

# NAG Fortran Library Routine Document

## G08AFF

**Note:** before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

### 1 Purpose

G08AFF performs the Kruskal–Wallis one-way analysis of variance by ranks on  $k$  independent samples of possibly unequal sizes.

### 2 Specification

```
SUBROUTINE G08AFF(X, LX, L, K, W, H, P, IFAIL)
INTEGER          LX, L(K), K, IFAIL
real           X(LX), W(LX), H, P
```

### 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:

- The pooled sample of all the observations is ranked. Average ranks are assigned to tied scores.
- The ranks of the observations in each sample are summed, to give the rank sums  $R_i$ , for  $i = 1, 2, \dots, k$ .
- 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.

G08AFF 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  $\chi_{k-1}^2$  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 G08ACF) or the Mann–Whitney  $U$  test (see G08AHF) 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

- 1:  $X(LX)$  – *real* array *Input*  
*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.
- 2:  $LX$  – INTEGER *Input*  
*On entry:* the total number of observations,  $N$ .  
*Constraint:*  $LX = \sum_{i=1}^k L(i)$ .
- 3:  $L(K)$  – INTEGER array *Input*  
*On entry:*  $L(i)$  must contain the number of observations  $l_i$  in sample  $i$ , for  $i = 1, 2, \dots, k$ .  
*Constraint:*  $L(i) > 0$ , for  $i = 1, 2, \dots, k$ .
- 4:  $K$  – INTEGER *Input*  
*On entry:* the number of samples,  $k$ .  
*Constraint:*  $K \geq 2$ .
- 5:  $W(LX)$  – *real* array *Workspace*
- 6:  $H$  – *real* *Output*  
*On exit:* the value of the Kruskal–Wallis test statistic,  $H$ .
- 7:  $P$  – *real* *Output*  
*On exit:* the approximate significance,  $p$ , of the Kruskal–Wallis test statistic.
- 8:  $IFAIL$  – INTEGER *Input/Output*  
*On entry:*  $IFAIL$  must be set to 0,  $-1$  or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.  
*On exit:*  $IFAIL = 0$  unless the routine detects an error (see Section 6).  
 For environments where it might be inappropriate to halt program execution when an error is detected, the value  $-1$  or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, for users not familiar with this parameter the recommended value is 0. **When the value  $-1$  or 1 is used it is essential to test the value of  $IFAIL$  on exit.**

## 6 Error Indicators and Warnings

If on entry  $IFAIL = 0$  or  $-1$ , explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

$IFAIL = 1$

On entry,  $K < 2$ .

$IFAIL = 2$

On entry,  $L(i) \leq 0$  for some  $i$ ,  $i = 1, 2, \dots, k$ .

$IFAIL = 3$

On entry,  $LX \neq \sum_{i=1}^k L(i)$ .

IFAIL = 4

On entry, all the observations were equal.

## 7 Accuracy

For estimates of the accuracy of the significance  $p$ , see G01ECF. The  $\chi^2$  approximation is acceptable unless  $k = 3$  and  $l_1, l_2$  or  $l_3 \leq 5$ .

## 8 Further Comments

The time taken by the routine is small, and increases with  $N$  and  $k$ .

If  $k = 2$ , the Median test (see G08ACF) or the Mann–Whitney  $U$  test (see G08AHF) 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

**Note:** the listing of the example program presented below uses *bold italicised* terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*      G08AFF Example Program Text
*      Mark 14 Revised.  NAG Copyright 1989.
*      .. Parameters ..
INTEGER          K, LMAX
PARAMETER        (K=5,LMAX=35)
INTEGER          NIN, NOUT
PARAMETER        (NIN=5,NOUT=6)
*      .. Local Scalars ..
real           H, P
INTEGER          I, IFAIL, II, LX, NHI, NI, NLO
*      .. Local Arrays ..
real          W1(LMAX), X(LMAX)
INTEGER          L(K)
*      .. External Subroutines ..
EXTERNAL         G08AFF
*      .. Executable Statements ..
WRITE (NOUT,*) 'G08AFF Example Program Results'
*      Skip heading in data file
READ (NIN,*)
READ (NIN,*) L
WRITE (NOUT,*)
WRITE (NOUT,*) 'Kruskal-Wallis test'
WRITE (NOUT,*)
WRITE (NOUT,*) 'Data values'
WRITE (NOUT,*)
WRITE (NOUT,*) '  Group      Observations'
LX = 0
DO 20 I = 1, K
    LX = LX + L(I)
20 CONTINUE
IF (LX.LE.LMAX) THEN
    READ (NIN,*) (X(I),I=1,LX)
    IFAIL = 0
    NLO = 1
    DO 40 I = 1, K
        NI = L(I)
        NHI = NLO + NI - 1
        WRITE (NOUT,99999) I, (X(II),II=NLO,NHI)
        NLO = NLO + NI
    40 CONTINUE
*

```

```

      CALL G08AFF(X,LX,L,K,W1,H,P,IFAIL)
*
      WRITE (NOUT,*)
      WRITE (NOUT,99998) 'Test statistic      ', H
      WRITE (NOUT,99997) 'Degrees of freedom ', K - 1
      WRITE (NOUT,99998) 'Significance      ', P
      END IF
      STOP
*
99999 FORMAT (1X,I5,5X,10F4.0)
99998 FORMAT (1X,A,F9.3)
99997 FORMAT (1X,A,I9)
      END

```

## 9.2 Program Data

G08AFF 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 34 32 31 33 31
28 30 24 29 30

```

## 9.3 Program Results

G08AFF Example Program 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

---