NAG Toolbox

nag_correg_coeffs_kspearman_overwrite (g02bn)

1 Purpose

nag_correg_coeffs_kspearman_overwrite (g02bn) computes Kendall and/or Spearman nonparametric rank correlation coefficients for a set of data; the data array is overwritten with the ranks of the observations.

2 Syntax

```
[x, rr, ifail] = nag_correg_coeffs_kspearman_overwrite(x, itype, 'n', n, 'm', m)
[x, rr, ifail] = g02bn(x, itype, 'n', n, 'm', m)
```

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

At Mark 22: **n** was made optional.

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 of the jth variable.

The quantities calculated are:

(a) Ranks

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 magnitudes 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 actual observations x_{ij} are replaced by the ranks y_{ij}.

(b) Nonparametric rank correlation coefficients

(i) Kendall's tau:

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

where $\operatorname{sign} u = 1$ if $u > 0$,

$\operatorname{sign} u = 0$ if $u = 0$,

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

and $T_j = \sum t_j(t_j - 1)$, 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.

(ii) Spearman's:

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

where $T^*_j = \sum t_j\left(t^2_j - 1\right)$, 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.

4 References

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

5 Parameters

5.1 Compulsory Input Parameters

1: **x**(ldx, **m**) – REAL (KIND=nag_wp) array

ldx, the first dimension of the array, must satisfy the constraint $ldx \geq$ **n**.

x(i,j) must be set to x_{ij}, the value of the ith observation on the jth variable, for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$.

2: **itype** – INTEGER

The type of correlation coefficients which are to be calculated.

itype $= -1$
Only Kendall's tau coefficients are calculated.

itype $= 0$
Both Kendall's tau and Spearman's coefficients are calculated.

itype $= 1$
Only Spearman's coefficients are calculated.

Constraint: **itype** $= -1$, 0 or 1.

5.2 Optional Input Parameters

1: **n** – INTEGER

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

n, the number of observations or cases.

Constraint: **n** ≥ 2.

2: **m** – INTEGER

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

m, the number of variables.

Constraint: **m** ≥ 2.

5.3 Output Parameters

1: **x**$(ldx,$ **m**$)$ – REAL (KIND=nag_wp) array

x(i,j) contains the rank y_{ij} of the observation x_{ij}, for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$.

2: **rr**$(ldrr,$ **m**$)$ – REAL (KIND=nag_wp) array

The requested correlation coefficients.

If only Kendall's tau coefficients are requested (**itype** $= -1$), **rr**(j,k) contains Kendall's tau for the jth and kth variables.

If only Spearman's coefficients are requested (**itype** $= 1$), **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** $= 0$), 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 $j = 1,2,\ldots,m$ and $k = 1,2,\ldots,m$.

(Diagonal terms, **rr**(j,j), are unity for all three values of **itype**.)

3: **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** < 2.

ifail $= 2$

On entry, **m** < 2.

ifail $= 3$

On entry, $ldx <$ **n**,
or $ldrr <$ **m**.

ifail $= 4$

On entry, **itype** < -1,
or **itype** > 1.

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_correg_coeffs_kspearman_overwrite (g02bn) depends on n and m.

9 Example

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

9.1 Program Text

```
function g02bn_example

fprintf('g02bn example results\n\n');

x = [1.7,  1, 0.5;
     2.8,  4, 3.0;
     0.6,  6, 2.5;
     1.8,  9, 6.0;
     0.99, 4, 2.5;
     1.4,  2, 5.5;
     1.8,  9, 7.5;
     2.5,  7, 0.0;
     0.99, 5, 3.0];
[n,m] = size(x);
fprintf('Number of variables (columns) = %d\n', m);
fprintf('Number of cases     (rows)    = %d\n\n', n);
disp('Data matrix is:-');
disp(x);

itype = nag_int(0);
[x, rr, ifail] = g02bn(x, itype);

fprintf('\nMatrix of ranks:-\n');
disp(x);
fprintf('Matrix of rank correlation coefficients:\n');
fprintf('Upper triangle -- Spearman''s\n');
fprintf('Lower triangle -- Kendall''s tau\n\n');
disp(rr);
```

9.2 Program Results

```
g02bn example results

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

Data matrix is:-
    1.7000    1.0000    0.5000
    2.8000    4.0000    3.0000
    0.6000    6.0000    2.5000
    1.8000    9.0000    6.0000
    0.9900    4.0000    2.5000
```

```
    1.4000    2.0000    5.5000
    1.8000    9.0000    7.5000
    2.5000    7.0000         0
    0.9900    5.0000    3.0000

Matrix of ranks:-
    5.0000    1.0000    2.0000
    9.0000    3.5000    5.5000
    1.0000    6.0000    3.5000
    6.5000    8.5000    8.0000
    2.5000    3.5000    3.5000
    4.0000    2.0000    7.0000
    6.5000    8.5000    9.0000
    8.0000    7.0000    1.0000
    2.5000    5.0000    5.5000

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

    1.0000    0.2246    0.1186
    0.0294    1.0000    0.3814
    0.1176    0.2353    1.0000
```