

# Reflection on Data Error Identification Methods for Field Survey Data on Commuter Train Passenger Travel Behavior

Hitapriya Suprayitno<sup>1,a)</sup> & Ria Asih Aryani Soemitro<sup>1,b)</sup>

<sup>1)</sup>*Civil Engineering Department, Institut Teknologi Sepuluh Nopember (ITS), Surabaya.*

Correspondance : <sup>a)</sup>suprayitno.hita@gmail.com & <sup>b)</sup>ria@ce.its.ac.id

## ABSTRACT

Good understanding of Demand Behavior is important for Infrastructure and Facility Planning. Therefore, Field Survey for Travel Behavior Characteristics on Commuter Train Passenger is very important. The data collected and processed must be correct data. Meanwhile errors could easily happen in data collection and tabulation. How is the method to identify the data error. Experiment indicates several following methods : using spread-sheet software is strongly suggested for easiness to develop the whole process, establish numerical data tabulation for data error identification easiness, establish code for data back-tracked, develop field survey data table, develop data error possibilities tabel, develop a error logical tests, develop spreadsheet logical test function, do error identification calculation.

**Keywords :** infrastructure and facility asset management, facility planning, facility demand, commuter train, passenger travel behavior, field survey data error identification .

## INTRODUCTION

One important part of Infrastructure & Facility Asset Management is a good understanding of the Infrastructure & Facility Function and its Demand (Soemitro & Suprayitno 2018; Suprayitno & Soemitro 2018). For Commuter Train as Transportation Facility, its Main Function is to flow the Passenger from station A to station B, with enough capacity, fluently, safely, comfortly, affordably. Therefore, Commuter Train Passenger's Travel Behavior Characteristics is important to be well understood.

A good understanding of Travel Behavior Characteristics are very useful for : defining the train station's influence area, defining the access and egress mode and their distances, defining the parking capacity needed, defining the public transport feeder needed, defining facilities at station, and other needs (Pratiwi & Suprayitno 2016; Suprayitno & Upa 2018; Susanti, Soemitro & Suprayitno 2017; Susanti, Soemitro & Suprayitno 2017a; Susanti, Soemitro & Suprayitno 2018).

The Travel Behavior Characteristics are gotten from Field Survey. Correct and accurate Travel Behavior Picture can only be gotten from complete, correct, and enough number of sample. Meanwhile, personal experiences shows that Data Errors can easily occurs. Therefore, "How the Data Error can be identified and corrected ?", is an important question to be addressed in Data Collecting and Processing.

Travel Behavior Surveys in Australia, especially in Melbourne were done by using face-to-face interview or self completion drop-off questionnaire. Data collection by telephone, by internet and by GPS done in Toronto, Chicago, Germany were observed and evaluated. Among those three techniques, there is no perfect one, there advantages and

disadvantages with regard to representativeness, response rates, data accuracy and costs. Sampling Method generally used in Melbourne can be considered as well enough, but response rates can potentially be improved by using mix of methods between internet and telephone interview for data collection to reach different demographic groups (Inbakaran & Kroen 2011). Meanwhile, in Indonesia direct face-to-face interview is still the most used, and it seems still the most appropriate in Indonesia.

Research on Data Error Identification has been developed on various fields. Among others, it can be found concerning Language Grammatical Error, Earth Science Model, Laboratory Data Capture Apparatus, Distributor Pattern in Business, and Ergonomics Science. Even the Theory, Method and Validation, on Error Identification, has been developed also (Baber & Stanton 2002; Baker 2017; Cisco 2013; Kohmar 2016; Wang et al 2002; Schmaltz et al 2017).

This paper present a Reflection on Field Survey Data Errors Identification Procedures on Commuter Train Passenger Travel Behavior.

## RESEARCH METHOD

The research was executed by following these steps : statement of the background, defining the objective, executed related reference review, method development, and finally ended by a conclusion.

The method was developed with assumption that the Field Survey Data were tabulated on an Spreadsheet File, which the most used for micro computer is the Microsoft Excell.

Afterward, the method was developed following these steps : field survey procedure, data quality concept, data collected, probability of data correctness error, data correctness checking method, data error identification method.

## LITERATURE REVIEW

Two basic literature review on Travel Behaviour and Data Error Identification are presented below.

### Travel Behavior Research on Urban Public Mass Transport Passenger

Travel Behavior survey on Commuter Train Passenger and Urban Bus Passenger in several different train-lines and bus-lines are already executed. Several of them are mentioned below (Silaen, Nasution & Suwanto 2018; Suprayitno et al 2006; Suprayitno & Upa 2016; Upa, Suprayitno & Ryansyah 2018; Suprayitno & Ryansyah 2018; Suprayitno, Saraswati & Ratnawati 2018; Susanti, Soemitro & Suprayitno 2017; Susanti, Soemitro & Suprayitno 2017a; Susanti, Soemitro & Suprayitno 2018; Susanti, Soemitro & Suprayitno 2018a).

