**Program: detpattconsec**

**Author: Kruti Biren Shah**

The program detpattconsec has five arguments.

1.The variable name (a)

2.The id variable (c)

3.The size of a set of consecutive values (d)

4.The length of an answer value (e)

5.The threshold of repetitions (f)

**Example**

2 surveyors with ids 101 and 102 respectively, are tasked with conducting the assessments of 7 students. The assessment consists of 5 multiple choice questions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| q1 | q2 | q3 | q4 | q5 | surveyor\_id |
| 1 | 1 | 1 | 2 | 4 | 101 |
| 2 | 2 | 2 | 1 | 2 | 101 |
| 3 | 3 | 3 | 3 | 2 | 101 |
| 4 | 4 | 4 | 1 | 3 | 102 |
| 4 | 4 | 4 | 1 | 3 | 102 |
| 4 | 4 | 4 | 1 | 3 | 102 |
| 2 | 1 | 2 | 1 | 2 | 101 |

To detect if there are any patterns in sets of 2 consecutive responses, run the below codes. In this example, we assume that if a pattern is being repeated more than once, there is a possibility of data fabrication by the surveyor. In other words, the threshold of repetitions (f) is 1 in this case. Given that all the answer values are of length 1, e takes the value of 1.

clear

use "dummy\_data.dta" // importing the dummy dataset.

egen patt=concat(q1-q5) //combining the values of q1-q5

detpattconsec patt surveyor\_id 3 1 1

|  |  |  |  |
| --- | --- | --- | --- |
| surveyor\_id | repfreq | pattname | group |
| 101 | 1 | 212 | pattconsec\_q3\_g3 |
| 102 | 2 | 413 | pattconsec\_q3\_g3 |

From the table, we observe that surveyor 101 has recorded questions 3,4,5 as 2,1,2 twice. Similarly, surveyor 102 has recorded questions 3,4,5 as 4,1,3 three times. For a clearer understanding, see the table below. Note that the variable repfreq denotes the number of duplicates of the corresponding pattern. In the variable, group, the value pattconsec\_qi\_gj indicates that the corresponding pattern starts from question i and that it contains groups of j consecutive questions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| q1 | q2 | q3 | q4 | q5 | surveyor\_id |
| 1 | 1 | 1 | 2 | 4 | 101 |
| 2 | 2 | 2 | 1 | 2 | 101 |
| 3 | 3 | 3 | 3 | 2 | 101 |
| 4 | 4 | 4 | 1 | 3 | 102 |
| 4 | 4 | 4 | 1 | 3 | 102 |
| 4 | 4 | 4 | 1 | 3 | 102 |
| 2 | 1 | 2 | 1 | 2 | 101 |

The output is stored in the excel sheet “Consec\_cheats\_\_patt\_`suffix’” where suffix is a local macro containing the date of running the code.

**Uses**

1.To detect cheating in class quizzes

2. To detect patterns in self-enumerated surveys

3. To detect data fabrication by surveyors