# Investigating the Normality of Data Set Collected to Resolve the Accessibility Issues of Badin Sub-Region, Sindh, Pakistan

MIR AFTAB HUSSAIN TALPUR\*, FAREED AHMED MEMON\*\*, AND SHABIR HUSSAIN KHAHRO\*\*\*

#### **RECEIVED ON 28.01.2015 ACCEPTED ON 28.05.2015**

## ABSTRACT

To highlight the accessibility issues of rural households, the data were collected from the rural subregion of Badin, Sindh, Pakistan. The purpose of the data-collection was to offer policy proposal's inputs, as concerned planning agencies could be able to know the ground realities and implement their planned decisions accordingly. Therefore, the aim of this study was to investigate about the normality of the data, collected through the questionnaire survey from the sampled population of the study area. Total hundred questionnaires were completed from sampled rural households. Mostly, the household's demographic and socioeconomic aspects were targeted in this regard, including travel- time and distance characteristics. The data were successfully analyzed in SPSS-17.0 and results were generated accordingly. The results exhibited the symmetry in household-size aspect due to clustering; whereas, household income, traveldistance and time depicted enormous variations.

Key Words: Accessibility, Normality, Rural Sub-Region, Demography, Travel Time, Travel Distance.

## **1. INTRODUCTION**

ransportation networks are articulated for the welfare of humankind, as they could access their jobs effectively. Better accessibility standards are significant for sound living environments, particularly in rural sub-regions [1]. For any of the transport-related organizations, the issues are not just the paving of roads and their widening; it is a matter of catering backward communities with appropriate and equivalent opportunities to access essential activities. Efficient transportation systems, along with reachable and suitable network services are indispensable for planned regional growth. The transport accessibility problems can be regarded as a root-cause for the most of the socioeconomic issues of rural inhabitants. As, miserable socioeconomic conditions, poverty, hunger, unavailability of transportation facilities, together with the shortage of basic health and primary education institutions are dominating the distressed sub-regions of developing countries with the changing instances. The backward sub-regions are termed as conventional agriculture-based societies, together with low-labor productivity. Commonly exposed to policy limitations; urbanization has been an often phenomenon in developing countries [2]. Because of the absence of convenient road access and low-cost freight transport services, peasants often faced the difficulties related to accessibility and mobility. At scattered locations, agriculturalists lack market information and struggled to meet with the demands of consumers from food safety,

\* Assistant Professor, Department of City & Regional Planning, Mehran University of Engineering & Technology, Jamshoro.
 \*\* Assistant Professor, Department of Civil Engineering, Mehran University of Engineering & Technology, Jamshoro.

\*\*\* Lecturer, Department of Civil Engineering, Mehran University of Engineering & Technology, Jamshoro.

and quality perspectives. As a result, they struggle to cater standardized agricultural products on a continuous basis [3]. It is a general perception and closed to reality that remote sub-regions often agonized from higher costs of the products than the market or accessible locations. The transportation cost is a major factor to highlight the linkages between accessibility and agro-based development [4]. Costs always related to purchase inputs or market products and often approximated, such as the closeness of markets to the road facilities [5,6].

The decision-makers and policy-planners, due to political interests and socioeconomic reasons, neglected rural subregions of the developing countries. The policy proposals were often developed to rectify urban issues; whereas, rural settlements frequently left on their own. It was procured that rural development agencies also struggled to cope-up with the mounting disastrous situations in rearward sub-regions, and inept to deliver solutions to the difficulties of the rural population. Without any doubts, policy plans are required to deal with transport accessibility issues [7]. The problem of rural inaccessibility can also be resolved with the timely execution of policy proposals. The transportation policy development and formulation of credible transportation planning models certainly depend on the data, and previous records about the trips, travel patterns, demographic and socioeconomic characteristics of the rural population, together with the transportation network and land-use zoning information [8, 9]. Consequently, data were collected from the study area to resolve the accessibility-related issues of the rural population. To talk about the accessibility measures and their implementation is out of the scope of this research; nonetheless, just to conduct the normality test on the data set is the aim of this study, which were collected to highlight the accessibility problems of the rural population of the Badin sub-region.