- Passenger Travel Behavior of Economy Class Train in Gerbangkertosusila Region.
- Passenger Travel Behavior of Susi Commuter Train in Surabaya.
- Passenger Travel Behavior of Mamminasata BRT Lyne 2 and Lyne 3, in Makassar.
- Passenger Travel Behavior of Trans Koetaradja Bus Lyne 1, in Banda Aceh.
- Comparaison and Synthese of Travel Behavior between Mamminasata BRT Passenger and Trans Koetaradja BRT Passenger.
- Comparaison and Synthese of Travel Behavior between Commuter Train Passenger and BRT Pasenger.
- Public Preference fo New BRT Network Plan Trans Mebidang in Medan.

### Data Error Identification

Data Error can happen in all aspects of life for all phenomena. Several of them are, for example : error on language grammar, on laboratory measurement apparatus, on bussiness, on modeling, on ergonomics science, on product design, etc (Baber & Stanton 2002; Baker 2017; Cisco 2013; Kohmar 2016; Wang et al 2002; Schmaltz et al 2017). As illustrations, certain of them is presented below.

Laboratory Apparatus Measurement gross error identification method has been developed by using a theory which is called Grey System Theory (Wang et al 2002). In Automated Evaluation of Scientific Writing (AESW), the method commonly used for grammar error identification is the attention-based encoder-decoder model. This method can be used for correction generation instead only error identification. A new method was developed, character-based encoder-decoder, which is proofed to be better for AESW (Schmaltz et al 2017). The Product Design field has developed Task Analysis for Error Identification (TAEI). It based on communication of user and products. This can represents a form of problem solving. Eeach state of dialogue offer the user potential of action. Therefore analysing action can be used for design or ergonomic error identification (Baber & Stanton 2002).

It can be noted easily, that in all cases presented above, the error identification task is different from the correction generation task. Correction must be based on Error Identification result.

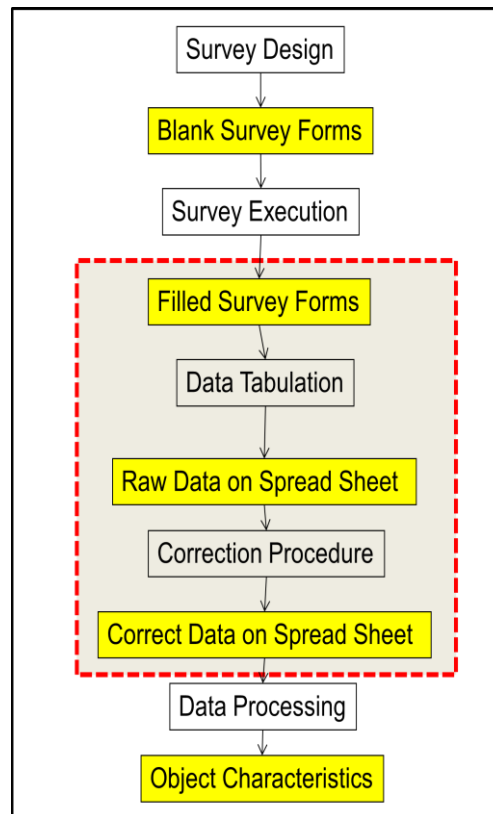
## **METHOD DEVELOPMENT ON DATA ERROR IDENTIFICATION**

### **Method Development Step**

The Method development was done by following these steps : formulating Field Survey Procedure, formulating Data Quality Concept, thinking Example of Typical Data Collected, formulating Data Error Probability, formulating Error Identification Procedure, and ended by a Method Trial.

### **Field Survey Procedure**

The Field Survey Procedure generally folow the following steps. It is started by survey design, survey execution, data tabulation, data correction procedure, and finished by data processing. These steps are presented in Figure 1 below.



**Figure 1.** Field Survey Procedure

### Field Survey Data Quality Concept

Field Survey is executed to collect primary data on certain characteristics of an object. The data collected must be able to well picture the characteristics in question. Therefore, the data collected has certain quality parameter as follows :

- **Data Completeness :**  
Every data needed, must be successfully collected.  
This problem is related on Survey Questionnaire Design and Survey Execution
- **Data Correctness :**  
All data tabulated and proceed on spread sheet must be the correct data.  
Suppose, the Survey Questionnaire is correct, this could be be a problem on Field Data Collection and Data Tabulation. Data Error Identification procedure must be established.
- **Sample Size :**  
The whole data must be able to accurately picture the Surveyed Characteristics.  
Even if all of the Tabulated Data are all correct, the accuracy still cannot be guaranteed, unless enough sample size can be collected. Too small sample with correct data can produce different characteristics from the reality.

Accuracy Quality depend absolutely on Data Completeness, Data Correctness, and Sample Size. This paper discuss only the Error Identification Method..

