NAG Library Routine Document

G11AAF

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

G11AAF computes χ^2 statistics for a two-way contingency table. For a 2×2 table with a small number of observations exact probabilities are computed.

2 Specification

```
SUBROUTINE G11AAF (NROW, NCOL, NOBS, LDNOBS, EXPT, CHIST, PROB, CHI, G,    &
                   DF, IFAIL)
INTEGER            NROW, NCOL, NOBS(LDNOBS,NCOL), LDNOBS, IFAIL
REAL (KIND=nag_wp) EXPT(LDNOBS,NCOL), CHIST(LDNOBS,NCOL), PROB, CHI,       &
                   G, DF
```

3 Description

For a set of n observations classified by two variables, with r and c levels respectively, a two-way table of frequencies with r rows and c columns can be computed.

$$\begin{array}{cccc|c} n_{11} & n_{12} & \dots & n_{1c} & n_{1.} \\ n_{21} & n_{22} & \dots & n_{2c} & n_{2.} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n_{r1} & n_{r2} & \dots & n_{rc} & n_{r.} \\ \hline n_{.1} & n_{.2} & \dots & n_{.c} & n \end{array}$$

To measure the association between the two classification variables two statistics that can be used are, the Pearson χ^2 statistic, $\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij}-f_{ij}\right)^2}{f_{ij}}$, and the likelihood ratio test statistic, $\displaystyle 2\sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} \times \log\left(n_{ij}/f_{ij}\right)$, where f_{ij} are the fitted values from the model that assumes the effects due to the classification variables are additive, i.e., there is no association. These values are the expected cell frequencies and are given by

$$f_{ij} = n_{i.}n_{.j}/n.$$

Under the hypothesis of no association between the two classification variables, both these statistics have, approximately, a χ^2-distribution with $(c-1)(r-1)$ degrees of freedom. This distribution is arrived at under the assumption that the expected cell frequencies, f_{ij}, are not too small. For a discussion of this point see Everitt (1977). He concludes by saying, '... in the majority of cases the chi-square criterion may be used for tables with expectations in excess of 0.5 in the smallest cell'.

In the case of the 2×2 table, i.e., $c = 2$ and $r = 2$, the χ^2 approximation can be improved by using Yates' continuity correction factor. This decreases the absolute value of $\left(n_{ij} - f_{ij}\right)$ by $\frac{1}{2}$. For 2×2 tables with a small value of n the exact probabilities from Fisher's test are computed. These are based on the hypergeometric distribution and are computed using G01BLF. A two tail probability is computed as $\min(1, 2p_u, 2p_l)$, where p_u and p_l are the upper and lower one-tail probabilities from the hypergeometric distribution.

4 References

Everitt B S (1977) *The Analysis of Contingency Tables* Chapman and Hall

Kendall M G and Stuart A (1973) *The Advanced Theory of Statistics (Volume 2)* (3rd Edition) Griffin

5 Arguments

1: NROW – INTEGER *Input*

On entry: r, the number of rows in the contingency table.

Constraint: NROW ≥ 2.

2: NCOL – INTEGER *Input*

On entry: c, the number of columns in the contingency table.

Constraint: NCOL ≥ 2.

3: NOBS(LDNOBS, NCOL) – INTEGER array *Input*

On entry: the contingency table NOBS(i,j) must contain n_{ij}, for $i = 1,2,\ldots,r$ and $j = 1,2,\ldots,c$.

Constraint: NOBS$(i,j) \geq 0$, for $i = 1,2,\ldots,r$ and $j = 1,2,\ldots,c$.

4: LDNOBS – INTEGER *Input*

On entry: the first dimension of the arrays NOBS, EXPT and CHIST as declared in the (sub) program from which G11AAF is called.

Constraint: LDNOBS $\geq$ NROW.

5: EXPT(LDNOBS, NCOL) – REAL (KIND=nag_wp) array *Output*

On exit: the table of expected values. EXPT(i,j) contains f_{ij}, for $i = 1,2,\ldots,r$ and $j = 1,2,\ldots,c$.

6: CHIST(LDNOBS, NCOL) – REAL (KIND=nag_wp) array *Output*

On exit: the table of χ^2 contributions. CHIST(i,j) contains $\dfrac{(n_{ij}-f_{ij})^2}{f_{ij}}$, for $i = 1,2,\ldots,r$ and $j = 1,2,\ldots,c$.

