

Studies on the Characteristics of Sampling for Transport Planning

H.S.Sathish

*Asst.Professor, Department of Civil Engg, B.M.S. College of Engineering
Bull Temple Road, Bangalore-560019, Karnataka, India*

Dr.H.S.Jagadeesh

*Professor, Department of Civil Engg, B.M.S. College of Engineering
Bull Temple Road, Bangalore-560019, Karnataka, India*

Dr. R.Sathya Murthy

*Professor, Department of Civil Engg, B.M.S. College of Engineering
Bull Temple Road, Bangalore-560019, Karnataka, India*

Abstract : Sampling is necessary in transport planning process because of the wide variety of data that can be collected by the Home Interview technique and the high cost involved, it is impractical and unnecessary to interview all the residents of the study area. Since travel patterns tend to be uniform in a particular zone, sampling is very much necessary. The size of the sample is usually determined on the basis of the population of the study area. The sample data must be collected at random, data should be complete and it must be statistically stable. In this study data was collected for 30 selected wards of Bangalore city. The main objective of the study was to evaluate the statistics related to sampling and to establish the confidence levels for the attributes, viz: size of the family, per capita income, distance travelled, travel time and travel cost respectively. The histograms, frequency distribution curves and Ogive curves are the tools adopted and statistical parameters are determined for the various attributes. Also the coefficient of skewness and kurtosis are determined for the various attributes. It was concluded that the statistical parameters analyzed that as the sample size increases, the standard deviation for the size of the family as attribute increases. This indicates that as the sample size increases the scatter also increases for this attribute. The standard error of estimate (e) at 95% confidence level has a decreasing trend with the increase in sample size for all the attributes considered.

Key words : Population, Sample, Transport planning, Statistical Attributes, Ogive curve, Histograms, Coefficient of Skewness, Kurtosis.

I. INTRODUCTION

General:

The metropolitan cities of India have recorded a phenomenal growth of population during the past few decades. The rapid growth of industrial and economic activities in and around these cities, coupled with inadequate employment opportunities, educational and other facilities in the towns and rural areas, have resulted in increasing migration to the urban areas.

As the cities expand to accommodate the growth in population and activities, the spatial separation between the population and employment locations increases and increasing needs for travel modes are felt. The daily needs of the residents for work, education, business, shopping and recreational travel have created enormous demands on the transport system.

Urban transport is important for productive and progressive environment. It is a means to an end, a service provided towards assisting the urban society to satisfy its needs. It has a vital role to play in the balanced growth of cities and has to be planned for the present travel demand and increasing demands.

Bangalore is the fifth largest metropolis in India and is one of the fastest growing cities in Asia. It is also the capital state of Karnataka. It is globally recognized as IT capital of India and also as a well-developed industrial city. Before Bangalore city had 100 wards, due to the increasing population it has been reorganized into 198 wards.

The Bangalore city which was 28.75 sq km in 1901 increased to 174.7 sq km in 1971 to 272 sq km in 1986 and presently it has expanded to nearly 800sq km. The city which was originally developed as a garden city has slowly transformed into an industrial and software hub of India. The establishment of Silicon Park on the outskirts of the city has converted the city and its surroundings into Silicon Valley of the country.

Objectives of the Present Study:

The main objective of the present study is to study the statistical characteristics of sampling and to establish the confidence levels for the attributes.

II. PRESENT INVESTIGATIONS

Methodology:

The location for the statistical analysis is performed by considering 30 wards of Bangalore. The selection of Wards is based on the socio economic parameters and land use pattern. The wards selected for the present study is shown in Table 1.1. Based on the literature survey the sample size 1.5% was selected for the investigation of each attribute.