### Structure of Data Table or Basis Data

In general Data Table or Basis Data has a structure as explained afterward. Data of a Respondent is written in one line on the Data Table. Thus 100 Respondents will produce 100 lines of Data Table. Each Repondent characteristics is written in each defined column. The

columns are always started by a a column indicating ID Number, follow by columns to fill the Repondents Characteristics. In Data Base System, technically, each column is called Field and each line is called Record (Schurmann 2006).

**Table 1** Basis Data Structure

	<i>field 1</i>	<i>field 2</i>	...	...	<i>field n</i>	
	<b>No</b>	<b>ID</b>	<b>Data 1</b>	<b>Data 2</b>	<b>Data 3</b>	<b>Data 4</b>
record 1	1					
record 2	2					
...	...					
...	...					
record n	n					

**Example of Typical Collected Data**

As an example, a Typical Data collected, on Commuter Train Passenger Travel Behavior, are presented below. Numerical data, such as : age, travel distance, travel time, vehicle possession, and others should be collected as numerical data. Other data should be posed in questionnaire as a multiple choice data. Example of Typical Data Collected is presented in Figure 2 below.

**Table 2.** Example of Typical Collected Data

<b>Train</b>	<b>Trip Maker</b>	<b>Trip</b>
trip code number	name	purpose
direction		hour
departing time	age	access trip
etc	gender	origin
	education	access station
	profession	access distance
	vehicle possession	access mode
	etc	egress trip
		destination
		egress station
		egress distance
		egress mode
		etc

**Data Error Possibility**

Data Collected must be as correct and as accurate as possible. In general, even if the Field Survey had been designed at maximum corretness and accuracy, Risk of Data Error is still there. Such error has been experienced by the author. Therefore, Data Error Identification Method need to be established. The Data Error can be classified into two categories :

- Error on data extraction from the passenger.
- Error on data entry to the Data Table from the Questionnaire Form

On the case of Commuter Train Passenger Travel Behavior Field Survey, the data error can be classified into 4 categories : Double Counting, Trip Maker Characsterstics, Trip Characteristics, and correlation between Trip Maker ~ Trip Characteristics. Example of Error Possibilities for each these catagories are presented in the following Table 3.

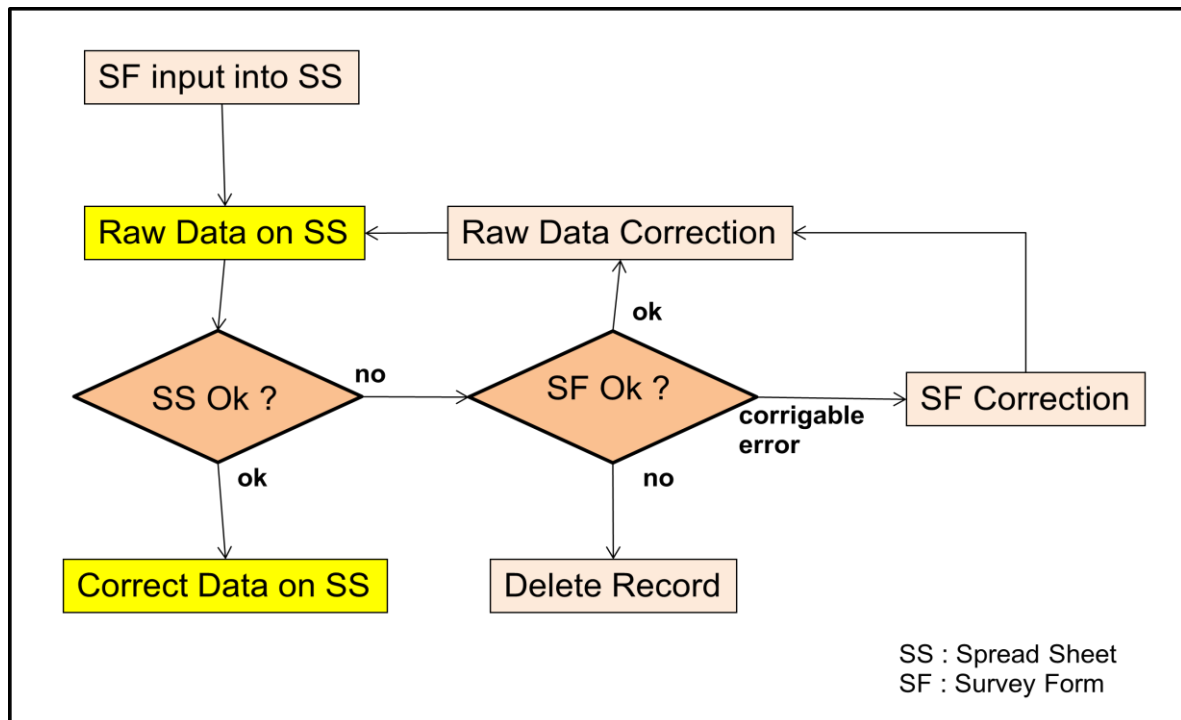
**Table 3** Example of Data Errors Possibilities

