NAG Toolbox for MATLAB

g08al

1   Purpose

g08al performs the Cochran Q-test on cross-classified binary data.

2   Syntax

```
[q, prob, ifail] = g08al(n, x, 'k', k)
```

3   Description

Cochran's Q-test may be used to test for differences between k treatments applied independently to n individuals or blocks (k related samples of equal size n), where the observed response can take only one of two possible values; for example a treatment may result in a 'success' or 'failure'. The data is recorded as either 1 or 0 to represent this dichotomization.

The use of this 'randomized block design' allows the effect of differences between the blocks to be separated from the differences between the treatments. The test assumes that the blocks were randomly selected from all possible blocks and that the result may be one of two possible outcomes common to all treatments within blocks.

The null and alternative hypotheses to be tested may be stated as follows.

H_0 : the treatments are equally effective, that is the probability of obtaining a 1 within a block is the same for each treatment.

H_1 : there is a difference between the treatments, that is the probability of obtaining a 1 is not the same for different treatments within blocks.

The data is often represented in the form of a table with the n rows representing the blocks and the k columns the treatments. Let R_i represent the row totals, for $i = 1, 2, \ldots, n$, and C_j represent the column totals, for $j = 1, 2, \ldots, k$. Let x_{ij} represent the response or result where $x_{ij} = 0$ or 1.

		Treatments				
Blocks	1	2		k	Row Totals
1	x_{11}	x_{12}	$\cdots$	x_{1k}	R_1
2	x_{21}	x_{22}	$\cdots$	x_{2k}	R_2
$\vdots$			$\vdots$		$\vdots$
n	x_{n1}	x_{n2}	$\cdots$	x_{nk}	R_n
Column Totals	C_1	C_2		C_k	N = Grand Total

If $p_{ij} = \Pr(x_{ij} = 1)$, for $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$, then the hypotheses may be restated as follows

H_0 : $p_{i1} = p_{i2} = \ldots = p_{ik}$, for each $i = 1, 2, \ldots, n$.

H_1: $p_{ij} \neq p_{ik}$, for some j and k, and for some i.

The test statistic is defined as

$$Q = \frac{k(k-1)\sum_{j=1}^{k}\left(C_j - \frac{N}{k}\right)^2}{\sum_{i=1}^{n}R_i(k - R_i)}.$$

When the number of blocks, n, is large relative to the number of treatments, k, Q has an approximate χ^2-distribution with $k - 1$ degrees of freedom. This is used to find the probability, p, of obtaining a statistic

greater than or equal to the computed value of Q. Thus p is the upper-tail probability associated with the computed value of Q, where the χ^2-distribution is used to approximate the true distribution of Q.

4 References

Conover W J 1980 *Practical Nonparametric Statistics* Wiley

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

5 Parameters

5.1 Compulsory Input Parameters

1: **n – int32 scalar**

n, the number of blocks.

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

2: **x(ldx,k) – double array**

ldx, the first dimension of the array, must be at least **n**.

The matrix of observed zero-one data. $\mathbf{x}(i,j)$ must contain the value x_{ij}, for $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$.

Constraint: $\mathbf{x}(i,j) = 0.0$ or 1.0, for all $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$.

5.2 Optional Input Parameters

1: **k – int32 scalar**

Default: The dimension of the array **x**.

k, the number of treatments.

Constraint: $\mathbf{k} \geq 2$.

5.3 Input Parameters Omitted from the MATLAB Interface

ldx

5.4 Output Parameters

1: **q – double scalar**

The value of the Cochran Q-test statistic.

2: **prob – double scalar**

The upper tail probability, p, associated with the Cochran Q-test statistic, that is the probability of obtaining a value greater than or equal to the observed value (the output value of **q**).

3: **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, $\mathbf{n} < 2$,
> or $\quad \mathbf{k} < 2$,
> or $\quad \mathbf{ldx} < \mathbf{n}$.

ifail $= 2$

> On entry, $\mathbf{x}(i,j) \neq 0.0$ or 1.0 for some i and j, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$.

ifail $= 3$

> The approximation process used to calculate the tail probability has failed to converge. The result returned in **prob** may still be a reasonable approximation.

7 Accuracy

The use of the χ^2-distribution as an approximation to the true distribution of the Cochran Q-test statistic improves as k increases and as n increases relative to k. This approximation should be a reasonable one when the total number of observations left, after omitting those rows containing all 0 or 1, is greater than about 25 and the number of rows left is larger than 5.

8 Further Comments

None.

9 Example

```
n = int32(12);
x = [1, 1, 1;
     1, 1, 1;
     0, 1, 0;
     1, 1, 0;
     0, 0, 0;
     1, 1, 1;
     1, 1, 1;
     1, 1, 0;
     0, 0, 1;
     0, 1, 0;
     1, 1, 1;
     1, 1, 1];
[q, prob, ifail] = g08al(n, x)

q =
    2.8000
prob =
    0.2466
ifail =
          0
```