Working Paper

A set of quality controls for stated preference data

Author(s):
Talati, Amit; König, Arnd; Axhausen, Kay W.

ETH Library

```
while (t<=(p-1))
        {
                flag = 0;

                if (b[t][choice-1]==1)
                {
                        flag=1;
                        t+=num_alt;
                        y=1;
                }
                else if(b[t][choice-1]==0)
                {
                        if(y==1)
                        {
                                flag=0;
                                break;
                        }
                }

                if(b[t][choice-1]==0 && y==0)
                {
                        t++;
                }
        }
```

A set of quality controls for Stated Preference data

A Talati

A König

KW Axhausen

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Arbeitsbericht Verkehrs- und Raumplanung

Titel der Arbeit

Amit Talati
IVT
ETH
CH-8093 Zürich

Arnd König
IVT
ETH
CH-8093 Zürich

KW Axhausen
IVT
ETH
CH-8093 Zürich

Telefon: 01-633 45 22
Telefax: 01-633 10 57
e-Mail- Adresse
arnd@ivt.baug.ethz.ch

Telefon: 01-633 39 522
Telefax: 01-633 10 57
e-Mail- Adresse
arnd@ivt.baug.ethz.ch

Telefon: 01-633 39 43
Telefax: 01-633 10 57
e-Mail- Adresse
axhausen@ivt.baug.ethz.ch

July 2002

Kurzfassung

Text

Schlagworte

Schlagwort 1, Schlagwort 2, ... Schlagwort n

Zitierungsvorschlag

Talati, A., A. König und K.W. Axhausen (2002) A set of quality control routines for Stated Preference data, *Arbeitsbericht Verkehrs- und Raumplanung,* **125**, Institut für Verkehrsplanung und Transportsysteme, ETH Zürich, Zürich.

Working paper

Title of the paper

Amit Talati
IVT
ETH
CH-8093 Zürich

Arnd König
IVT
ETH
CH-8093 Zürich

KW Axhausen
IVT
ETH
CH-8093 Zürich

Telephone: 01-633 45 22
Telefax: 01-633 10 57
e-Mail-Address
b9amit@civil.iitb.ac.in

Telephone: 01-633 39 522
Telefax: 01-633 10 57
e-Mail-Address
arnd@ivt.baug.ethz.ch

Telephone: 01-633 39 43
Telefax: 01-633 10 57
e-Mail-Address
axhausen@ivt.baug.ethz.ch

July 2002

Abstract

Stated Preference techniques have become a very popular tool for applied transport research. But there is still a lack of tools to test the quality of SP data. The main purpose of this paper is to discuss a set of routines for quality control checks. The focus is on testing three of quality elements; choice making, lexicographic answers and consistency of choice. Few respondents tend to ignore the survey and choose only a particular alternative for all questions, thus making no choice at all. On the other hand, some individuals are biased toward a particular factor (ex. Price) and make no trade between the attributes, hence making a lexicographic choice. The routines are designed in C++, which work on the data entered by the user, in a text format (.txt). The output is a set of text files giving the results and analysis of the tests.

Keywords

Stated Preference technique – Choice – Lexicography – Inconsistency – Working – ETH Zurich – Institute of Transportation, Traffic, Highway- and Railway- Engineering

Preferred citation style

Talati, A., A. König and K.W. Axhausen (2002) A set of quality control routines for Stated Preference data, *Arbeitsbericht Verkehrs- und Raumplanung,* **125**, Institut für Verkehrsplanung und Transportsysteme, ETH Zürich, Zürich.

