nag_shapiro_wilk_test (g01ddc)

1. Purpose

nag_shapiro_wilk_test (g01ddc) calculates Shapiro and Wilk's W statistic and its significance level for testing Normality.

2. Specification

```
#include <nag.h>
#include <nagg01.h>

void nag_shapiro_wilk_test(Integer n, double x[], Boolean calc_wts,
        double a[], double *w, double *pw, NagError *fail)
```

3. Description

This routine calculates Shapiro and Wilk's W statistic and its significance level for any sample size between 3 and 2000. It is an adaptation of the Applied Statistics Algorithm AS 181, see Royston (1982a). The full description of the theory behind this algorithm is given in Royston (1982b).

Given a set of observations $x_1, x_2, \ldots, x_n$ sorted into either ascending or descending order (nag_double_sort (m01cac) may be used to sort the data), nag_shapiro_wilk_test calculates the value of Shapiro and Wilk's W statistic defined as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_i\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

where $\bar{x} = \frac{1}{n}\sum_{1}^{n} x_i$ is the sample mean and a_i, for $i = 1, 2, \ldots, n$ are a set of 'weights' whose values depend only on the sample size n.

On exit, the values of a_i, for $i = 1, 2, \ldots, n$ are only of interest should the user wish to call the routine again to calculate W and its significance level for a different sample of the same size.

4. Parameters

n

Input: the sample size, n.
Constraint: $3 \leq \mathbf{n} \leq 2000$.

x[n]

Input: the ordered sample values, x_i; for $i = 1, 2, \ldots, n$.

calc_wts

Input: **calc_wts** must be set to **TRUE** if the user wishes nag_shapiro_wilk_test to calculate the elements of **a**.

calc_wts should be set to **FALSE** if the user has saved the values in **a** from a previous call to nag_shapiro_wilk_test.

If in doubt, set **calc_wts** equal to **TRUE**.

a[n]

Input: if **calc_wts** has been set to **FALSE** then before entry **a** must contain the n weights as calculated in a previous call to nag_shapiro_wilk_test, otherwise **a** need not be set.
Output: the n weights required to calculate W.

w

Output: the value of the statistic, W.

pw
Output: the significance level of W.

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 3: **n** = ⟨*value*⟩.

NE_INT_ARG_GT
On entry, **n** must not be greater than 2000: **n** = ⟨*value*⟩.

NE_NON_MONOTONIC
On entry, the sequence in array **x** is non-monotonic. First anomaly detected at **x**[⟨*value*⟩] = ⟨*value*⟩.

NE_ALL_ELEMENTS_EQUAL
On entry, all the values in the array **x** must not be equal.

NE_INTERNAL_ERROR
An internal error has occurred in this function. Check the function call and any array sizes. If the call is correct then please consult NAG for assistance.

6. Further Comments

The time taken by the routine depends roughly linearly on the value of n.

For very small samples the power of the test may not be very high.

The contents of the array A should not be modified between calls to nag_shapiro_wilk_test for a given sample size, unless **calc_wts** is reset to **TRUE** before each call of nag_shapiro_wilk_test.

The Shapiro and Wilk W test is very sensitive to ties. If the data has been rounded the test can be improved by using Sheppard's correction to adjust the sum of squares about the mean. This produces an adjusted value of W,

$$WA = W\frac{\sum(x_{(i)}-\bar{x})^2}{\left\{\sum_{i=1}^{n}(x_{(i)}-\bar{x})^2-\frac{n-1}{12}\omega^2\right\}}$$

where ω is the rounding width. WA can be compared with a standard normal distribution, but a further approximation is given by Royston (1986).

6.1. Accuracy

There may be a loss of significant figures for large n.

6.2. References

Royston J P (1982a) Algorithm AS181: The W Test for Normality *Appl. Statist.* **31** 176–180.
Royston J P (1982b) An extension of Shapiro and Wilk's W Test for Normality to large samples *Appl. Statist.* **31** 115–124.
Royston J P (1986) A Remark on AS181: The W Test for Normality *Appl. Statist.* **35** 232–234.

7. See Also

nag_ranks_and_scores (g01dhc)

8. Example

A program to test the following 2 samples (each of size 20) for Normality.

Sample Number	Data
(1)	0.11, 7.87, 4.61, 10.14, 7.95, 3.14, 0.46, 4.43, 0.21, 4.75, 0.71, 1.52, 3.24, 0.93, 0.42, 4.97, 9.53, 4.55, 0.47, 6.66
(2)	1.36, 1.14, 2.92, 2.55, 1.46, 1.06, 5.27, -1.11, 3.48, 1.10, 0.88, -0.51, 1.46, 0.52, 6.20, 1.69, 0.08, 3.67, 2.81, 3.49

The elements of **a** are calculated only in the first call of nag_shapiro_wilk_test and are re-used in the second call.

8.1. Program Text

```
/* nag_shapiro_wilk_test(g01ddc) Example Program.
 *
 * Copyright 1996 Numerical Algorithms Group.
 *
 * Mark 4, 1996.
 *
 */

#include <nag.h>
#include <nag_stdlib.h>
#include <stdio.h>
#include <nagg01.h>
#include <nagm01.h>

main()
{

#define NMAX 20

  /* Local variables */
  double a[NMAX];
  Integer i, j, n;
  double w, x[NMAX], pw;
  Boolean calwts;

  Vprintf("g01ddc Example Program Results\n");

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

  calwts = TRUE;
  Vscanf("%ld ", &n);
  if (n > 0 && n <= NMAX)
    {
      for (j = 1; j <= 2; ++j)
        {
          for (i = 1; i <= n; ++i)
            Vscanf("%lf ", &x[i - 1]);

          m01cac(x, (size_t)n, Nag_Ascending,NAGERR_DEFAULT);
          g01ddc(n, x, calwts, a, &w, &pw, NAGERR_DEFAULT);

          Vprintf("\n For sample number %2ld, value of W statistic = %7.4f\n",
                  j,w);
          Vprintf("                        Significance level is %8.4f\n", pw);
          calwts = FALSE;
        }
    }
  exit(EXIT_SUCCESS);
}
```

8.2. Program Data

```
g01ddc Example Program Data
20
0.11  7.87  4.61 10.14  7.95  3.14  0.46  4.43  0.21  4.75
0.71  1.52  3.24  0.93  0.42  4.97  9.53  4.55  0.47  6.66
1.36  1.14  2.92  2.55  1.46  1.06  5.27 -1.11  3.48  1.10
0.88 -0.51  1.46  0.52  6.20  1.69  0.08  3.67  2.81  3.49
```

8.3. Program Results

```
g01ddc Example Program Results

 For sample number  1, value of W statistic =  0.8992
                      Significance level is   0.0408

 For sample number  2, value of W statistic =  0.9583
                      Significance level is   0.5171
```