NAG C Library Function Document

nag_kruskal_wallis_test (g08afc)

1 Purpose

nag_kruskal_wallis_test (g08afc) performs the Kruskal–Wallis one-way analysis of variance by ranks on $k$ independent samples of possibly unequal sizes.

2 Specification

```c
#include <nag.h>
#include <nagc08.h>

void nag_kruskal_wallis_test (Integer k, const Integer l[], const double x[],
   Integer lx, double *h, double *p, NagError *fail)
```

3 Description

The Kruskal–Wallis test investigates the differences between scores from $k$ independent samples of unequal sizes, the $i$th 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:

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

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 group and the summation is over all tied groups.

nag_kruskal_wallis_test 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 $\chi^2_{k-1}$ distribution). $H_0$ is rejected by a test of chosen size $\alpha$ 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_median_test (g08acc)) or the Mann–Whitney $U$ test (see nag_mann_whitney (g08amc)) is more appropriate).

4 Parameters

1:  $k$ – Integer

   **Input**

   *On entry:* the number of samples, $k$.

   *Constraint:* $k \geq 2$.

2:  $l[i]$ – const Integer

   **Input**

   *On entry:* $l[i-1]$ must contain the number of observations $l_i$ in sample $i$, for $i = 1, 2, \ldots, k$.

   *Constraint:* $l[i-1] > 0$, for $i = 1, 2, \ldots, k$. 


[NP3491/6]
3: \( x[\mathbf{l}x] \) – const double \hspace{1cm} \text{Input}

\text{On entry:} the elements of \( x \) must contain the observations in the \( k \) groups. The first \( l_1 \) elements must contain the scores in the first group, the next \( l_2 \) those in the second group, and so on.

4: \( \mathbf{lx} \) – Integer \hspace{1cm} \text{Input}

\text{On entry:} the total number of observations, \( N \).

\text{Constraint:} \( \mathbf{lx} = \sum_{i=1}^{k} \mathbf{l}[i - 1] \).

5: \( \mathbf{h} \) – double * \hspace{1cm} \text{Output}

\text{On exit:} the value of the Kruskal–Wallis test statistic, \( H \).

6: \( \mathbf{p} \) – double * \hspace{1cm} \text{Output}

\text{On exit:} the approximate significance, \( p \), of the Kruskal–Wallis test statistic.

7: \( \text{fail} \) – NagError * \hspace{1cm} \text{Input/Output}

The NagError error parameter (see the Essential Introduction).

5 \hspace{1cm} \text{Error Indicators and Warnings}

\textbf{NE\_INT\_ARG\_LT}

\text{On entry,} \( k \) must not be less than 2: \( k = <\text{value}> \).

\textbf{NE\_ARRAY\_CONS}

The contents of array \( l \) are not valid.

\text{Constraint:} \( \mathbf{l}[i - 1] > 0 \), for \( i = 1, 2, \ldots, k \).

\textbf{NE\_INT}

\text{On entry,} \( \mathbf{lx} = <\text{value}> \).

\text{Constraint:} \( \mathbf{lx} = \sum_{i=1}^{k} \mathbf{l}[i - 1] \), for \( i = 1, 2, \ldots, k \).

\textbf{NE\_X\_IDEN}

\text{On entry, all elements of} \( x \) \text{ are equal.}

\textbf{NE\_ALLOC\_FAIL}

Memory allocation failed.

\textbf{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 \hspace{1cm} \text{Further Comments}

The time taken by the routine is small, and increases with \( N \) and \( k \).

If \( k = 2 \), the Median test (see \text{nag\_median\_test} (g08acc)) or the Mann–Whitney \( U \) test (see \text{nag\_mann\_whitney} (g08amc)) is more appropriate.

6.1 \hspace{1cm} \textbf{Accuracy}

For estimates of the accuracy of the significance \( p \), see \text{nag\_prob\_chi\_sq} (g01ecc). The \( \chi^2 \) approximation is acceptable unless \( k = 3 \) and \( l_1, l_2 \) or \( l_3 \leq 5 \).
6.2 References

7 See Also
nag_prob_chisq (g01ec)
nag_median (g08a)
nag_mann_whitney (g08am)

8 Example
This example is taken from Moore et al. 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.

8.1 Program Text
/* nag_kruskal_wallis_test (g08afc) Example Program. *
 * Copyright 2000 Numerical Algorithms Group. *
 * Mark 6, 2000. */

#include <stdio.h>
#include <nag.h>
#include <nag_stdlib.h>
#include <nagg08.h>

int main (void)
{
  double h, p, *x=0;
  Integer count, i, ii, k, *l=0, lx, nhi, ni, nlo;
  Integer exit_status=0;
  NagError fail;

  INIT_FAIL(fail);
  Vprintf("g08afc Example Program Results\n");

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

  k=5;
  if (!((l =NAG_ALLOC(k, Integer)))
  {
    Vprintf("Allocation failure\n");
    exit_status = -1;
    goto END;
  }
  for (i=1; i<k; i++)
  Vscanf("%ld", &l[i-1]);
  Vprintf("\n");
  Vprintf("%s\n", "Kruskal-Wallis test");
  Vprintf("\n");
  Vprintf("%s\n", "Data values");
  Vprintf("\n");
```c
Vprintf("%s\n"," Group Observations");

lx = 0;
for (i = 1; i <= 5; ++i)
    lx += l[1 - i];

if (! (x = NAG_ALLOC(lx, double)))
    {
    Vprintf("Allocation failure\n");
    exit_status = -1;
    goto END;
    }
for (i = 1; i <= lx; ++i)
    Vscanf("%lf", &x[i - 1]);

nlo = 1;
for (i = 1; i <= k; ++i)
    {
    ni = l[i - 1];
    nhi = nlo + ni - 1;
    Vprintf(" %5ld ", i);
    count = 1;
    for (ii = nlo; ii <= nhi; ++ii)
        {
        Vprintf("%4.0f\%s", x[ii - 1], count%10?"\n":"
");
        count++;
        }
    nlo += ni;
    Vprintf("\n");
    }
g08afc(k, l, x, lx, &h, &p, &fail);
if (fail.code != NE_NOERROR)
    {
    Vprintf("Error from g08afc.\n\n", fail.message);
    exit_status = 1;
    goto END;
    }
Vprintf("\n");
Vprintf("%s9.3f\n","Test statistic ", h);
Vprintf("%s91d\n","Degrees of freedom ", k-1);
Vprintf("%s9.3f\n","Significance ", p);
END:
if (l) NAG_FREE(l);
if (x) NAG_FREE(x);
return exit_status;
}

8.2 Program Data

g08afc Example Program Data
5 8 6 8 8
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 32 31 31 33 31
28 30 24 29 30
```
8.3 Program Results

g08afc Example Program Results

Kruskal-Wallis test

Data values

<table>
<thead>
<tr>
<th>Group</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23 27 26 19 30</td>
</tr>
<tr>
<td>2</td>
<td>29 25 33 36 32 28 30 31</td>
</tr>
<tr>
<td>3</td>
<td>38 31 28 35 33 36</td>
</tr>
<tr>
<td>4</td>
<td>30 27 28 22 33 34 34 32</td>
</tr>
<tr>
<td>5</td>
<td>31 33 31 28 30 24 29 30</td>
</tr>
</tbody>
</table>

Test statistic 10.537
Degrees of freedom 4
Significance 0.032