<b>Group</b>	<b>Error</b>
Double Counting	recording the same person more than once
Trip Maker Characteristics	age ~ education
	income ~ mode possession etc
Trip Characteristics	trip direction ~ train direction
	access distance ~ egress distances not logic
	walking distance > 5 km etc
Trip Maker ~ Trip Characteristics	age ~ mode used
	etc

### Field Survey Data Error Correction Procedure

Field Survey procedure is presented in Figure 1 above. There is a step which is called Correction Procedure. Data Error Identification is part of Data Error Correction Procedure. Therefore to develop Error Identification Procedure must be based on Data Error Correction Procedure.

The Data Correction Procedure has as input Raw Data Table and has as output of Correct Data Table, the two written on spreadsheet software. Based on Raw Data on Spreadsheet, the Data Error Correction Procedure is started to Check whether the each Raw Data on Spreadsheet is correct or not. Checks are done record by record. If all data on a Record is all correct, the Record can be put into Correct Data on Spreadsheet. On the other hand, if a Record has a certain Data Error, this Record has to be confronted to Check whether Data on the Survey Form is Correct or not. In this Check there are three possibilities. First, the Survey Form Data is actually correct, only the inputting data which is wrong, then the Record must be corrected on Raw Data Correction step, the process is continued by having Raw Data on Spreadsheet and then Check whether the whether each Data on Spreadsheet is correct or not. Second, the Survey Form Data is not correct, but in some way can be corrected, then the procedure is continued by Survey Form Correction, which is followed by Raw Data Correction. Third, the Survey Form Data is totally in-correctable error, the Record must be dump or the Record must be deleted. The Data Correction Procedure is presented in Figure 3 below.

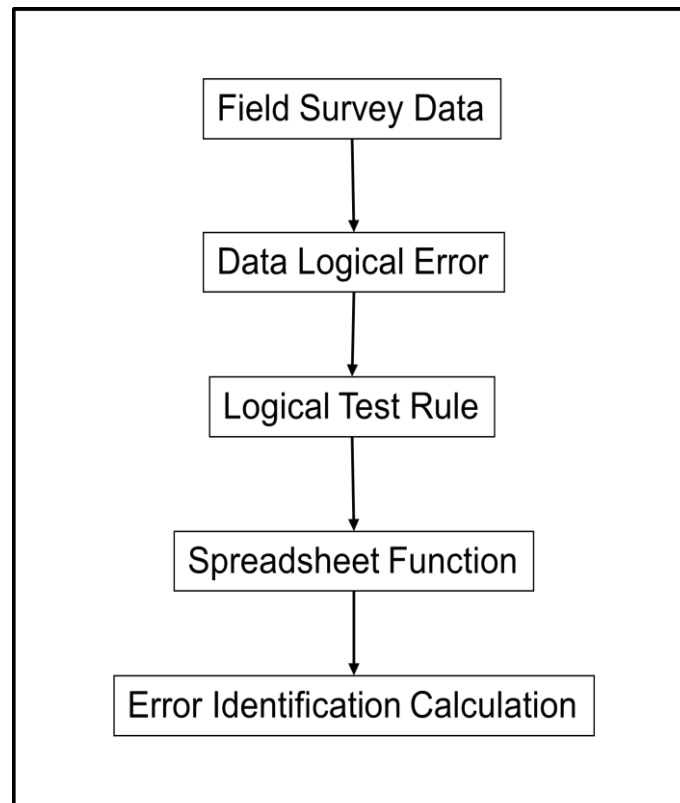


**Figure 2.** Data Correctness Checking Procedure

**Data Error Identification Method**

All of Possible Error must be able to be identified. How is the method to identify such error. All data must be tabulated in a spreadsheet software by using numeric code as much as possible to ease the Data Error Identification program.

The Data Error Identification procedure was developed to follow the following steps. After establishing a Field Survey Data Table, the first step is to develop Data Logical Error, followed by developing Logical Test Rule. Now, based on Logical Test Rule, Spreadsheet Logical Test Function can be written, and ended by executing Error Identification Calculation. Example of the whole Error Identification Process is presented in sub-chapter Method Trial below. The steps are presented in Figure 3 below.



**Figure 3.** Data Error Identification Step

### **Method Trial**

#### The Method Trial

The Method Trial was executed by following these steps : experiment case (data code and data collected), data logical error, logical test rule, spread sheet function, error identification calculation (identification of double counting error, identification of trip maker characteristics error, identification of trip error, identification of trip maker ~ trip correlation error).

#### Experiment Case

Virtual Experiment Case was established and taken to do the Method Trial. The case is about surveying Passenger Travel Behavior Data travelling in a Commuter Train with 6 stations, serving station 1 to station 6. The Case is very simplified. The Trip Maker Data is limited only for name, age and accomplished education. The Trip Data are limited only on the Access and Egress Trip, with each denoting the zone, station, distance and mode.