1   Introduction

Stated preference (SP) techniques refer to a number of different approaches all of which use peoples` statements of how they would respond to different situations. During the last ten to fifteen years Stated Preference techniques has become a very powerful tool for applied transport research; see Lourviere, Hensher, and Swait (2000). While being powerful, SP experiments are a complicated tool for applied transport research. It is difficult for the respondents to make realistic choices and the design can be too demanding and complicated for the respondents. Thus the respondents may tend to a simplified response routine. Though the surveys today are made more realistic and less complicated for the respondents, there is still a lack of tools to test the SP-results and to find a "critical level of rejection". This is both a question of indicators for "non-trading" respondents and tests for "internal" or "external" inconsistency in the different surveys.
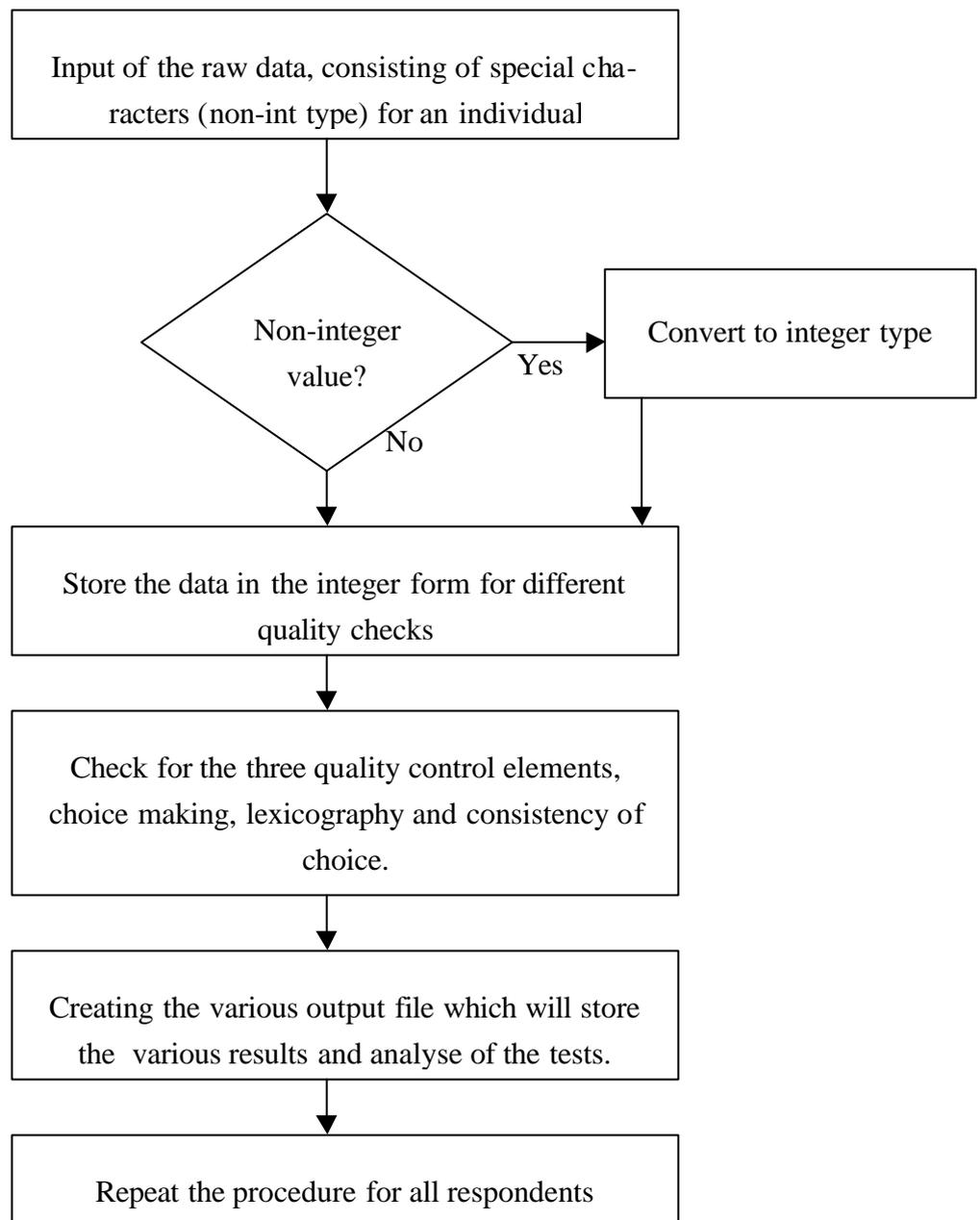
The set of routines, designed in C++, work on an input data file provided by the user. The user needs to supply the data file, in a text format, with few mandatory items about the data, the number of attributes, the starting column of the first attribute in the data entered and the column for choice variable. The routines work with or without any further information. Quality of the data is checked based on three parameters. Firstly, a check is made if the individual has taken the survey seriously and has made a choice at all. Secondly the data is checked for any lexicographic answers, to see if the respondent is trading between the factors included in the survey. Finally, a consistency check is performed to see whether the respondent makes a realistic choice, i.e., if the respondent has chosen the alternative with highest utility.

