nag_ken_spe_corr_coeff (g02brc)

1. Purpose

nag_ken_spe_corr_coeff (g02brc) calculates Kendall and Spearman rank correlation coefficients.

2. Specification

```
#include <nag.h>
#include <nagg02.h>

void nag_ken_spe_corr_coeff(Integer n, Integer m, double x[], Integer tdx,
   Integer svar[], Integer sobs[], double corr[], Integer tdc, NagError *fail)
```

3. Description

This function calculates both the Kendall rank correlation coefficients and the Spearman rank correlation coefficients.

The data consists of n observations for each of m variables:

$$x_{ij},\ i = 1,2,\ldots,n;\ j = 1,2,\ldots,m \quad (m,n \geq 2)$$

where x_{ij} is the ith observation on the jth variable. The function eliminates any variable x_{ij} for $i = 1,2,\ldots,n$ where the argument **svar**$[j-1] = 0$, and any observation x_{ij} for $j = 1,2,\ldots,m$ where the argument **sobs**$[i-1] = 0$.

The observations are first ranked as follows:

For a given variable, j say, each of the observations x_{ij} for which **sobs**$[i-1] > 0$, for $i = 1,2,\ldots,n$ 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 observations on that same variable for which **sobs**$[i-1] > 0$.

The smallest of these valid observations for variable j is assigned the rank 1, the second smallest observation for variable j the rank 2, and so on until the largest such observation is given the rank n_s, where n_s is the number of observations for which **sobs**$[i-1] > 0$.

If a number of cases all have the same value for a given variable, j, then they are each given an 'average' rank — e.g. if in attempting to assign the rank $h+1$, k observations for which **sobs**$[i-1] > 0$ 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 for which **svar**$[j-1] > 0$.

Let y_{ij} be the rank assigned to the observation x_{ij} when the jth variable is being ranked. For those observations, i, for which **sobs**$[i-1] = 0$, $y_{ij} = 0$, for $j = 1,2,\ldots,m$.

For variables j,k the following are computed:

(a) Kendall's tau correlation coefficients:

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

where n_s is the number of observations for which **sobs**$[i-1] > 0$,

and sign $u = 1$ if $u > 0$,
sign $u = 0$ if $u = 0$,
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.

(b) Spearman's rank correlation coefficients:

$$R_{jk} = \frac{n_s(n_s^2-1) - 6\sum_{i=1}^{n}(y_{ij}-y_{ik})^2 - \frac{1}{2}(T_j+T_k)}{\sqrt{[n_s(n_s^2-1)-T_j][n_s(n_s^2-1)-T_k]}} \quad j,k = 1,2,\ldots,m;$$

where n_s is the number of observations for which **sobs**$[i-1] > 0$,

and $T_j = \sum t_j(t_j^2-1)$ 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. Parameters

n

Input: the number of observations in the data set.
Constraint: $\mathbf{n} \geq \mathbf{2}$.

m

Input: the number of variables.
Constraint: $\mathbf{m} \geq \mathbf{2}$.

x[n][tdx]

Input: $\mathbf{x}[i-1][j-1]$ must contain the ith observation on the jth variable, for $i = 1,2,\ldots,n$; $j = 1,2,\ldots,m$.

tdx

Input: the second dimension of the array **x** as declared in the function from which nag_ken_spe_corr_coeff is called.
Constraint: $\mathbf{tdx} \geq \mathbf{m}$.

svar[m]

Input: **svar**$[j-1]$ indicates which variables are to be included, for the jth variable to be included, **svar**$[j-1] > 0$. If all variables are to be included then a NULL pointer (Integer*)0 may be supplied.
Constraint: **svar**$[j-1] \geq 0$, for $j = 1,2,\ldots,m$ and there is at least 1 positive element.

sobs[n]

Input: **sobs**$[i-1]$ indicates which observations are to be included, for the ith observation to be included, **sobs**$[i-1] > 0$. If all observations are to be included then a NULL pointer (Integer*)0 may be supplied.
Constraint: **sobs**$[i-1] \geq 0$, for $i = 1,2,\ldots,n$ and there are at least 2 positive elements.

corr[m][tdc]

Output: the upper n_s by n_s part of **corr** contains the correlation coefficients, the upper triangle 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, **corr**$[j-1][k-1]$ contains the Spearman rank correlation coefficient, and **corr**$[k-1][j-1]$ contains Kendall's tau, for $j,k = 1,2,\ldots,n_s$. The diagonal will be set to 1.

tdc

Input: the second dimension of the array **corr** as declared in the function from which nag_ken_spe_corr_coeff is called.
Constraint: $\mathbf{tdc} \geq \mathbf{m}$.

fail

The NAG error parameter, see the Essential Introduction to the NAG C Library.

5. Error Indications and Warnings

NE_INT_ARG_LT

On entry, **n** must not be less than 2: $\mathbf{n} = \langle value \rangle$.
On entry, **m** must not be less than 2: $\mathbf{m} = \langle value \rangle$.

