NAG Toolbox

nag_nonpar_test_kruskal (g08af)

1 Purpose

nag_nonpar_test_kruskal (g08af) performs the Kruskal–Wallis one-way analysis of variance by ranks on k independent samples of possibly unequal sizes.

2 Syntax

```
[h, p, ifail] = nag_nonpar_test_kruskal(x, l, 'lx', lx, 'k', k)
[h, p, ifail] = g08af(x, l, 'lx', lx, 'k', k)
```

3 Description

The Kruskal–Wallis test investigates the differences between scores from k independent samples of unequal sizes, the ith sample containing l_i observations. The hypothesis under test, H_0, often called the null hypothesis, is that the samples come from the same population, and this is to be tested against the alternative hypothesis H_1 that they come from different populations.

The test proceeds as follows:

(a) The pooled sample of all the observations is ranked. Average ranks are assigned to tied scores.

(b) The ranks of the observations in each sample are summed, to give the rank sums R_i, for $i = 1, 2, \ldots, k$.

(c) The Kruskal–Wallis' test statistic H is computed as:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{l_i} - 3(N+1), \qquad \text{where } N = \sum_{i=1}^{k} l_i,$$

i.e., N is the total number of observations. If there are tied scores, H is corrected by dividing by:

$$1 - \frac{\sum (t^3 - t)}{N^3 - N}$$

where t is the number of tied scores in a sample and the summation is over all tied samples.

nag_nonpar_test_kruskal (g08af) returns the value of H, and also an approximation, p, to the probability of a value of at least H being observed, H_0 is true. (H approximately follows a χ^2_{k-1} distribution). H_0 is rejected by a test of chosen size α if $p < \alpha$. The approximation p is acceptable unless $k = 3$ and l_1, l_2 or $l_3 \leq 5$ in which case tables should be consulted (e.g., O of Siegel (1956)) or $k = 2$ (in which case the Median test (see nag_nonpar_test_median (g08ac)) or the Mann–Whitney U test (see nag_nonpar_test_mwu (g08ah)) is more appropriate).

4 References

Moore P G, Shirley E A and Edwards D E (1972) *Standard Statistical Calculations* Pitman

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

5 Parameters

5.1 Compulsory Input Parameters

1: **x**(**lx**) – REAL (KIND=nag_wp) array

The elements of **x** must contain the observations in the **k** samples. The first l_1 elements must contain the scores in the first sample, the next l_2 those in the second sample, and so on.

2: **l**(**k**) – INTEGER array

l(i) must contain the number of observations l_i in sample i, for $i = 1, 2, \ldots, k$.

Constraint: **l**$(i) > 0$, for $i = 1, 2, \ldots, k$.

5.2 Optional Input Parameters

1: **lx** – INTEGER

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

N, the total number of observations.

Constraint: $\mathbf{lx} = \sum_{i=1}^{k} \mathbf{l}(i)$.

2: **k** – INTEGER

Default: the dimension of the array **l**.

k, the number of samples.

Constraint: $\mathbf{k} \geq 2$.

5.3 Output Parameters

1: **h** – REAL (KIND=nag_wp)

The value of the Kruskal–Wallis test statistic, H.

2: **p** – REAL (KIND=nag_wp)

The approximate significance, p, of the Kruskal–Wallis test statistic.

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, $\mathbf{k} < 2$.

ifail $= 2$

On entry, **l**$(i) \leq 0$ for some i, $i = 1, 2, \ldots, k$.

ifail $= 3$

On entry, $\mathbf{lx} \neq \sum_{i=1}^{k} \mathbf{l}(i)$.

ifail $= 4$

On entry, all the observations were equal.

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

For estimates of the accuracy of the significance p, see nag_stat_prob_chisq (g01ec). The χ^2 approximation is acceptable unless $k = 3$ and l_1, l_2 or $l_3 \leq 5$.

8 Further Comments

The time taken by nag_nonpar_test_kruskal (g08af) is small, and increases with N and k.

If $k = 2$, the Median test (see nag_nonpar_test_median (g08ac)) or the Mann–Whitney U test (see nag_nonpar_test_mwu (g08ah)) is more appropriate.

9 Example

This example is taken from Moore *et al.* (1972). There are 5 groups of sizes 5, 8, 6, 8 and 8. The data represent the weight gain, in pounds, of pigs from five different litters under the same conditions.

9.1 Program Text

```
function g08af_example

fprintf('g08af example results\n\n');

x = [23; 27; 26; 19; 30;
     29; 25; 33; 36; 32; 28; 30; 31;
     38; 31; 28; 35; 33; 36;
     30; 27; 28; 22; 33; 34; 34; 32;
     31; 33; 31; 28; 30; 24; 29; 30];
l = [nag_int(5);8;6;8;8];

fprintf('Kruskal-Wallis test\n\n');
fprintf('Data values\n\n');
fprintf('  Group    Observations');

ix = 1;
for j = 1:numel(l)
  fprintf('\n%5d    ', j);
  fprintf('%4.0f', x(ix:ix+l(j)-1));
  ix = ix + l(j);
end

[h, p, ifail] = g08af(x, l);

fprintf('\n\nTest statistic        %8.3f\n', h);
fprintf('Degrees of freedom    %4d\n', numel(l)-1);
fprintf('Significance          %8.3f\n', p);
```

9.2 Program Results

```
     g08af example results

Kruskal-Wallis test

Data values

  Group    Observations
    1      23  27  26  19  30
    2      29  25  33  36  32  28  30  31
    3      38  31  28  35  33  36
    4      30  27  28  22  33  34  34  32
    5      31  33  31  28  30  24  29  30

Test statistic          10.537
Degrees of freedom       4
Significance             0.032
```