The results and analyses of the data are stored in the files named Result.txt, Personal_information.txt and Analyis.txt. The program automatically creates these files in the same directory from where the program is run.

2   Task

The task for making the routines for the quality check of the SP data can be summed up in the following flowchart.

Figure 1: Program structure

```
┌─────────────────────────────────────────────┐
│ Input of the raw data, consisting of special cha-│
│ racters (non-int type) for an individual        │
└─────────────────────────────────────────────┘

              ◇ Non-integer value?  ──Yes──→  ┌──────────────────────┐
                                               │ Convert to integer type │
                      │No                      └──────────────────────┘

┌─────────────────────────────────────────────┐
│ Store the data in the integer form for different│
│ quality checks                                  │
└─────────────────────────────────────────────┘

┌─────────────────────────────────────────────┐
│ Check for the three quality control elements,   │
│ choice making, lexicography and consistency of  │
│ choice.                                         │
└─────────────────────────────────────────────┘

┌─────────────────────────────────────────────┐
│ Creating the various output file which will store│
│ the  various results and analyse of the tests.  │
└─────────────────────────────────────────────┘

┌─────────────────────────────────────────────┐
│ Repeat the procedure for all respondents        │
└─────────────────────────────────────────────┘
```

3   The three quality control elements

The SP techniques are a powerful but complicated tool for transport research. The complexity of the design may lead to simplified response routines. In this context, it becomes very important to check the quality of the SP data. The three main elements for the quality control are checks on choice making, lexicography and consistency of answers.

3.1   Choice making

The first question arising while analysing the SP data is that whether the person, who is interviewed makes a choice at all. It is difficult for the respondents to make realistic choices and the design can be too demanding and complicated for the respondents. The respondent may not choose at all. The criteria for this test is, if the respondent selects a particular alternative for all the questions in the questionnaire. Following is the flowchart for the test.

3.2   Lexicographic answers

Lexicography refers to the problem with the respondents sorting/choosing the alternatives according to a particular factor without considering any other standard factor. For example, a person may value the price of a service above all and makes a choice according to this factor only without trading it of against other factors. The reason for a high number of lexicographic answers may be due to the following reasons:

- The choice of alternatives may be so complicated that the interviewees simplify their priorities.

- One or more factors may dominate, that is to say for example that the price alone means more than any other factor put together.

3.3   Consistency check

Consistent choices are the choices with a high probability of occurring based on general choice model estimated from the survey. Every alternative has a utility attached to it. To maximise the output a respondent must choose the alternative with the maximum utility. The utility is calculated from the parameters entered by the user. If only one set of parameters are given, then the routines checks for the consistency within a range of 5% of the parameters.