All Data are presented in Data Code except for name. The Data Code are presented in Table 4 below, while the Field Survey Data are presented in Table 5 below.



**Table 4.** Data Code

No	Education		Origin		Origin Station		Mode	
1	1	Primary School	1	Zone 1	1	First Station	1	Walk
2	2	Middle School	2	Zone 2	2	Second Station	2	Riding Motorcycle
3	3	High School	3	Zone 3	3	Third Station	3	Driving Car
4	4	Higher Education	4	Zone 4	4	Fourth Station		
5			5	Zone 5	5	Fifth Station		
6			6	Zone 6	6	Last Station		

**Table 5.** Field Survey Data

No	ID	Trip Maker			Trip							
		Name	Age	Edu	Access				Egress			
					Orig.	A.Sta.	A.Dist.	A.Mode	Dest.	E.Sta.	E.Dist.	E.Mode
							km				km	
1	1	A	20	1	5	5	3	2	6	6	1	1
2	2	B	15	4	1	1	1	1	6	6	1	1
3	3	C	32	4	2	1	5	3	5	5	1	1
4	4	D	32	2	1	2	1	1	5	5	1	1
5	5	E	32	2	3	3	5	2	4	4	1	1
6	6	F	5	0	3	3	5	2	4	3	1	1
7	7	G	20	1	5	5	3	2	6	6	1	1

Data Logical Error and Logical Test Rule

After the Field Survey Data are tabulated, for preparing the Error Identification Calculation, two steps has to be executed : developing the Data Logical Error and the Logical Test Rule. These two are presented below.

*Data Logical Error*

Data Logical Error step is to Identify Different Logical Error Existence in related data value. One of them, for example, is double counting error : a certain passenger is counted more than once. Another example is the Train heading to the north but the trip heading to the south.

In general the Data Logical Error can be classified into : double counting, trip maker characteristics data error, trip characteristics data error and data correlation between trip maker and trip characteristics data. An example of Data Logical Error is presented in Table 5 below.

**Table 5.** Example of Data Logical Error

Error Group	Error	Logical Error
Duble Counting	Double Counting	same trip maker characteristics
Trip Maker	Age~Education	education is not in correspondance with age
Trip	Trip Direction	trip direction against train direction
	Access Mode~Distance	access mode & access distance not logical
Trip Maker ~ Trip	Age~Mode	mode is not in correspondance with age
	Age~Distance~Mode	age is not in correspondance with mode and distance

*Logical Test Rule*

The Logical Error tabulated above has to be formulated in Logical Test Rule to be able to be programmed. One Logical Error for each Error Group is taken, for which the Logical Test Rule is formulated. The Logical Test Rule is presented in Table 6 below.

**Table 6.** Example of Logical Test Rule

No	Error	Logical Test																		
1	Double Counting	2 records with exactly the same field value																		
2	Correlation between Age~Education	the age is not correspondance to the education level																		
		<table border="1"> <thead> <tr> <th><u>Education</u></th> <th><u>Code</u></th> <th><u>Age</u></th> </tr> </thead> <tbody> <tr> <td>no education</td> <td>0</td> <td>any</td> </tr> <tr> <td>primary</td> <td>1</td> <td>&gt;13</td> </tr> <tr> <td>middle</td> <td>2</td> <td>&gt;16</td> </tr> <tr> <td>high</td> <td>3</td> <td>&gt;19</td> </tr> <tr> <td>higher education</td> <td>4</td> <td>&gt;23</td> </tr> </tbody> </table>	<u>Education</u>	<u>Code</u>	<u>Age</u>	no education	0	any	primary	1	>13	middle	2	>16	high	3	>19	higher education	4	>23
		<u>Education</u>	<u>Code</u>	<u>Age</u>																
		no education	0	any																
		primary	1	>13																
middle	2	>16																		
high	3	>19																		
higher education	4	>23																		
3	Trip Direction	Trip Direction is not correspondance to Train Direction																		
		Train Trip toward Small Station Code																		
		Origin Station Code < Destination Code																		
		Train Trip toward Big Station Code vice versa																		
4	Correlation between Age~Mode	the age is not correspondant to the mode used																		
		<table border="1"> <thead> <tr> <th><u>Mode</u></th> <th><u>Code</u></th> <th><u>Age</u></th> </tr> </thead> <tbody> <tr> <td>Walk</td> <td>1</td> <td>any</td> </tr> <tr> <td>Riding Motorcycle</td> <td>2</td> <td>&gt;17</td> </tr> <tr> <td>Driving Car</td> <td>3</td> <td>&gt;19</td> </tr> </tbody> </table>	<u>Mode</u>	<u>Code</u>	<u>Age</u>	Walk	1	any	Riding Motorcycle	2	>17	Driving Car	3	>19						
		<u>Mode</u>	<u>Code</u>	<u>Age</u>																
		Walk	1	any																
Riding Motorcycle	2	>17																		
Driving Car	3	>19																		