Table 1.1 Names of wards of Bangalore City selected for the study

Sl. No	Name of the Ward	Sl. no	Name of the Ward	Sl. no	Name of the Ward
1	Peenya Indl. Area	11	K R Puram	21	HBR Layout
2	Yelahanka Town	12	Marathahalli	22	Kadugondanahalli
3	Thanisandra	13	Jalahalli	23	C V Raman Nagar
4	Byatarayanapura	14	Yeshwanthpura	24	Shanthi Nagar
5	Vidyaranyapura	15	Mahalakshimpuram	25	Vijayanagar
6	Bommanahalli	16	Rajarajeshwari Nagar	26	Koramangala
7	Uttarahalli	17	Malleswaram	27	Hanumanth Nagar
8	Puttenahalli	18	Gandhinagar	28	BTM Layout
9	Arakere	19	Chamrajapet	29	Padmanabha Nagar
10	Ramamurthy Nagar	20	Hebbala	30	Rajajinagar

Data Collection :

Wardwise population data was collected from the City Municipal Corporation. Based on the literature survey the sample size of 1.5% was selected. The House Hold Interview surveys were conducted in all the wards. 115 houses were surveyed in each ward. The information sought for was obtained through the questionnaire which included household information, person information and trip information of the residents. Information obtained are number of members in the household, vehicle ownership, income, sex, age, working status and number of trips made on the travel day.

Previous day trip details made by a member of the household was also collected. The information usually collected are :Identification of the person, who made the trip, address of the origin of the trip, address of the destination of the trip, time at which the trip started, time at which the trip ended, purpose of the trip and mode of transportation.

III. ANALYSIS OF THE DATA

Stratification of data :

The data obtained from House Hold Interview was stratified to cater a range of 0-15%, 10-25% and 100%. The total number of households interviewed was 3450 and the stratification was 450 households, 875 households and 3450 households respectively. The percentage of stratified data works out to be 13% for 450 households, 21% for 750 households and 100% for 3450 households.

Attributes :

The attributes contributing for the trip generation as well as trip attraction of the locality considered for analysis are the socio economic factors:

Size of the household, 2. Income of the household, 3. Distance to work, 4. Travel time and 5. Travel cost

Trend Analysis :

To analyze the trend of the occurrence of events or attributes the histogram and frequency distribution tables were prepared for each of the attribute as given in table 2.1 to 2.4. From tables, the histograms were constructed as given in following figures 1 to 5. Also the cumulative frequency curves (Ogives) are plotted as given in figure 6 to 10. From the frequency tables the statistic all parameters viz mean value, standard deviation and coefficient of variation have been arrived at for each attribute as given in table 2.12 to 2.14.

Table.2.1 Frequency distribution for Households and Size of the family

Size of the family	No. of Household		
	450	875	3450
1	10	17	58
2	171	312	1063
3	166	276	1069
4	68	165	745
5	23	64	290
6	9	25	131
7	3	16	94

Table.2.2 Frequency distribution for Households and Income

Sl.No	Income, Rs	No. of Household		
		450	875	3450
1	2,000	10	12	58
2	2,000-5,000	171	178	1063
3	5,000-10,000	166	169	1069
4	10,000-15,000	68	78	745
5	15,000-20,000	23	13	290
6	>20,000	9	10	131

Table.2.3 Frequency distribution for Households and travel time

Range of distance travelled (km)	No. of Residents			Travel time (min)
	450	875	3450	
Less than 1	86	67	161	Less than 10
1-3	125	117	633	10-20
3-6	182	158	1272	20-30
6-9	97	99	811	30-40
9-12	52	31	369	40-50
12-15	10	10	121	50-60
More than 15	18	11	83	More than 60

Table.2.4 Frequency distribution for Households and travel cost

Travel cost (Rs)	No. of household residents		
	450	875	3450
Less than 10	62	146	823
10-20	115	211	1077
20-30	153	253	832
30-40	73	139	371
40-50	21	67	159
50-60	11	26	90
More than 60	9	33	98

Histograms and frequency Curves

The histograms are plotted for the various attributes with the number of households with frequency curves. Following are the histograms and frequency curves for the stratified data of 450 households. Similarly histograms and frequency curves are drawn for the sample size of 875 and 3450 respectively. These are shown in Figure 1 to Figure 5.