From the rural-regional perspective, accessibility dealt with respect to socioeconomic and demographic aspects. The focus of this study was to examine the data, as to know the consequences of the lower household's income and higher household sizes regarding the travel decisionmaking process of rural households. Therefore, this research purposed to perform the normality test on the data, collected through the questionnaire survey from the study area. The collected data comprised of household income, household size, travel distance and time. The test was successfully performed on the data and the results were retrieved. Symmetrical household income and size can be seen from the data, due to cluster sampling, as the target was lower-income group. Whilst the travel distance and commuting time data showed variations, because of dispersed household activities.

### 2. MATERIALAND METHODS

The sub-region of Badin was selected as a study area, which can be characterized as one of the dejected subregions of Pakistan [10]. To acquire reliable data, varieties of available data-collection and survey techniques were reviewed [11]. Before executing a survey exercise, sample sizes were determined for the rural population, which can be shown in Fig. 1.

Total 10% (maximum level) of the rural population can be selected as a sample size [12]. Equally, it is apprehended from Fig. 1 that the total number of inhabitants in the Badin sub-region were approximately 1,177,997 in the year 2012. According to the report, i.e. "District Vision Badin," it was found that about 75% of the total population termed as the rural population [13]. Accordingly, 75% of the total population will be 883,500, which was the targeted number. As, ten percent of the total rural population can be selected for the questionnaire survey; hence, it will be 88,350 inhabitants. This is the sample size, which was considered for the questionnaire survey from the study area, together with interviews from the household heads. Then, the questionnaire survey was successfully carried out and data were retrieved, focusing household size, household income, travel distance and time.

Mehran University Research Journal of Engineering & Technology, Volume 35, No. 1, January, 2016 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

The location of survey areas can be seen in **Fig. 2** with light blue rectangles. The collected data were scrutinized and analyzed with the help of *SPSS-17.0* [14]. The main objective of the survey was to know that how rural household's travel and socioeconomic features affect their travel decision-making process, which is not discussed in this study. Because, this paper only emphasized on the normality test conducted on the data, which were obtained through the questionnaire survey.

Fig. 2 shows the survey sites chosen for the data-collection process. The number of hundred questionnaires was decided to be filled from the study area. Although, there were no standards available in this regard; hence, sampling methodologies provided liberty to select the practicable number of questionnaires to be filled from the study area, which must meet with standard population sample size. In addition, in this study, the sum of the hundred questionnaires was procured after a brief literature review, which authenticated the proportion made for questionnaire survey [15,16]. In this regard, the number of twenty questionnaires was decided to be filled from each Taluka of the Badin sub-region. The statistics about Badin Taluka can be seen in Fig. 3. Fig. 3 verifies that number of twenty questionnaires were completed, and data about household income, household size, work distance and travel time were retrieved successfully. As shown, the nineteenth data point for household size recorded as twenty-four, which confirmed that the joint family system in the sub-region is still prevailing.

The data about Golarchi Taluka can be depicted in Fig. 4.

Symmetrical data can be observed from Fig. 4, except datapoint eight, which shows enormous household income, i.e. 300,000. That was the result of the existence of a household, which was mainly dependent on agriculturelands, and their income was higher as compared to other household units.

The data-figures of Matli Taluka can be apprehended in Fig. 5.