General Microsoft Excell Function for Error Identification

Data are tabulated in Spreadsheet to facilitate the Data Processing. Based on Data Error Logical test, the data error can be detected easily by using Spreadsheet Logical Test Function : IF. The syntax of IF Function is presented below.

$$IF = (\text{logical\_test}; \text{value\_if\_right}; \text{value\_if\_wrong}).$$

Identification of Double Counting Error

By using Logical Test mentioned in Table 6 on Double Counting Data Error Logique, a Microsoft Execell Logical Test Function was written as below. The Excell Function is written in Record-Wise test. Therefore, the complete test must be done (N-1) times to check the whole possibilities of double counting existence, where N is the number of sample.

$$=IF(AND(\text{age}=\text{age}; \text{edu}=\text{edu}; \text{orig}=\text{orig}; \text{a.sta}=\text{a.sta}; \text{a.dist}=\text{a.dist}; \text{a.mode}=\text{a.mode}; \text{dest}=\text{dest}; \text{e.sta}=\text{e.sta}; \text{e.dist}=\text{e.dist}; \text{e.mode}=\text{e.mode}); "X"; "ok").$$

$$=IF(AND(D6=D$5; E6=E$5; F6=F$5; G6=G$5; H6=H$5; I6=I$5; J6=J$5; K6=K$5; L6=L$5; M6=M$5); "X"; "ok").$$

The checking procedure found that there is a Double Counting Error for Respondent no 1 and Respondent no 7. The error lies on the fact that the data recorded for these two individuals are all exactly the same. A probability of double counting for these 2 records is very strong. It must be checked wether a double counting has been done or not. The Double Counting Data Error calculation is presented in Table 6 below.

**Table 6.** Error Identification – Double Counting

No	ID	Trip Maker			Trip								Error Check
		Name	Age	Edu	Access				Egress				
					Orig.	A.Sta.	A.Dist.	A.Mod	Dest.	E.Sta.	E.Dist.	E.Mod	
			<i>year</i>				<i>km</i>				<i>km</i>		
1	1	A	20	1	5	5	3	2	6	6	1	1	Ref.
2	2	B	20	4	1	1	1	1	6	6	1	1	ok
3	3	C	32	4	2	1	5	3	5	5	1	1	ok
4	4	D	32	2	1	2	1	1	5	5	1	1	ok
5	5	E	32	2	3	3	5	2	4	4	1	1	ok
6	6	F	5	0	3	3	5	2	4	3	1	1	ok
7	7	G	20	1	5	5	3	2	6	6	1	1	X

Identification of Trip Maker Characteristics Error : Age ~ Education

By using Logical Test mentioned in Table 6 on Correlation between Age~Education Error, a Microsoft Excell Logical Test Function was written as below. The Excell Function is written in Field-Wise test.

=IF(AND(age=0;edu<12);"ok";IF(AND(age=1;edu>13);"ok";IF(AND(age=2;edu>15);"ok";IF(AND(age=3;edu>18);"ok";IF(AND(age=4;edu>23);"ok";"X")))))

=IF(AND(E5=0;D5<12);"ok";IF(AND(E5=1;D5>13);"ok";IF(AND(E5=2;D5>15);"ok";IF(AND(E5=3;D5>18);"ok";IF(AND(E5=4;D5>23);"ok";"X")))))

The Data Error Identification calculation found that there is a certain error for Respondent no 2. The error lies in the fact that the Respondent no 2 was recorded having an age of 20 years, meanwhile he finished already his higher education. The Error Identification calculation is presented in Table 7 below.

**Table 7.** Error Identification – Trip Maker Characteristics : Age ~ Education

No	ID	Trip Maker			Trip								Error Check
		Name	Age	Edu	Access				Egress				
					Orig.	A.Sta.	A.Dist.	A.Mod	Dest.	E.Sta.	E.Dist.	Mode	
			<i>year</i>				<i>km</i>				<i>km</i>		
1	1	A	20	1	5	5	3	2	6	6	1	1	ok
2	2	B	20	4	1	1	1	1	6	6	1	1	X
3	3	C	32	4	2	1	5	3	5	5	1	1	ok
4	4	D	32	2	1	2	1	1	5	5	1	1	ok
5	5	E	32	2	3	3	5	2	4	4	1	1	ok
6	6	F	5	0	3	3	5	2	4	3	1	1	ok
7	7	G	20	1	5	5	3	2	6	6	1	1	ok

Identification of Trip Characteristics Error : Trip Direction ~ Train Direction

By using Logical Test mentioned in Table 6 on Trip Direction Error, a Microsoft Excell Logical Test Function was written as below. The Excell Function is written in Field-Wise test.