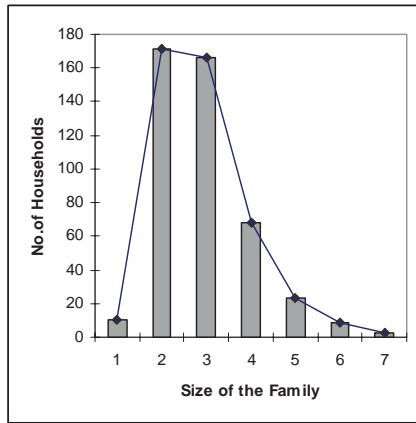


Fig 1: Histogram and Frequency Curve for Size of the family of 450 households sample.

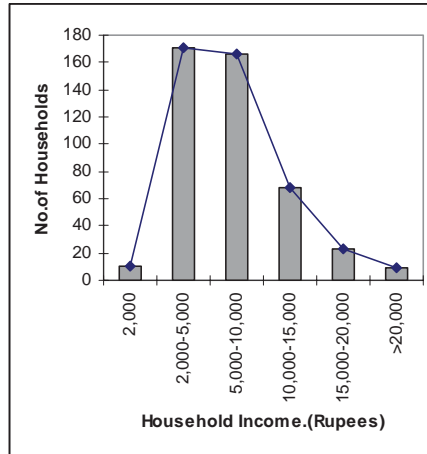


Fig 2: Histogram and Frequency Curve for the Income of 450 households sample

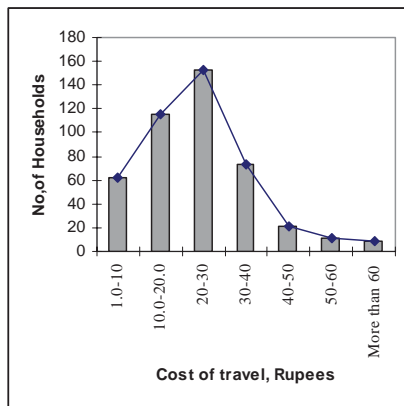


Fig 3: Histogram and Frequency Curve for travel cost of 450 households sample

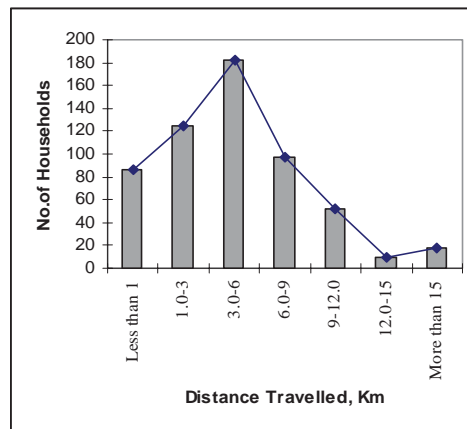


Fig.4: Histogram and Frequency Curve for Distance traveled time of 450 household sample

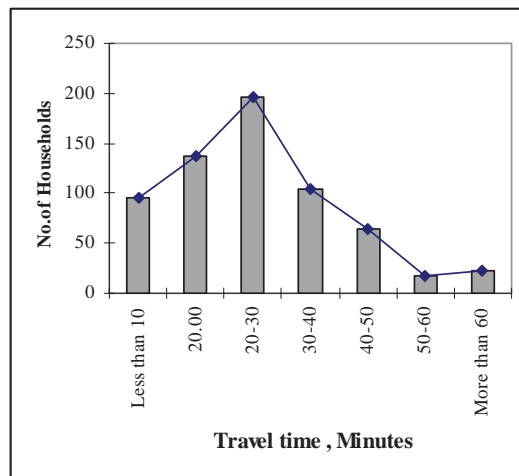


Fig 5: Histogram and Frequency Curve for travel time of 450 households sample
Ogive Curves- Frequency distribution

From the frequency tables given in tables 2.1 to 2.4, the cumulative frequency curves or the Ogive curves have been plotted for various attributes. Ogive curves for the sample size of 450 are shown below.

