# NAG Fortran Library Routine Document

# G01ADF

**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

G01ADF calculates the mean, standard deviation and coefficients of skewness and kurtosis for data grouped in a frequency distribution.

## 2    Specification

```
SUBROUTINE G01ADF(K, X, IFREQ, XMEAN, S2, S3, S4, N, IFAIL)
INTEGER        K, IFREQ(K), N, IFAIL
real           X(K), XMEAN, S2, S3, S4
```

## 3    Description

The input data consist of a univariate frequency distribution, denoted by $f_i$, for $i = 1, 2, \ldots, k - 1$, and the boundary values of the classes $x_i$, for $i = 1, 2, \ldots, k$. Thus the frequency associated with the interval $(x_i, x_{i+1})$ is $f_i$, and the routine assumes that all the values in this interval are concentrated at the point

$$y_i = (x_{i+1} + x_i)/2, \quad i = 1, 2, \ldots, k - 1.$$

The following quantities are calculated:

(a)   total frequency,

$$n = \sum_{i=1}^{k-1} f_i.$$

(b)   mean,

$$\bar{y} = \frac{\sum_{i=1}^{k-1} f_i y_i}{n}.$$

(c)   standard deviation,

$$s_2 = \sqrt{\frac{\sum_{i=1}^{k-1} f_i (y_i - \bar{y})^2}{(n-1)}}, \quad n \geq 2.$$

(d)   coefficient of skewness,

$$s_3 = \frac{\sum_{i=1}^{k-1} f_i (y_i - \bar{y})^3}{(n-1) \times s_2^3}, \quad n \geq 2.$$

(e)   coefficient of kurtosis,

$$s_4 = \frac{\sum_{i=1}^{k-1} f_i (y_i - \bar{y})^4}{(n-1) \times s_2^4} - 3, \quad n \geq 2.$$

The routine has been developed primarily for groupings of a continuous variable. If, however, the routine is to be used on the frequency distribution of a discrete variable, taking the values $y_1, \ldots, y_{k-1}$, then the boundary values for the classes may be defined as follows:

(i)   for $k > 2$,

$$
\begin{aligned}
x_1 &= (3y_1 - y_2)/2 \\
x_j &= (y_{j-1} + y_j)/2, \qquad j = 2, \ldots, k - 1 \\
x_k &= (3y_{k-1} - y_{k-2})/2
\end{aligned}
$$

(ii) for $k = 2$,

$$x_1 = y_1 - a \quad \text{and} \quad x_2 = y_1 + a \quad \text{for any } a > 0.$$

## 4 References

None.

## 5 Parameters

1:     K – INTEGER                                                                                                    *Input*

*On entry*: the number of class boundaries, which is one more than the number of classes of the frequency distribution, $k$.

*Constraint*: K $> 1$.

2:     X(K) – ***real*** array                                                                                         *Input*

*On entry*: the elements of X must contain the boundary values of the classes in ascending order, so that class $i$ is bounded by the values in X($i$) and X($i+1$), for $i = 1, 2, \ldots, k-1$.

*Constraint*: X($i$) $<$ X($i+1$), for $i = 1, 2, \ldots, k-1$.

3:     IFREQ(K) – INTEGER array                                                                                   *Input*

*On entry*: the $i$th element of IFREQ must contain the frequency associated with the $i$th class, for $i = 1, 2, \ldots, k-1$. IFREQ($k$) is not used by the routine.

*Constraint*: IFREQ($i$) $\geq 0$, for $i = 1, 2, \ldots, k-1$ and $\sum_{i=1}^{k-1}$ IFREQ($i$) $> 0$.

4:     XMEAN – ***real***                                                                                          *Output*

*On exit*: the mean value, $\bar{y}$.

5:     S2 – ***real***                                                                                               *Output*

*On exit*: the standard deviation, $s_2$.

6:     S3 – ***real***                                                                                               *Output*

*On exit*: the coefficient of skewness, $s_3$.

7:     S4 – ***real***                                                                                               *Output*

*On exit*: the coefficient of kurtosis, $s_4$.

8:     N – INTEGER                                                                                                  *Output*

*On exit*: the total frequency, $n$.

9:     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, $K \leq 1$.