=IF(OR(a.sta=e.sta;a.sta>e.sta);"X";"ok")

$$=IF(OR(G5=K5;G5>K5);"X";"ok")$$

The Error Identification found that there is a certain error for Respondent no 6. The error lies in the fact that the Respondent no 6 was recorded boarding and alighting on the same station. The Error Identification Calculation is presented in Table 8 below.

**Table 8.** Error Identification – Trip Characteristics : Trip Direction

No	ID	Trip Maker			Trip								Error Check
		Name	Age	Edu	Access				Egress				
					Orig.	A.Sta.	A.Dist.	A.Mod	Dest.	E.Sta.	E.Dist.	E.Mod	
	year		km			km							
1	1	A	20	1	5	5	3	2	6	6	1	1	ok
2	2	B	20	4	1	1	1	1	6	6	1	1	ok
3	3	C	32	4	2	1	5	3	5	5	1	1	ok
4	4	D	32	2	1	2	1	1	5	5	1	1	ok
5	5	E	32	2	3	3	5	2	4	4	1	1	ok
6	6	F	5	0	3	3	5	2	4	3	1	1	X
7	7	G	20	1	5	5	3	2	6	6	1	1	ok

Identification of Trip Maker ~ Trip Characteristic Correlation Error : Age ~ Mode

By using Logical Test mentioned in Table 6 on Correlation between Age~Mode Error, a Microsoft Excell Logical Test Function was written as below. The Excell Function is written in Field-Wise test.

$$=IF(AND(mode=2;age<18);"X";if(AND(mode=3;age<20);"X";"ok"))$$

$$=IF(AND(I5=2;D5<18);"X";IF(AND(I5=3;D5<20);"X";"ok"))$$

The Error Identification found that there is a certain error for respondent no 6. The error lies in the fact that a child of 5 years old riding a motorcycle to to the origin station. The Error Identification Calculation is presented in Table 9 below.

**Table 9.** Error Identification – Trip Maker ~ Trip Characteristics : Age ~ Mode

No	ID	Trip Maker			Trip								Error Check
		Name	Age	Edu	Access				Egress				
					Orig.	A.Sta.	A.Dist.	A.Mode	Dest.	E.Sta.	E.Dist.	E.Mode	
	year		km			km							
1	1	A	20	1	5	5	3	2	6	6	1	1	ok
2	2	B	15	4	1	1	1	1	6	6	1	1	ok
3	3	C	32	4	2	1	5	3	5	5	1	1	ok
4	4	D	32	2	1	2	1	1	5	5	1	1	ok
5	5	E	32	2	3	3	5	2	4	4	1	1	ok
6	6	F	5	0	3	3	5	2	4	3	1	1	X
7	7	G	20	1	5	5	3	2	6	6	1	1	ok

**CONCLUSIONS**

The research has been successfully finished. Data Error Identification Method has been established. It must be noted that the data on this experiment are very much simplified. The objective is merely to establish a good method. Real case need a more accurate and complicated rules. Several main conclusions are written as follows :

- Data Error Identification Method, in principle, consists of the following steps : Field Survey Data tabulation, Data Logical Error identification, Error Logical Test Rule development, Spreadsheet Function programming, Data Error Identification calculation.
- For Commuter Train Passenger Travel Behavior case, Data Error can occur as double counting, on trip maker characteristics, on trip characteristics, on correlation between trip maker ~ trip characteristics.
- Data Tabulation must be done on spreadsheet, to facilitate data processing.
- Spreadsheet IF Logical Test Function can be used for Error Identification Calculation.
- Be careful, there are two types of IF Logical Test : Record Wise and Field Wise.

For the purposes of indicating ease, it is better than this method is named. The method is named DC&CC Error Identification. DC & CC stand for Double Counting and Characteristics Conformity Error.

Further researches are still needed to be done, among others are on : trying the developed method on real case, conducting experiment on minimum sample size, conducting experiment on minimum sample size calculation for various proportion number and proportion cases, developing the whole data correction method.

**Notes.** This research is part of the Main Research on Infrastructure Demand Modeling, started with a Transport Demand Modelling Accuracy Error Identification Method. Research Series on Trip Production Modelling and Trip Length Distribution Modelling Accuracy are still being undergone.

