

NAG Toolbox for MATLAB

g08ac

1 Purpose

g08ac performs the Median test on two independent samples of possibly unequal size.

2 Syntax

```
[i1, i2, p, ifail] = g08ac(x, n1, 'n', n)
```

3 Description

The Median test investigates the difference between the medians of two independent samples of sizes n_1 and n_2 , denoted by:

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

and

$$x_{n_1+1}, x_{n_1+2}, \dots, x_n,$$

where $n = n_1 + n_2$.

The hypothesis under test, H_0 , often called the null hypothesis, is that the medians are the same, and this is to be tested against the alternative hypothesis H_1 that they are different.

The test proceeds by forming a 2×2 frequency table, giving the number of scores in each sample above and below the median of the pooled sample:

	Sample 1	Sample 2	Total
Scores \leq pooled median	i_1	i_2	$i_1 + i_2$
Scores \geq pooled median	$n_1 - i_1$	$n_2 - i_2$	$n - (i_1 + i_2)$
Total	n_1	n_2	n

Under the null hypothesis, H_0 , we would expect about half of each group's scores to be above the pooled median and about half below, that is, we would expect i_1 , to be about $n_1/2$ and i_2 to be about $n_2/2$.

g08ac returns:

- (a) the frequencies i_1 and i_2 ;
- (b) the probability, p , of observing a table at least as 'extreme' as that actually observed, given that H_0 is true. If $n < 40$, p is computed directly ('Fisher's exact test'); otherwise a χ^2_1 approximation is used (see g01af).

H_0 is rejected by a test of chosen size α if $p < \alpha$.

4 References

Siegel S 1956 *Non-parametric Statistics for the Behavioral Sciences* McGraw-Hill

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 ($= n - n_1$) elements to the data values in the second sample.

- 2: **n1 – int32 scalar**
 The size of the first sample n_1 .
Constraint: $1 \leq \mathbf{n1} < \mathbf{n}$.

5.2 Optional Input Parameters

- 1: **n – int32 scalar**
Default: The dimension of the array **x**.
 the total of the two sample sizes, $n (= n_1 + n_2)$.
Constraint: $\mathbf{n} \geq 2$.

5.3 Input Parameters Omitted from the MATLAB Interface

w

5.4 Output Parameters

- 1: **i1 – int32 scalar**
 The number of scores in the first sample which lie below the pooled median, i_1 .
- 2: **i2 – int32 scalar**
 The number of scores in the second sample which lie below the pooled median, i_2 .
- 3: **p – double scalar**
 The tail probability p corresponding to the observed dichotomy of the two samples.
- 4: **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, $\mathbf{n} < 2$.

ifail = 2

On entry, $\mathbf{n1} < 1$,
 or $\mathbf{n1} \geq \mathbf{n}$.

7 Accuracy

The probability returned should be accurate enough for practical use.

8 Further Comments

The time taken by g08ac is small, and increases with n .

9 Example



```
x = [13;  
     6;  
     12;  
     7;  
     12;  
     7;  
     10;  
     7;  
     10;  
     7;  
     10;  
     7;  
     10;  
     8;  
     9;  
     8;  
     17;  
     6;  
     16;  
     8;  
     15;  
     8;  
     15;  
     10;  
     15;  
     10;  
     14;  
     10;  
     14;  
     11;  
     14;  
     11;  
     13;  
     12;  
     13;  
     12;  
     13;  
     12;  
     12];  
n1 = int32(16);  
[i1, i2, p, ifail] = g08ac(x, n1)
```

```
i1 =  
      13  
i2 =  
      6  
p =  
 8.8086e-04  
ifail =  
      0
```