Similar household size can be observed from Fig. 5, together with parallel travel time. Mostly, lower-income

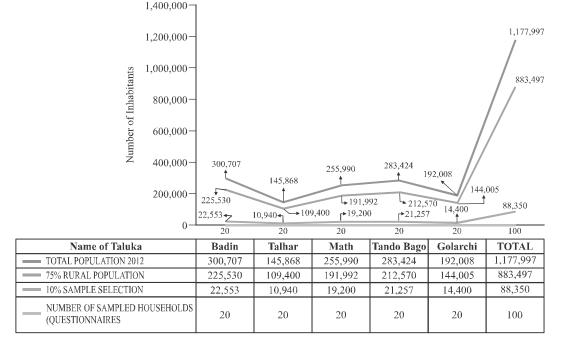


FIG. 1. SAMPLED DEMOGRAPHIC FEATURES (ESTIMATED)

households were situated nearer to their work locations, having the maximum commuting time of an hour.

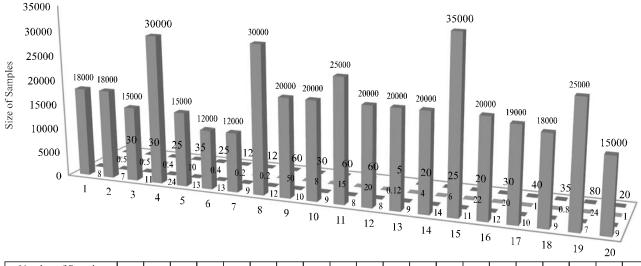
The information about the Talhar Taluka can be referred in Fig. 6.



FIG. 2. QUESTIONNAIRE SURVEY LOCATIONS, BADIN SUB-REGION

Mehran University Research Journal of Engineering & Technology, Volume 35, No. 1, January, 2016 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

The household size in Fig. 6 looks quite similar, while travel time and travel distance showed variations. The reason is some of the trips were recorded for long distances and travel times, because of health emergencies. Meanwhile, household income confirmed parallel values, due to clustering.



Number of Samples	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Household Income (Rs)	18000	18000	15000	30000	15000	12000	12000	30000	20000	20000	25000	20000	20000	20000	35000	20000	19000	18000	25000	15000
<ul> <li>Household Size</li> </ul>	8	7	-11	24	13	13	9	12	10	9	8	8	9	14	-11	12	10	9	7	9
Work Distance (Km)	0.5	0.5	0.4	10	0.4	0.2	0.2	50	8	15	20	0.12	4	6	22	20	1	0.8	24	1
<ul> <li>Travel Time (Mins)</li> </ul>	30	30	25	35	25	12	12	60	30	60	60	5	20	25	20	30	40	35	80	20

FIG. 3. QUESTIONNAIRE DATA (BADIN TALUKA)

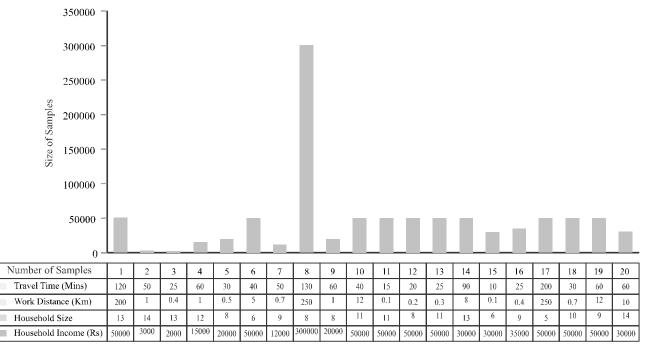


FIG. 4. QUESTIONNAIRE DATA (GOLARCHI TALUKA)

The figures about Tando Bago Taluka given in Fig. 7. Fig. 7 illustrates a similarity in household income, because the low-income household was selected to collect data that how lower-income can affect the travel decisionmaking process of rural households. In addition, household size also has not shown any huge variations.

After the process of data-collection, the same were analyzed, which can be seen in the results and discussion section hereafter. It should be noted that the analytical process was conducted on the data as a whole, comprised of hundred questionnaires.