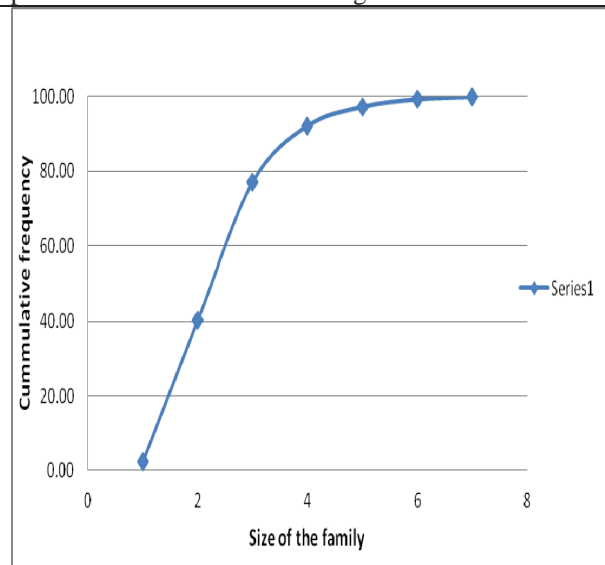


Fig 6: Ogive curve for size of the family of 450 households

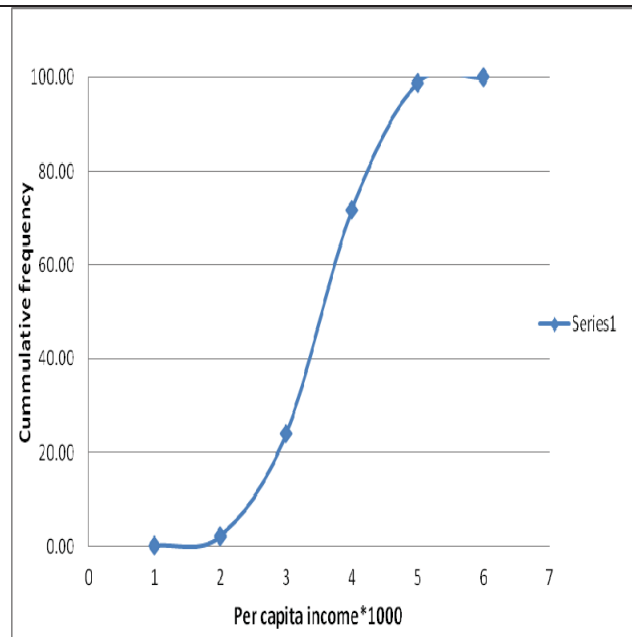


Fig 7: Ogive curve for per household income of 450 households

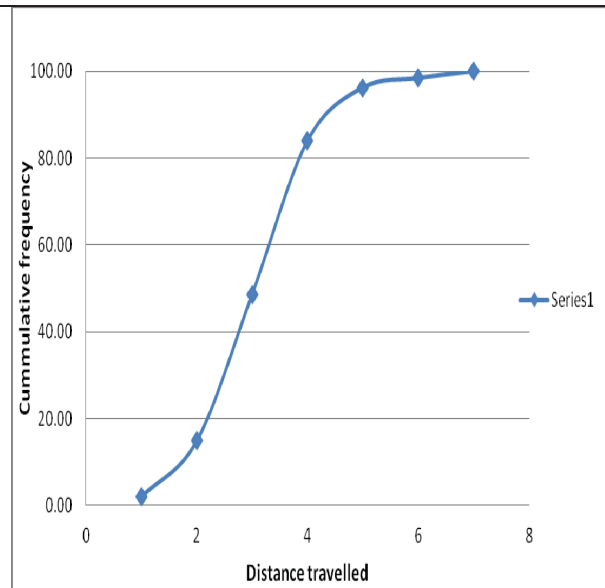


Fig 8: Ogive curve for distance travelled of 450 households

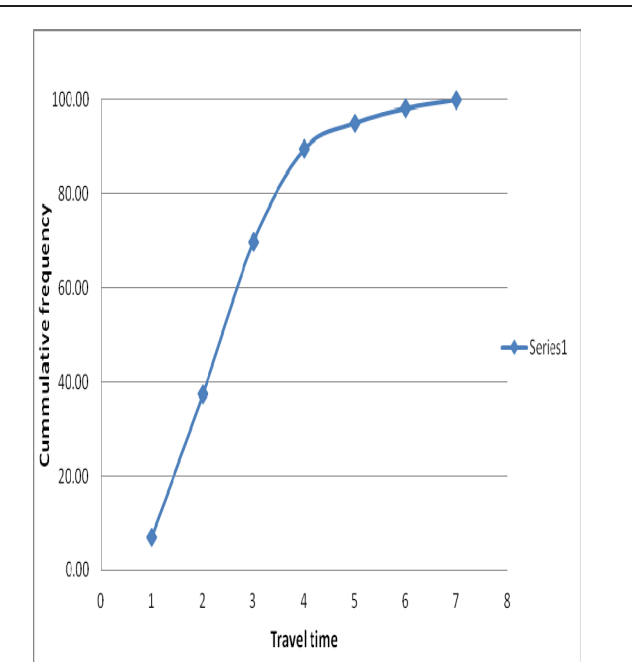


Fig 9: Ogive curve for travel time of 450 households

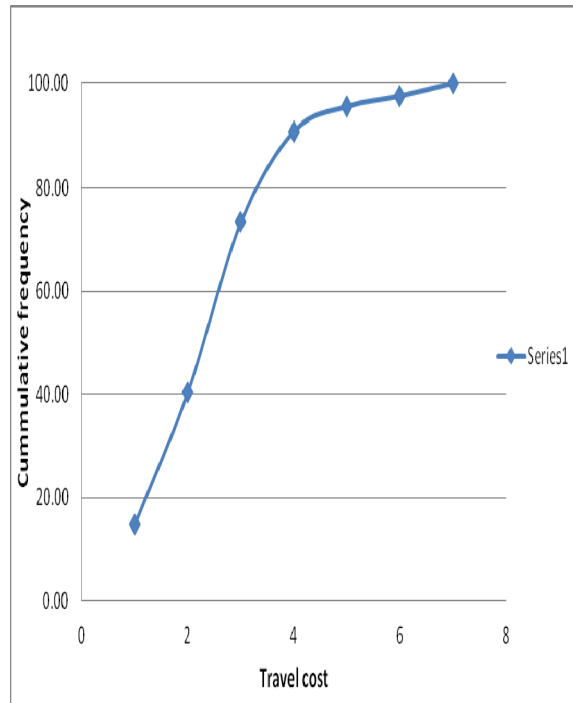


Fig 10: Ogive curve for travel cost of 450 households

From the figures 6 to 10 and tables 2.1 to 2.4 the following statistics are obtained : 10th percentile, 50th percentile and 90th percentile . The following tables 2.5 to 2.11 are the percentiles for the attributes considered for stratified data of 450,875 and 3450 households.

Table2.5: Percentiles of 450 households

Percentile	Size of family	Income *1000 (Rs)	Distance Travelled(Km)	Travel Time (min)	Travel Cost (Rs)
10th	1.2	2.5	1.8	1.1	0
50th	2.2	3.7	3	2.4	2.3
90th	4	4.7	4.3	4	4

Table2.6 : Percentiles of 875 households

Percentile	Size of family	Income *1000 (Rs)	Distance Travelled(Km)	Travel Time (min)	Travel Cost (Rs)
10th percentile	1.2	2.2	1.8	1	0
50th percentile	2.4	3.2	3	2.1	2.2
90th percentile	4	4.5	4.8	4	4.5

Table 2.7: Percentiles of 3450 households

Percentile	Size of family	Income *1000 (Rs)	Distance Travelled(Km)	Travel Time (min)	Travel Cost (Rs)

10th percentile	1.2	2.2	1.3	1	0
50th percentile	2.6	3.1	2.8	2	1.8
90th percentile	4.2	4.2	4.5	4	4

4.6 Statistical Parameters:

The statistical parameters considered for analysis are Mean, Standard Deviation and Coefficient of variation. The above mentioned parameters were calculated for various attributes taken for the stratified data of 450 households, 875 households and 3450 households. The following tables 2.8 to 2.10 show the values of statistical parameters for different attributes.