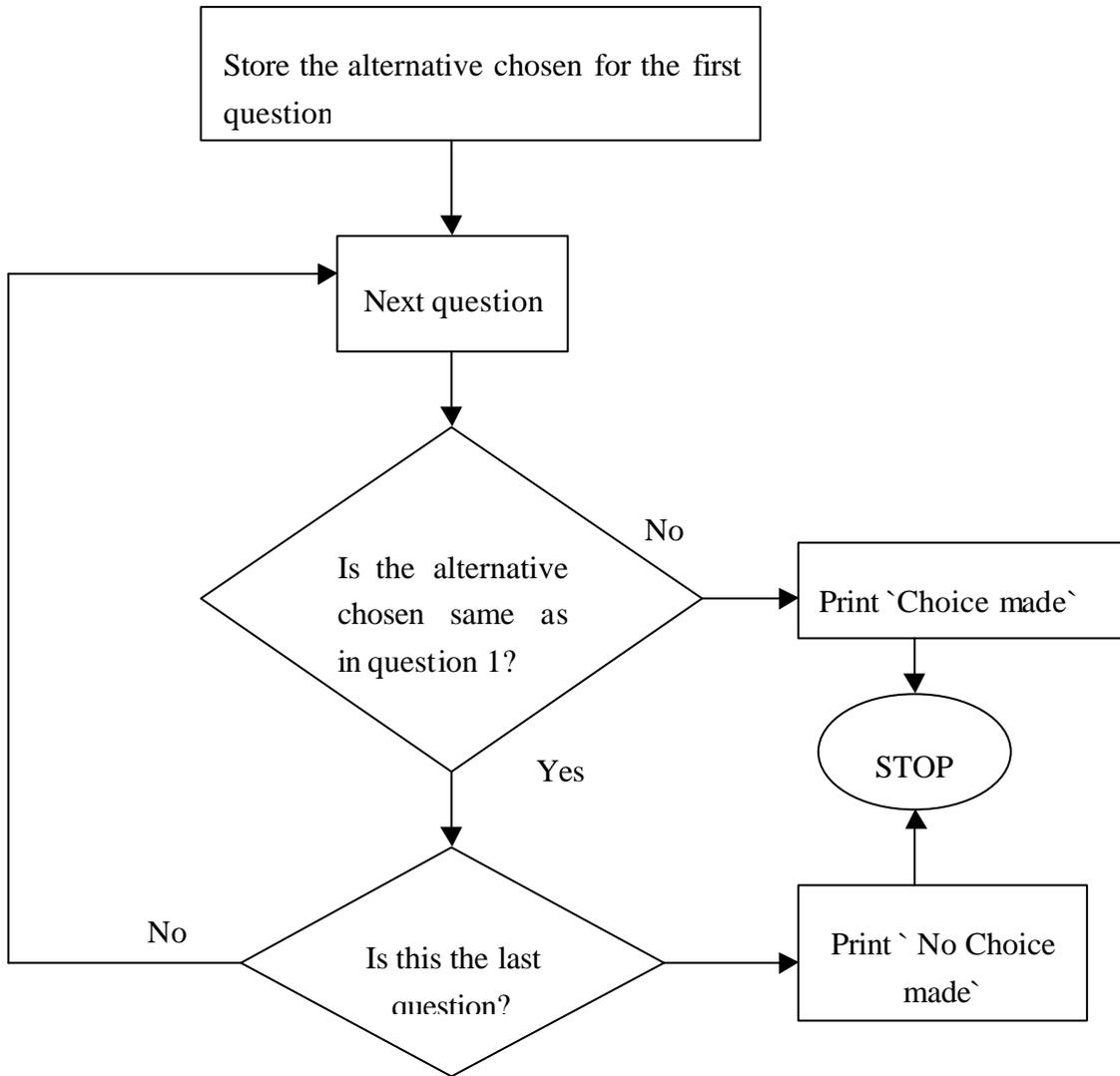Figure 2: Flowchart for the test on choice making