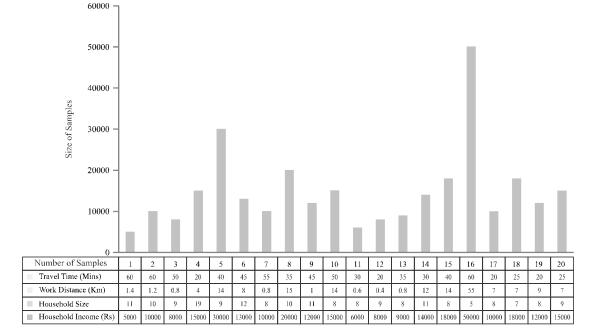


FIG. 5. QUESTIONNAIRE DATA (MATLI TALUKA)

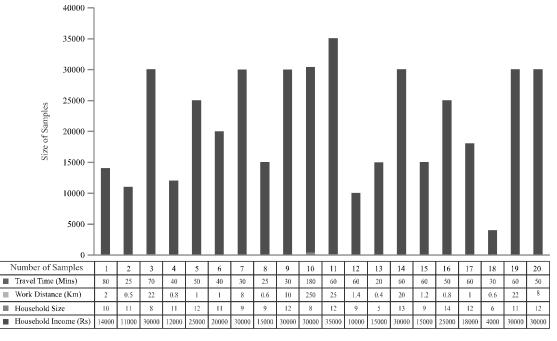


FIG. 6. QUESTIONNAIRE DATA (TALHAR TALUKA)

#### 3. **RESULTS AND DISCUSSION**

To check the normality of the data, the common method is to run descriptive statistics. This will give the values of "skewness" and "kurtosis." Skewness and kurtosis provide numerical values, which can assist in deciding the normality of the data. Skewness values range from +2 to -2 or when data are normally distributed. In addition, some authors have taken the standard values as +1 to -1. The left skewed curve shows the sign of positive values, while the right skewed curve depicts negative values. Kurtosis measures the flatness and height of curves. The standard value of kurtosis is zero. However, a value less than 1.96 can be regarded as normal. The larger value of kurtosis is 2.58 and 3.29 can be considered as a very large kurtosis value [17].

#### 3.1 **Evocative Statistics**

The different tests were experimented by using SPSS-17.0 in order to figure out the different parameters of the collected data, including household income, household size, travel distance to work places and travel timeduration. The statistical results can be shown in Table 1, which supported in calculating the normal distribution of the aforementioned data aspects.

Table 1 shows the statistical results of the data set collected through the questionnaire survey. The notation "N" confirms that number of hundred questionnaires was filled from the Badin sub-region. Mean values for household size reflects that higher household size neighborhoods were selected with the help of cluster sampling. The skewness values for household size, travel distance and travel time were closed to normal ranges, while kurtosis values analyzed as higher than the standards. In addition, huge variations can be seen in household size, i.e. from four to twenty-four, and average household income from two thousand to three hundred thousand (Pakistani Rupees), which verified the imbalance socioeconomic regional structure.

Before giving the arguments about the normal curve distribution of these parameters, it is worth mentioninghere to describe the standards of normal distribution or ideal curve characteristics. There are five divergent standards available [18], which makes a normal curve:

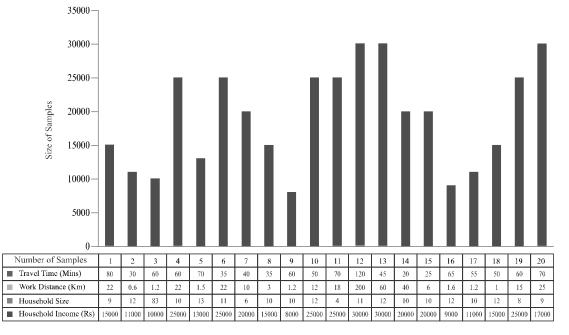


FIG. 7. QUESTIONNAIRE DATA (TANDO BAGO TALUKA)

Mehran University Research Journal of Engineering & Technology, Volume 35, No. 1, January, 2016 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

- (1) One hump in the middle of the curve;
- (2) Symmetry in mean, mode and median;
- (3) Symmetrical shape (not skewed);
- (4) Asymptotic (the extreme points never touches x-axes);
- (5) Zero-level Kurtosis, i.e. neither flat nor peaked curve.