## REFERENCES

- Baber, C. & Stanton, N.A. (2002). "Task Analysis for Error Identification : Theory, Method and Validation". *Theoretical Issues in Ergonomics Science*, April 2002.
- Baker, A. (2017). "Quality Assurance and Error Identification for the Community Earth Science Model". *SC 2017 Workshop – Software Correctness for HPC Applications*. National Center for Atmospheric Research. Boulder, Colorado.
- Cisco (2013). "Error Identification and Search". *Distributor and 2-Tier Partner Job Aid*. CSCC – Cisco Service Contract Center. China.
- Inbakaran, C. & Kroen, A. (2011). "Travel Survey – Review of International Survey Methods". *Proceedings Australasian Transport Research Forum 2011, 28-30 September 2011*. Adelaide, Australia.
- Kochmar, Ekaterina (2016). "Error Detection in Content Word Combinations". *Technical Report Number 886*. Computer Laboratory. University of Cambridge. Cambridge.
- Pratiwi, J.O. & Suprayitno, H. (2016). "Calculation Method to Analyze the Concordance between Supply and Demand of Bus Public Transport for Morning Commuter Trip at Gubeng Station". *IPTEK Journal of Proceeding Series 2016 (2) 1*.
- Schmaltz, Kim, Rush & Shieber (2017). "Sentence-Level Grammatical Error Identification as Sequence to Sequence Correction". *Working Paper*. Harvard University. Cambridge, Massachusetts.
- Schurmann, C. (2006). "Introduction to Database Design". *Working Paper*. IT University of Copenhagen. Copenhagen.
- Silaen, S.K., Nasution, A.D. & Suwanto, H. (2018). "Public Preference for New Service Network Plan BRT Trans Mebidang (Route : Pancurbatu – Sambu Market Center)". *IOP Conference Series : Materials Science and Engineering 420 (2018) 012010*.
- Soemitro, R.A.A. & Suprayitno, H. (2018). "Pemikiran Awal tentang Konsep Dasar Manajemen Aset Fasilitas". *Jurnal Manajemen Aset Infrastruktur & Fasilitas, Vol. 2, Sup. 1, Juni 2018, hal. : 1-13*.

- Suprayitno, H., Irlambang, D., Pramono, Y.A. & Wibowo, A.G. (2006). “Karakteristik Penumpang dan Perjalanan pada KA Penumpang Ekonomi di Wilayah Gerbangkertosusila”. *Jurnal Teknobisnis*, Vol. 2, No. 1, Juli 2006, hal. : 41-48.
- Suprayitno, H. & Upa, V.A. (2016). “Mamminasata BRT User Trip Characteristics for the Design of Demand Modelling Method for a New BRT Line”. *IPTEK, The Journal of Technology and Science*, Vol. 27, No. 3, December 2016, pp. : 47-52.
- Suprayitno, H., Ratnasari, V. & Saraswati, N. (2018). “Behavior of Trip Length Distribution Pattern due to the Variation of Trip Length Interval Determination”. *IPTEK, Journal of Engineering*, Vol. 4, No. 1, 2018, pp. : 7-11.
- Suprayitno, H. & Ryansyah, M. (2018). “Karakteristik Pelaku dan Perilaku Perjalanan Penumpang Bis Trans Koetaradja, Koridor 1 : Keudah-Darulsalam”. *Jurnal Aplikasi Teknik Sipil 16* (2), 55-62, 2018.
- Suprayitno, H. & Soemitro, R.A.A. (2018). “Preliminary Reflexion on Basic Principle of Infrastructure Asset Management”. *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, Vol. 2, No. 1, Maret 2018, hal. : 1-9.
- Suprayitno, H. & Upa, V.A. (2017). “Special Conventional Transport Model for a New BRT Line Passenger Demand Prediction (The General Modeling Method)”. *Journal of Technology and Social Science (JTSS)*, Vol. 1, No. 3, 2017, pp. : 10-18.
- Susanti, A., Soemitro, R.A.A. & Suprayitno, H. (2017).”Identifikasi Awal Layanan Feeder di Tiap-Tiap Stasiun yang Menjadi Pemberhentian KA Penumpang di Kota Surabaya”. *Rekats 17* (1), 2017.
- Susanti, A., Soemitro, R.A.A. & Suprayitno, H. (2017a).”Wilayah Pengaruh Kereta Api Komuter terhadap Pengguna Sepeda Motor di Kota Surabaya”. *Jurnal Transportasi 17* (3), 2017.
- Susanti, A., Soemitro, R.A.A. & Suprayitno, H. (2018).”Identifikasi Awal Wilayah Pengaruh KA Komuter ditinjau dari Penggunaan Mobil di Kota Surabaya”. *Bangunan 23* (1), 2018.
- Susanti, A., Soemitro, R.A.A. & Suprayitno, H. (2018a).”Comparative Analysis on Access and Egress Distance for Semi BRT Trans Mamminasata and Commuter Train Susi”. *Presented in CITIES, Surabaya, 24-25 October 2018*.
- Upa, V.A., Suprayitno, H. & Ryansyah, M. (2018). “Perbandingan dan Sintesis Karakteristik Perilaku Perjalanan Pengguna Bis Trans Mamminasata dan Bis Trans Koetaradja”. *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, Vol. 2, No. 2, September 2018, hal : 69-81.
- Wang, Z., Gao, Y., Tse, S. & Qin, P. (2002). “Gross Measurement Error Identification Using the Grey System Theory”. *Working Paper*. Henan University of Science and Technology, Louyang, 471039, China.