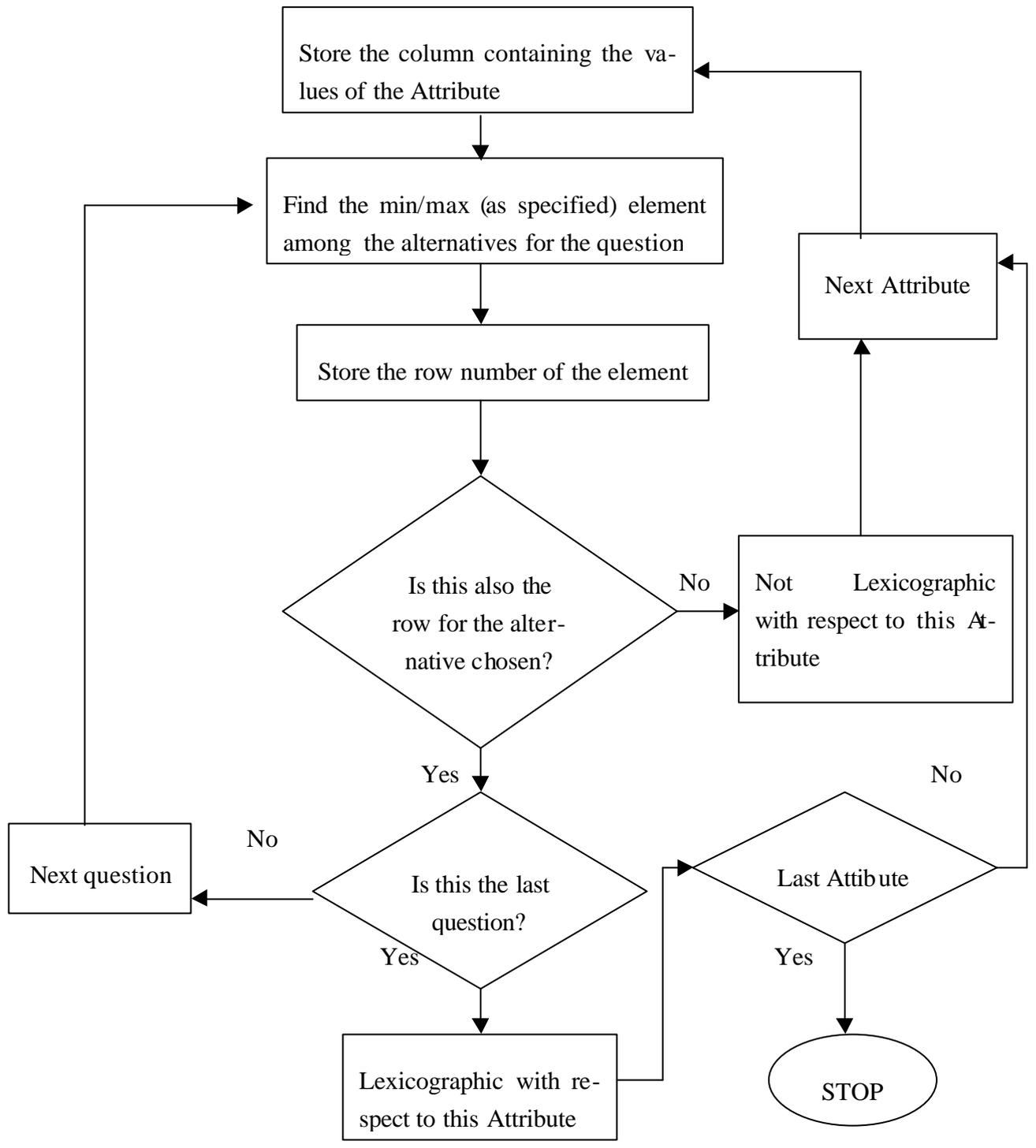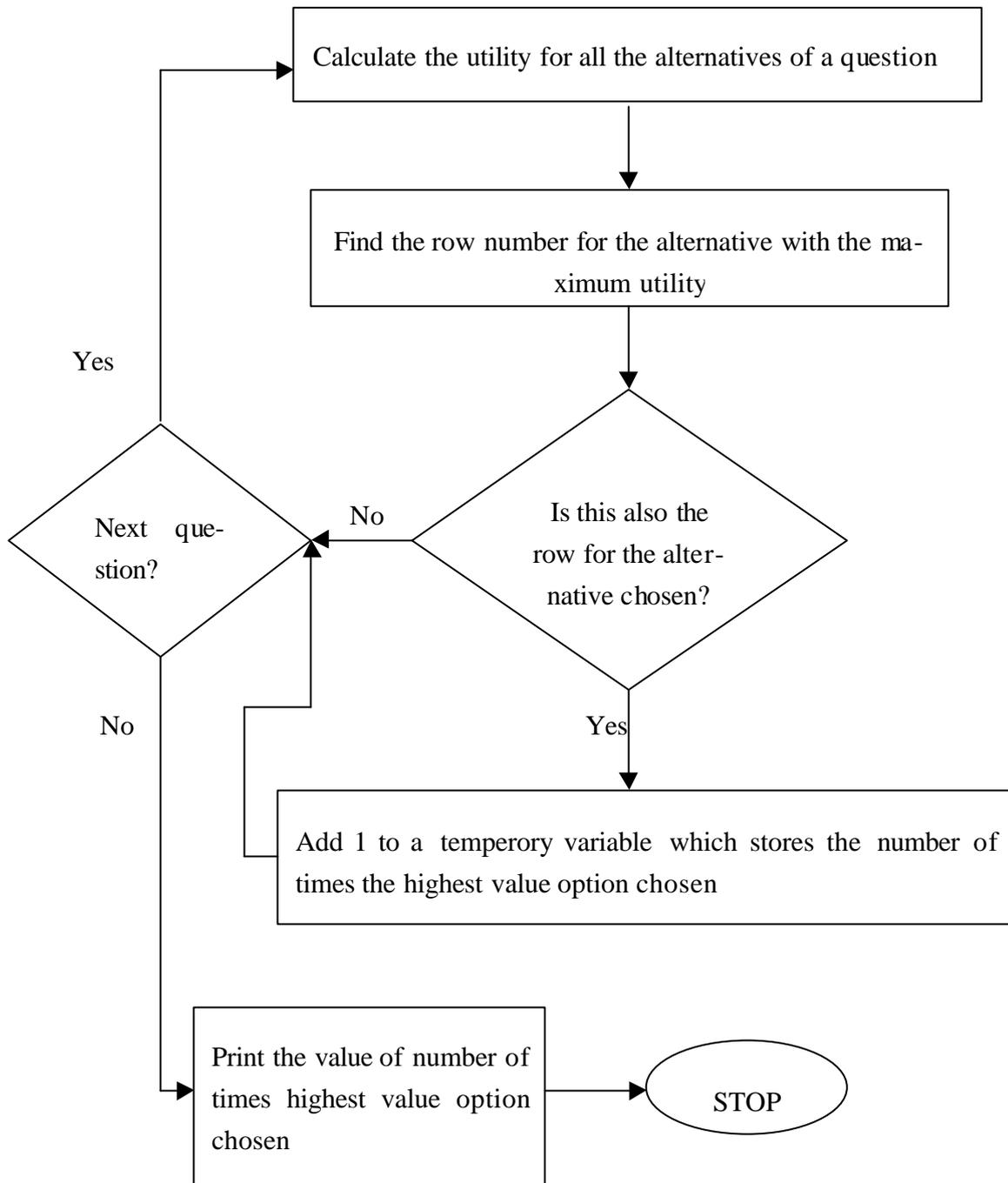Figure 3: Flowchart for lexicographic test