This should be noted that it is difficult to comply with all these standards, when the data is socioeconomic and collected through primary methods. The curve distribution test gives information about the normality of the data. No further normality tests were carried out on the data, in order to avoid data manipulation and its existence in original format. Therefore, the original data were used in the analysis to know the facts of the Badin sub-region. To perceive the information about the behavior of the normal distribution curve of the regional household data, analysis was carried out, which can be seen hereafter. Figs. 8-11 enlightened the frequency distribution of rural household's travel patterns and socioeconomic parameters, together with a normal curve comparison. The arguments about the rural-regional travel aspects are given hereafter.

#### 3.2 Regional Household Income Aspects

The income range in the study area was recorded from PKRS. 2,000-300,000/- (Fig. 8). The people in the study area were observed lower-income, as very less job, employment and work opportunities were available in the Badin sub-region. The income of the regional people was considered much lower, as compared to the country's inflation level, and most of the inhabitants were living their lives below the poverty line. The income level of the study areas showed larger variations. This is due to the imbalance socioeconomic conditions of the study area. The people with jobs or have agriculture lands possessed larger monthly earnings, as compared to the shopkeepers or low-level government employees.

As shown in Fig. 8, most of the rural household's income, i.e. nearly 62% recorded within the range of PKRS. 25,000/ -. The regional household income showed a bit left skewed distribution curve with the value of 7.612. Fig. 8 also depicts that more than half of the local inhabitants earn lesser than 50,000 per month. This amount can be considered as lower, when compared to the higher household sizes of the study area, together with the higher inflation rate.

Statistics	Regional Household Income (PKRS)	Regional Household Size	Regional Travel Distance (Kms)	Regional Travel Time Duration (Mins)		
Ν	100	100	100	100		
Mean	24700.00	10.02	19.141	47.39		
Median	20000.00	10.00	3.5	40		
Mode	30000	8	1	60		
Std. Deviation	30486.13	2.78	50.173	30.769		
Range	298000	20	250	195		
Skewness	7.612	1.393	3.961	2.417		
Standard Error of Skewness	0.241	0.241	0.241	0.241		
Kurtosis	68.356	6.107	14.981	8.655		
Standard Error of Kurtosis	0.478	0.478	0.478	0.478		
Minimum	2000	4	0.10	5		
Maximum	Maximum 300000		250	200		
Sum	Sum 2470000		1941.12	4379		

TABLE 1. REGIONAL HOUSEHOLD TRAVEL AND SOCIOECONOMIC STATISTICS

#### **3.3 Regional Household Characteristics**

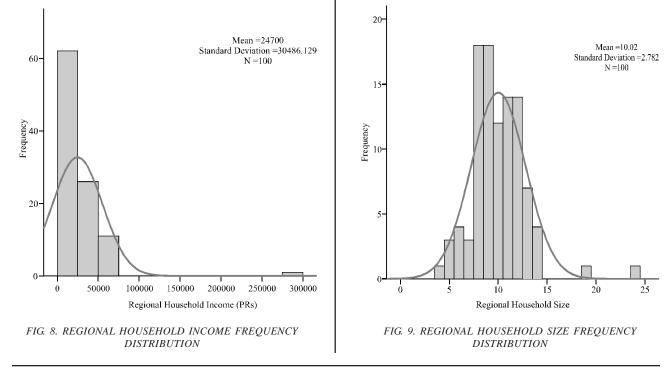
The regional household size ranges were observed from four to twenty-four inhabitants (Fig. 9). The questionnaire survey was conducted from intensive populated areas, in order to reach the minimum standard level of the population that should have been included, while performing these sorts of studies. Fig. 9 shows huge household size, this is mainly because of the joint family system in the study area, where one or two persons earn for the rest of the family. This is also one of the main reasons behind the catastrophic socioeconomic conditions in developing countries, including Pakistan.