Table 2.8 : Statistical Parameters for 450 Households (13% Sample size)

Parameter	Size of the family	Per Capita Income(Rs)	Distance travelled(Km)	Travel time(min)	Travel cost(Rs)
Mean	2.9156	13498.88	6.49	26.59	24.02
Std dev	1.0519	9092.68	3.16	12.69	14.03
Coeff.variation	36%	67%	49%	48%	58%
Std Error of estimate	0.097	840.12	0.2919	1.1724	1.2963

Table 2.9 : Statistical Parameters for 875 Households (21% Sample size)

Parameter	Size of the family	Per Capita Income(Rs)	Distance travelled(Km)	Travel time(min)	Travel cost(Rs)
Mean	3.1257	12107.65	6.55	24.46	25.49
Std dev	1.3369	7299	3.64	12.29	16.32
Coeff.variation	43%	60%	55%	50%	64%
Std Error of estimate	0.088	483.63	0.2411	0.8413	1.0813

Table 2.10: Statistical Parameters for 3450 Households (100% Sample size)

Parameter	Size of the family	Per Capita Income(Rs)	Distance travelled(Km)	Travel time(min)	Travel cost (Rs)
Mean	3.3023	11090.78	5.998	23.87	21.61
Std dev	1.4425	5289.46	3.79	13.10	15.71
Coeff.variation	44%	48%	63%	55%	73%
Std Error of estimate	0.048	176.50	0.1264	0.4371	0.524

Table 2.11: Statistical Parameters of distribution for 450, 875 and 3450 households

Attributes	No .of households Sample

	450		875		3450	
	Coefficient of skewness	Coefficient of kurtosis	Coefficient of skewness	Coefficient of kurtosis	Coefficient of skewness	Coefficient of kurtosis
Size of the family	0.285	0.866	0.142	12.34	0.066	40.65
Household income	-0.0909	3086.73	0.130	2634.11	0.1	2723.74
Distance travelled	0.04	0.804	0.2	0.865	0.0625	0.709
Travel time	0.103	-2.22	0.2666	0.285	0.333	0.24
Travel cost	-0.15	-0.204	0.022	1.756	0.1	-0.3075

It can be observed from the table that the Standard Deviation increases with increase in the number of households in the size of the family as attribute. The higher the Standard Deviation the greater will be the scatter. The coefficient of variation has been found to be a fairly stable measure of variability for homogeneous conditions.

Discussions :

The Histogram for 450 households (13% sample) for the attributes size of the family, per capita income, travel distance, travel time and travel cost as given in fig 1 to 5 show a normal tendency with symmetrical or skewed behavior. The attribute household income (fig 2) and distance travelled (fig 4) show a normal tendency with symmetrical behavior. Where as the attribute size of the family (fig 1) and travel time (fig 5) and travel cost (fig 3) have a tendency to show normal distribution skewed to the right i.e. positive skewness. The percentage skewness and coefficient of kurtosis have been discussed under cumulative frequency distributions.

Similarly histograms for 875 households (21% sample) and the histogram for 3450 households (100% sample) for the attributes size of the family, per capita income, travel distance, travel time and travel cost show a normal tendency with symmetrical or skewed behavior..

Discussions on Ogives i.e. cumulative frequency curves

The figures 6 to 10 for 450 household data indicate that the shape of the Ogive curve for the attributes household income (fig 7) and the distance travelled (fig 8) have a typical S shape indicating the normal frequency symmetrical distribution. The others indicate the skewness. Similarly for 875 and 3450 household data indicated S –shaped curves for household income and distance travelled shape indicating the normal frequency symmetrical distribution . Ogive curve for other attributes indicate the Skewness. The percentile distribution, the following parameters were established P10, P50 and P90, Q1 and Q3 and the coefficient of skewness and kurtosis were established from the standard equations.