Figure 4: Flowchart for consistency check

_____

Calculate the utility for all the alternatives of a question

Find the row number for the alternative with the maximum utility

Is this also the row for the alternative chosen?

Next question?

Yes

No

No

Yes

Add 1 to a temperory variable which stores the number of times the highest value option chosen

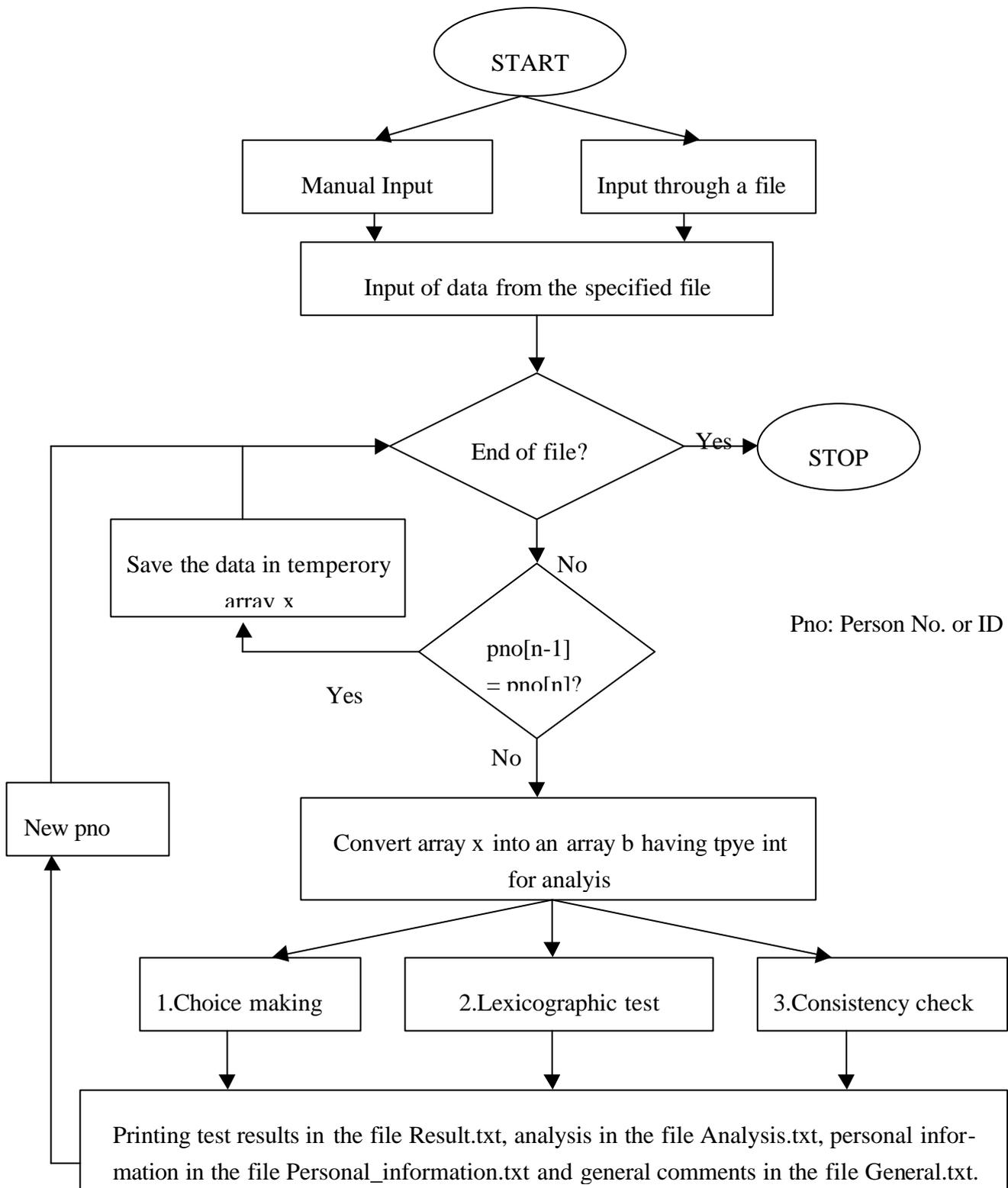Print the value of number of times highest value option chosen

STOP

_____

4   Algorithm

The first and foremost thing in analysing the SP data is to separate the data of an individual from rest, which can be used for further analysis. The current algorithm identifies an individual from his/her person ID, which user defines as one of the variables in the beginning. In case no person ID is given, the number of questions asked per person is used to separate the data for different respondents.

The step-by-step working of the algorithm can be easily understood from the flowchart for the algorithm (figure 5).

Figure 5: Algorithm

START

Manual Input

Input through a file

Input of data from the specified file

End of file?

Yes

STOP

No

Pno: Person No. or ID

Save the data in temperory array x

pno[n-1] = pno[n]?

Yes

No

New pno

Convert array x into an array b having tpye int for analyis

1.Choice making

2.Lexicographic test

3.Consistency check

Printing test results in the file Result.txt, analysis in the file Analysis.txt, personal information in the file Personal_information.txt and general comments in the file General.txt.

5   **Using the software**

The user starts the by running the file *Analyse.exe*. After the file has been started, a series of self explanatory questionnaire follow. After the user has entered the required information, the software automatically creates the output files, *General.txt, Personal_information.txt, Result.txt, Analysis.txt and ForExcel.txt*.

For the input of data the user has two options, one is to enter all details one by one as asked by the software, or there is an option for specifying an input file (in the prescribed format) that can directly be entered as the input. The format for the input file is given in *input_format.txt* file. The input file and the file it calls should be in the same directory from which the *Analyse.exe* file is run. The advantage of making an input file rather than entering each variable independently is that the user can analyse as many data sets with only minor changes in the input file.