IFAIL $= 2$

> On entry, the boundary values of the classes in X are not in ascending order.

IFAIL $= 3$

> On entry, $\sum_{i=1}^{k-1} \text{IFREQ}(i) = 0$ or $\text{IFREQ}(i) < 0$ for some $i$, for $i = 1, 2, \ldots, k - 1$.

IFAIL $= 4$

> The total frequency, $n$, is less than 2, hence the quantities $s_2$, $s_3$ and $s_4$ cannot be calculated.

## 7 Accuracy

The method used is believed to be stable.

## 8 Further Comments

The time taken by the routine increases linearly with $k$.

## 9 Example

In the example program, NPROB determines the number of sets of data to be analysed. For each analysis, the boundary values of the classes and the frequencies are read. After the routine has been successfully called, the input data and calculated quantities are printed. In the example, there is one set of data, with 14 classes.

### 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.

```
*     G01ADF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER         KMAX
      PARAMETER       (KMAX=50)
      INTEGER         NIN, NOUT
      PARAMETER       (NIN=5,NOUT=6)
*     .. Local Scalars ..
      real            S2, S3, S4, XMEAN
      INTEGER         I, IFAIL, J, K, KMIN1, N, NPROB
*     .. Local Arrays ..
      real            X(KMAX)
      INTEGER         IFREQ(KMAX)
*     .. External Subroutines ..
      EXTERNAL        G01ADF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G01ADF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) NPROB
      DO 20 J = 1, NPROB
```

```
          READ (NIN,*) KMIN1
          K = KMIN1 + 1
          IF (K.GE.2 .AND. K.LE.KMAX) THEN
              READ (NIN,*) (X(I),IFREQ(I),I=1,KMIN1), X(K)
              WRITE (NOUT,*)
              WRITE (NOUT,99999) 'Problem ', J
              WRITE (NOUT,99999) 'Number of classes ', KMIN1
              IFAIL = 1
*
              CALL G01ADF(K,X,IFREQ,XMEAN,S2,S3,S4,N,IFAIL)
*
              WRITE (NOUT,*)
              IF (IFAIL.EQ.0) THEN
                  WRITE (NOUT,*) 'Successful call of G01ADF'
                  WRITE (NOUT,*)
                  WRITE (NOUT,*) '        Class         Frequency'
                  WRITE (NOUT,*)
                  WRITE (NOUT,99998) (X(I),X(I+1),IFREQ(I),I=1,KMIN1)
                  WRITE (NOUT,*)
                  WRITE (NOUT,99997) ' Mean ', XMEAN
                  WRITE (NOUT,99996) ' Std devn', S2
                  WRITE (NOUT,99996) ' Skewness', S3
                  WRITE (NOUT,99996) ' Kurtosis', S4
                  WRITE (NOUT,99995) ' Number of cases', N
              ELSE
                  WRITE (NOUT,99999)
     +              'Unsuccessful call of G01ADF. IFAIL = ', IFAIL
              END IF
          ELSE
              STOP
          END IF
   20 CONTINUE
      STOP
*
99999 FORMAT (1X,A,I4)
99998 FORMAT (1X,2F10.2,I12)
99997 FORMAT (1X,A,F16.4)
99996 FORMAT (1X,A,F13.4)
99995 FORMAT (1X,A,I8)
      END
```

## 9.2   Program Data

```
G01ADF Example Program Data
 1
14
     9.3         3        12        19        14        52        16        96
    18         121        20        115        22        86        24        70
    26          49        28        31        30        16        32         6
    34           8        36         7      39.7
```

## 9.3   Program Results

```
 G01ADF Example Program Results

 Problem    1
 Number of classes   14

 Successful call of G01ADF

         Class         Frequency

     9.30     12.00          3
    12.00     14.00         19
    14.00     16.00         52
    16.00     18.00         96
    18.00     20.00        121
    20.00     22.00        115
    22.00     24.00         86
    24.00     26.00         70
```

```
26.00     28.00          49
28.00     30.00          31
30.00     32.00          16
32.00     34.00           6
34.00     36.00           8
36.00     39.70           7

Mean           21.4932
Std devn        4.9325
Skewness        0.7072
Kurtosis        0.5738
Number of cases     679
```