NE_2_INT_ARG_LT

On entry **tdx** = ⟨*value*⟩ while **m** = ⟨*value*⟩. These parameters must satisfy **tdx** ≥ **m**.
On entry **tdc** = ⟨*value*⟩ while **m** = ⟨*value*⟩. These parameters must satisfy **tdc** ≥ **m**.

NE_INT_ARRAY_1

Value ⟨*value*⟩ given to **svar**[⟨*value*⟩] not valid. Correct range for elements of **svar** is **svar**[i] ≥ 0.
Value ⟨*value*⟩ given to **sobs**[⟨*value*⟩] not valid. Correct range for elements of **sobs** is **sobs**[i] ≥ 0.

NE_SOBS_LOW

Too few observations have been selected. On entry, **sobs** must contain at least 2 positive elements.

NE_SVAR_LOW

No variables have been selected. On entry, **svar** must contain at least 1 positive element.

NE_ALLOC_FAIL

Memory allocation failed.

NE_INTERNAL_ERROR

An initial error has occured in this function. Check the function call and any array sizes.

6. Further Comments

None.

6.1. Accuracy

The computations are believed to be stable.

6.2. References

Siegel S (1956) *Non-Parametric Statistics for the Behavioral Sciences.* McGraw-Hill.

7. See Also

nag_corr_cov (g02bxc)

8. Example

A program to calculate the Kendall and Spearman rank correlation coefficients from a set of data.

8.1. Program Text

```
/* nag_ken_spe_corr_coeff(g02brc) Example Program
 *
 * Copyright 1994 Numerical Algorithms Group.
 *
 * Mark 3, 1994.
 */

#include <nag.h>
#include <stdio.h>
#include <nag_stdlib.h>
#include <nagg02.h>

#define NMAX 20
#define MMAX 20

main()
{
  Integer i, j, m, n, tdx, tdc;
  double x[NMAX][MMAX], corr[MMAX][MMAX];
  Integer svar[MMAX], sobs[NMAX];
  Integer *svarptr, *sobsptr;
  char s, w;
```

```
  tdx = MMAX;
  tdc = MMAX;

  Vprintf("g02brc Example Program Results\n");

  /* Skip heading in data file */
  Vscanf(" %*[^\n]");

  /* Read data */
  Vscanf("%ld%ld\n",&m,&n);
  if (m>=1 && m<=MMAX && n>=1 && n<=NMAX)
    {
      for (i=0; i<n; i++)
        for (j=0; j<m; j++)
          Vscanf("%lf",&x[i][j]);

      /* Read flag specifying if svar is to be supplied */
      Vscanf(" %c",&s);
      if (s == 'S' || s == 's')
        {
          /* Assign pointer to svar and read in values for svar */
          svarptr = svar;
          for (i=0; i<m; i++)
            Vscanf("%ld",&svar[i]);
        }
      else
        {
          /* Assign pointer to NULL and discard rest of line */
          svarptr = (Integer *)0;
          Vscanf(" %*[^\n]"); /* skip rest of line */
        }

      /* Read flag specifying if sobs is to be supllied */
      Vscanf(" %c",&w);
      if (w == 'W' || w == 'w')
        {
          /* Assign pointer to sobs and read in values for sobs */
          sobsptr = sobs;
          for (i=0; i<n; i++)
            Vscanf("%ld",&sobs[i]);
        }
      else
        {
          /* Assign pointer to NULL and discard rest of line */
          sobsptr = (Integer *)0;
          Vscanf(" %*[^\n]"); /* skip rest of line */
        }

      /* Calculate the Kendall and Spearman coefficients */
      g02brc(n, m, (double *)x, tdx, svarptr, sobsptr, (double *)corr, tdc,
             NAGERR_DEFAULT);

      Vprintf("\nCorrelation coefficients:\n\n");
      for (i=0; i<m; i++)
        {
          for (j=0; j<m; j++)
            Vprintf("%8.5f ", corr[i][j]);
          Vprintf("\n");
        }
    }
  else
    {
      Vfprintf(stderr, "One or both of m and n are out of range:\
 m = %-3ld while  n = %-3ld\n", m, n);
      exit(EXIT_FAILURE);
    }
  exit(EXIT_SUCCESS);
}
```

8.2. Program Data

```
g02brc Example Program Data
3 7
1.0 2.0 4.0
7.0 7.0 3.0
2.0 3.0 4.0
4.0 4.0 5.0
5.0 6.0 7.0
3.0 1.0 3.0
6.0 5.0 5.0
s 1 1 1
w 1 1 1 1 1 1 1
```

8.3. Program Results

```
g02brc Example Program Results

Correlation coefficients:

 1.00000  0.85714  0.12849
 0.71429  1.00000  0.33040
 0.10287  0.41148  1.00000
```