The input consists of number of variables, number of attributes, starting column for the first attribute, name of variables, type of variables, type of attribute (max or min type), name of input data file, values for the parameters of the utility function and/or number of steps the user wants for consistency check. All the files input must in text format (.txt). Since few of the variable names are protected, with variable names user also needs to enter the type of variable. For example, a variable symbolising person's ID, has a protected name person_num. List of various variable types can be found in the file *variable_names.txt*. Similarly, with the attribute name the user need to enter its type. This is for the lexicographic test: some attributes need to assume maximum value whereas others need to take minimum value. As an example, an user making a lexicographic choice with respect to price, will always choose the alternative with the minimum price. Attributes taking maximum value belong to type 1 and the others to type 0. The user needs to input either 0 or 1 (as applicable), after attribute name and variable type, to define its type. User also needs to enter the number of steps he/she wants for the consistency check. The steps are the number of values to be checked for the parameters for utility function between the maximum and the minimum value. If only one set of coefficients are given, routine creates a range of +5% to –5% of the parameters. If the number of steps is not given the routine takes a default value of 2.

After all the data has been inputted, the routines run the data file and automatically creates a set of output files in the same directory. The output files that are created are *General.txt, Per-*

sonal_information.txt, Result.txt, Analysis.txt and ForExcel.txt. Description of the above files can be found in the next section .

For proper functioning of the software, the input file should match the format in the file *input_format.txt.*

A special file *ForExcel.txt* is created for qualitative analysis of the data. The file needs to be directly opened in Microsoft Excel, and by choosing any two rows, a Bar-graph can be generated.

6   Output Files

After the execution of the program, the following files are generated:

(i)   **General.txt:** It contains the general information, with all the neccessary information about the raw data. It has the information entered by the user such as the variable names, their number, the attributes, number of individuals interviewed etc. It is useful for the user to check the correctness of the information entered.

(ii)   **Personal_information.txt:**   This file contains the information about the individuals, their Person ID, age, sex etc.

(iii)   **Result.txt:** It contains the result for the three tests.

(iv)   **Analysis.txt:** The analyis file contains the qualitative analyis of the results. It gives the relationship between the social demographic factors and the results. For example, it gives the relationship between the age distribution and the results.

(v)   **ForExcel.txt :** This file is meant for analysis of results though graphs. When exported to Microsoft Excel, it open directly in a format in which can be directly used to make bar-graphs for qualitative analysis. The file should be opened without any changes. To make a graph, choose the two rows (one "Title line" and the second with values) and select the option for making a graph. The file is formatted such that the different concerned lines to make a graph are separated by a blank line.

(vi)   **Data.txt:** It is a temporary file for the program. It contains the original data file converted into integer form. The original file may contain some "special characters" such as "-" etc. which may cause problems during the processing of the software, so they need to be converted to integers.

7   Assumptions and Limitations

The set of routines for the quality check has the following assumptions and limitations.

- All the input files should be in .txt (text file) format.

- Some of the variable names are protected. For this purpose, the user has to enter "type" of the variable along with the variable name.

- Though the number of questions asked per person may vary , number of alternatives for all the question has to be same.

- All the attributes must occur consecutively in the data set.

- The file containing the raw data cannot have special characters (non-integer) other than "." and "-".

8   Literature

Louviere, J.J., D.A. Hensher and J.D. Swait (2000) Stated Choice Methods: Analysis and Application, Cambridge University Press, Cambridge.

Die *Arbeitsberichte Verkehrs- und Raumplanung* dienen der schnellen Verbreitung der Ergebnisse der Arbeit der Mitarbeitenden und Gäste des Instituts. Die Verantwortung für Inhalt und Gestaltung liegt alleine bei den Autor/innen.

The *Working Papers Traffic and Spatial Planning* are intended for the quick dissemination of the results of the members and guests of the Institute. Their content is the sole responsibility of the authors.

Eine vollständige Liste der Berichte kann vom Institut angefordert werden:

A complete catalogue of the papers can be obtained from:

IVT ETHZ
ETH Hönggerberg (HIL)
CH - 8093 Zürich

Telephon: +41 1 633 31 05
Telefax: +41 1 633 10 57
E-Mail: sekretariat@ivt.baug.ethz.ch
WWW: www.ivt.baug.ethz.ch

Der Katalog kann auch abgerufen werden von:

The catalogue can also be obtained from:

http://www.ivt.baug.ethz.ch/veroeffent_arbeitsbericht.html