Discussions on the Statistical parameters

The basic statistic parameters viz Arithmetic mean, Standard deviation, Coefficient of variation and The standard error of estimate for 95% confidence level are evaluated and tabulated in tables 2.8 to 2.10. These tables indicate that as the sample size increases from 13% to 100%, the standard deviation, which is a measure of scatter also increases from 1.0519 for 450 households (13% sample) to 1.4425 for 3450 households (100% sample) in case of size of family attribute. This demonstrates that as the sample size increases the scatter also increases. Per capita income indicates a gradual reduction of coefficient of variation 67% for 450 households (13% sample) to 48% for 3450 households (100% sample) indicating the higher reliability can be obtained at large sample size. The standard error of estimate (e) evaluated for 95% confidence level has decreasing trend with the increase in sample size for all the attributes. The least values are for a sample size of 3450 households (100% sample). The attribute size of the family has a least value of error of 0.048. It is the only unbiased statistically stable attribute in this study.

Discussions on the coefficient of skewness and kurtosis

As per the table 2.11 the consolidated statistical parameters, coefficient of skewness for the five attributes considered show positive trend for three attributes and negative trend for income and travel cost for 13% sample. As the percentage samples increases to 875(21%) and 3450(100%), the negative skewness disappears and all attributes exhibit positive trend. The lowest skewness is observed for distance travelled and hence there will be less scatter.

The coefficient of kurtosis for the five attributes considered show positive trend for three attributes and negative trend for income and travel cost for 13% sample. As the percentage samples increases to 875(23%) and 3450(100%), the negative skewness disappears and all attributes exhibit positive trend. The lowest kurtosis is observed for travel time and hence there will be less scatter.

IV. CONCLUSIONS

The home interview study for the attributes size of the family, per capita income, distance travelled, travel time and travel cost for 450 households, 875 households and 3450 households are analyzed for their statistical parameters and the conclusions were drawn are given below.

1. From the statistical parameters analyzed it can be concluded that as the sample size increases, the standard deviation for the size of the family as attribute increases. This indicates that as the sample size increases the scatter also increases for this attribute.
2. The attribute per capita income has a gradual reduction of coefficient of variation with increase in sample size which indicates higher reliability at large sample size is obtained for this attribute.
3. The standard error of estimate (e) at 95% confidence level has a decreasing trend with the increase in sample size for all the attributes considered.
4. From the histograms of various attributes for stratified data, it can be concluded that the attributes per capita income and distance travelled show a normal tendency with symmetrical behavior where as the remaining attributes size of the family, travel time and travel cost are skewed to the right.
5. The coefficient of skewness has negative skewness for the attributes per capita income and travel cost for lower sample size. As the sample size increases the negative skewness disappears exhibiting positive trend. Similarly the coefficient of kurtosis has negative kurtosis and it exhibits positive as sample size increases.

REFERENCES

- [1] Wilbur Smith Associates, "Comprehensive Traffic and Transportation Study" for Bangalore Metropolitan Region June 2008, Bangalore
- [2] Cochran, "Sampling Techniques" Third edition, Wiley, 1999, New York.
- [3] Sarna, A.C. ,Surya Narayana, Rao, S.B.S, Suthar, H.H and Basant Lal, "An analysis of Household Travel survey data for Bombay Metropolitan Region", Indian Road Congress Journal, Volume 47-1, Aug 1986, Paper number 376. New Delhi
- [4] Cambridge Systematics, Inc., "Review of Activity based Models", Florida Department of Transportation ,May 2008, USA
- [5] Fitzmorice,G.,Laird,N., and Ware,J. "Applied Longitudinal Analysis" Wiley, 2004, IEEE.
- [6] Liu, G and Liang, K, "Sample size Calculations for Studies with Correlated Observations",1997, Biometrics 53(3).
- [7] 7). Parsons Brinckerhoff Quade and Douglas, Inc., "Regional Travel Household Survey" Final Report, 2000, Prepared for the New York Metropolitan Transportation Council (NYMTC), New York.