NAG Toolbox for MATLAB

g08ba

1 Purpose

g08ba performs Mood's and David's tests for dispersion differences between two independent samples of possibly unequal size.

2 Syntax

```
[r, w, v, pw, pv, ifail] = g08ba(x, n1, itest, 'n', n)
```

3 Description

Mood's and David's tests investigate the difference between the dispersions of two independent samples of sizes n_1 and n_2, denoted by

$$x_1, x_2, \ldots, x_{n_1}$$

and

$$x_{n_1+1}, x_{n_1+2}, \ldots, x_n, \qquad n = n_1 + n_2.$$

The hypothesis under test, H_0, often called the null hypothesis, is that the dispersion difference is zero, and this is to be tested against a one- or two-sided alternative hypothesis H_1 (see below).

Both tests are based on the rankings of the sample members within the pooled sample formed by combining both samples. If there is some difference in dispersion, more of the extreme ranks will tend to be found in one sample than in the other.

Let the rank of x_i be denoted by r_i, for $i = 1, 2, \ldots, n$.

(a) Mood's test.

The test statistic $W = \sum_{i=1}^{n_1} \left(r_i - \frac{n+1}{2} \right)^2$ is found.

W is the sum of squared deviations from the average rank in the pooled sample. For large n, W approaches normality, and so an approximation, p_w, to the probability of observing W not greater than the computed value, may be found.

g08ba returns W and p_w if Mood's test is selected.

(b) David's test.

The disadvantage of Mood's test is that it assumes that the means of the two samples are equal. If this assumption is unjustified a high value of W could merely reflect the difference in means. David's test reduces this effect by using the variance of the ranks of the first sample about their mean rank, rather than the overall mean rank.

The test statistic for David's test is

$$V = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (r_i - \bar{r})^2$$

where

$$\bar{r} = \frac{\sum_{i=1}^{n_1} r_i}{n_1}.$$

For large n, V approaches normality, enabling an approximate probability p_v to be computed, similarly to p_w.

g08ba returns V and p_v if David's test is selected.

Suppose that a significance test of a chosen size α is to be performed (i.e., α is the probability of rejecting H_0 when H_0 is true; typically α is a small quantity such as 0.05 or 0.01).

The returned value p ($= p_v$ or p_w) can be used to perform a significance test, against various alternative hypotheses H_1, as follows.

(i) H_1: dispersions are unequal. H_0 is rejected if $2 \times \min(p, 1-p) < \alpha$.

(ii) H_1: dispersion of sample 1 $>$ dispersion of sample 2. H_0 is rejected if $1 - p < \alpha$.

(iii) H_1: dispersion of sample 2 $>$ dispersion of sample 1. H_0 is rejected if $p < \alpha$.

4 References

Cooper B E 1975 *Statistics for Experimentalists* Pergamon Press

5 Parameters

5.1 Compulsory Input Parameters

1: **x(n) – double array**

The first n_1 elements of **x** must be set to the data values in the first sample, and the next n_2 ($= \mathbf{n} - n_1$) elements to the data values in the second sample.

2: **n1 – int32 scalar**

The size of the first sample, n_1.

Constraint: $1 < \mathbf{n1} < \mathbf{n}$.

3: **itest – int32 scalar**

The test(s) to be carried out.

itest $= 0$

Both Mood's and David's tests.

itest $= 1$

David's test only.

itest $= 2$

Mood's test only.

Constraint: **itest** $= 0$, 1 or 2.

5.2 Optional Input Parameters

1: **n – int32 scalar**

Default: The dimension of the arrays **x**, **r**. (An error is raised if these dimensions are not equal.) the total of the two sample sizes, n ($= n_1 + n_2$).

Constraint: $\mathbf{n} > 2$.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **r(n) – double array**

The ranks r_i, assigned to the data values x_i, for $i = 1, 2, \ldots, n$.

2: **w – double scalar**

Mood's test statistic, W, if requested.

3: **v – double scalar**

David's test statistic, V, if requested.

4: **pw – double scalar**

The lower tail probability, p_w, corresponding to the value of W, if Mood's test was requested.

5: **pv – double scalar**

The lower tail probability, p_v, corresponding to the value of V, if David's test was requested.

6: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail $= 1$

On entry, **n** ≤ 2.

ifail $= 2$

On entry, **n1** ≤ 1,
or **n1** $\geq$ **n**.

ifail $= 3$

On entry, **itest** < 0,
or **itest** > 2.

7 Accuracy

All computations are believed to be stable. The statistics V and W should be accurate enough for all practical uses.

8 Further Comments

The time taken by g08ba is small, and increases with n.

9 Example

```
x = [6;
     9;
     12;
     4;
     10;
     11;
```

```
     8;
     1;
     3;
     7;
     2;
     5];
n1 = int32(6);
itest = int32(0);
[r, w, v, pw, pv, ifail] = g08ba(x, n1, itest)
```

```
r =
     6
     9
    12
     4
    10
    11
     8
     1
     3
     7
     2
     5
w =
   75.5000
v =
    9.4667
pw =
    0.5830
pv =
    0.1986
ifail =
           0
```