIL.NE.0) THEN
         WRITE (NOUT,99999) 'Routine fails, IFAIL =', IFAIL
      ELSE
         WRITE (NOUT,*) 'Matrix of rank correlation coefficients:'
         WRITE (NOUT,*) 'Upper triangle -- Spearman''s'
         WRITE (NOUT,*) 'Lower triangle -- Kendall''s tau'
         WRITE (NOUT,*)
         WRITE (NOUT,99998) (I,I=1,M)
         WRITE (NOUT,99997) (I,(CORR(I,J),J=1,M),I=1,M)
      END IF
      STOP
*
99999 FORMAT (1X,A,I3)
99998 FORMAT (1X,3I12)
99997 FORMAT (1X,I3,3F12.4)
      END
```

9.2 Program Data

```
G02BQF Example Program Data
1.70      1.00      0.50
2.80      4.00      3.00
0.60      6.00      2.50
1.80      9.00      6.00
0.99      4.00      2.50
1.40      2.00      5.50
1.80      9.00      7.50
2.50      7.00      0.00
0.99      5.00      3.00
```

9.3 Program Results

```
G02BQF Example Program Results

Number of variables (columns) =  3
Number of cases     (rows)    =  9

Data matrix is:-

          1           2           3
  1    1.7000      1.0000      0.5000
  2    2.8000      4.0000      3.0000
  3    0.6000      6.0000      2.5000
  4    1.8000      9.0000      6.0000
  5    0.9900      4.0000      2.5000
  6    1.4000      2.0000      5.5000
  7    1.8000      9.0000      7.5000
  8    2.5000      7.0000      0.0000
  9    0.9900      5.0000      3.0000

Matrix of rank correlation coefficients:
Upper triangle -- Spearman's
Lower triangle -- Kendall's tau

          1           2           3
  1    1.0000      0.2246      0.1186
  2    0.0294      1.0000      0.3814
  3    0.1176      0.2353      1.0000
```