7: PROB – REAL (KIND=nag_wp) *Output*

On exit: if $c = 2$, $r = 2$ and $n \leq 40$ then PROB contains the two tail significance level for Fisher's exact test, otherwise PROB contains the significance level from the Pearson χ^2 statistic.

8: CHI – REAL (KIND=nag_wp) *Output*

On exit: the Pearson χ^2 statistic.

9: G – REAL (KIND=nag_wp) *Output*

On exit: the likelihood ratio test statistic.

10: DF – REAL (KIND=nag_wp) *Output*

On exit: the degrees of freedom for the statistics.

11: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this argument you should refer to Section 3.4 in How to Use the NAG Library and its Documentation for details.

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, because for this routine the values of the output arguments may be useful even if IFAIL $\neq 0$ on exit, the recommended value is -1. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

On exit: IFAIL $= 0$ unless the routine detects an error or a warning has been flagged (see Section 6).

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).

Note: G11AAF may return useful information for one or more of the following detected errors or warnings.

Errors or warnings detected by the routine:

IFAIL $= 1$

On entry, NROW < 2,
or NCOL < 2,
or LDNOBS $<$ NROW.

IFAIL $= 2$

On entry, a value in NOBS < 0, or all values in NOBS are zero.

IFAIL $= 3$

On entry, a 2×2 table has a row or column with both values 0.

IFAIL $= 4$

At least one cell has expected frequency, f_{ij}, ≤ 0.5. The χ^2 approximation may be poor.

IFAIL $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.9 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -399$

Your licence key may have expired or may not have been installed correctly.

See Section 3.8 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -999$

Dynamic memory allocation failed.

See Section 3.7 in How to Use the NAG Library and its Documentation for further information.

7 Accuracy

For the accuracy of the probabilities for Fisher's exact test see G01BLF.

8 Parallelism and Performance

G11AAF is not threaded in any implementation.

9 Further Comments

The routine G01AFF allows for the automatic amalgamation of rows and columns. In most circumstances this is not recommended; see Everitt (1977).

Multidimensional contingency tables can be analysed using log-linear models fitted by G02GBF.

10 Example

The data below, taken from Everitt (1977), is from 141 patients with brain tumours. The row classification variable is the site of the tumour: frontal lobes, temporal lobes and other cerebral areas. The column classification variable is the type of tumour: benign, malignant and other cerebral tumours.

23	9	6	38
21	4	3	28
34	24	17	75
78	37	26	141

The data is read in and the statistics computed and printed.

10.1 Program Text

```
    Program g11aafe

!     G11AAF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g11aaf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: chi, df, g, prob
      Integer                          :: i, ifail, ldnobs, ncol, nrow
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: chist(:,:), expt(:,:)
      Integer, Allocatable             :: nobs(:,:)
!     .. Executable Statements ..
      Write (nout,*) ' G11AAF Example Program Results'
      Write (nout,*)

!     Skip heading in data file
      Read (nin,*)

!     Read in the problem size
      Read (nin,*) nrow, ncol

      ldnobs = nrow
      Allocate (nobs(ldnobs,ncol),expt(ldnobs,ncol),chist(ldnobs,ncol))

!     Read in data
      Read (nin,*)(nobs(i,1:ncol),i=1,nrow)

!     Perform chi-squared test
      ifail = -1
      Call g11aaf(nrow,ncol,nobs,ldnobs,expt,chist,prob,chi,g,df,ifail)
      If (ifail/=0) Then
        If (ifail/=3) Then
```

```
          Go To 100
        End If
      End If

!     Display results
      Write (nout,99999) ' Probability =', prob
      Write (nout,99998) ' Pearson Chi-square statistic = ', chi
      Write (nout,99998) ' Likelihood ratio test statistic = ', g
      Write (nout,99997) ' Degrees of freedom = ', df

100   Continue

99999 Format (A,F7.4)
99998 Format (A,F8.3)
99997 Format (A,F4.0)
    End Program g11aafe
```

10.2 Program Data

```
G11AAF Example Program Data
3 3                         : NROW NCOL
23  9  6
21  4  3
34 24 17                    : End of NOBS
```

10.3 Program Results

```
  G11AAF Example Program Results

 Probability = 0.0975
 Pearson Chi-square statistic =    7.844
 Likelihood ratio test statistic =    8.096
 Degrees of freedom =   4.
```