The curve is shaped like a bell, which is the sign of normal distribution. The value of skewness is 1.393 that looks normal. The household size of the study area is calculated with a mean of 10 and standard deviation of 2.78. The normal curve can be seen in Fig. 9, as there were symmetrical proportions of household sizes observed in most of the settlements in the study area.

### 3.4 Regional Travel Distance Characteristics

Fig. 10 expresses the number of trips occurred within the range of about 12 km. The main reason behind is the shopkeepers, lower-rank officers, school or college students; and employees, who lives in close proximity to their workplaces. The rural households with low-income hesitate to pay for transportation services in the developing world and would like to reside nearer to their job opportunities. Fig. 10 also displays the higher distances; this is mainly because the higher-rank officers with their personal vehicles were out-back on a daily basis. Some trips were also observed with higher distances with respect to medical emergencies; as prominent health facilities were available at the regional headquarter settlement, which was situated at almost 250 km away from the scattered locations.

The left skewed curve is observed from Fig. 10, together with the skewness value of 3.961, as most of the regional household trips by distance vary from 0.10-250 km.



Mehran University Research Journal of Engineering & Technology, Volume 35, No. 1, January, 2016 [p-ISSN: 0254-7821, e-ISSN: 2413-7219] 79

#### **3.5** Regional Travel Time Characteristics

The higher standard deviation was observed from Fig. 11, due to the variations in travel distances and their timings. Most of the journey timings lay within the range of an hour. This was probably due to the unavailability of road facilities and transportation services. The private bus service plied on the rural roads without maintenance, resulting in lower speed and delays. The mean travel time is recorded about 48 minutes, which considered as long with respect to rural areas of the developing world.

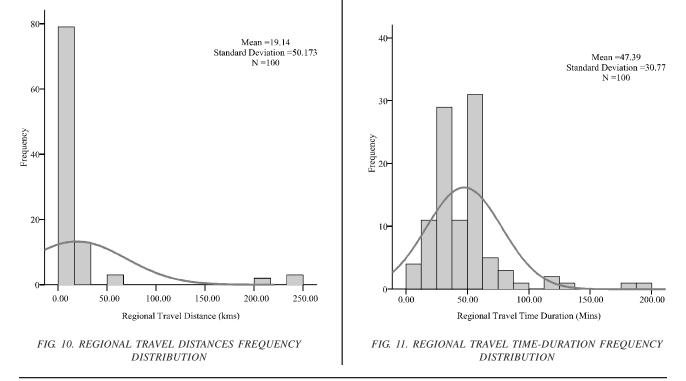
A hint of the bell-shaped left skewed curve is experienced from Fig. 11, as skewness is 2.147. From this figure, it can be argued that the most of the journey timings were recorded up to an hour. Because, rural people used to walk to their respective destinations, which took higher journey timings for shorter distances. This should be noted that in rare cases, socioeconomic data showed the normal curve behavior, which collected through questionnaire survey or by using any primary method.

#### 4. **RESEARCH GAP**

This study can assist policy makers and concerned planning agencies to implement their decisions, focusing the accessibility issues of rural households. For the Badin sub-region, Pakistan, this study abridged the gap of data unavailability regarding the travel and socioeconomic features of rural households. In addition, this study showed a way to decision-makers, through which they could be able to analyze the data and clarify the problems of rural households. In this way, the decision-making process would be easier and faster, which could assist local planning agencies to eliminate the accessibility issues of rural households.

### 5. CONCLUSIONS

This study is based on the data collected through the questionnaire survey from Badin sub-region Pakistan. The data were collected to eliminate the accessibility issues of rural households, as concerned planning agencies could be able to devise policy proposals accordingly. In this



Mehran University Research Journal of Engineering & Technology, Volume 35, No. 1, January, 2016 [p-ISSN: 0254-7821, e-ISSN: 2413-7219] 80

regard, rural household's travel and socioeconomic data, such as travel time, travel-distance, household income and household size were successfully collected with the help of hundred questionnaires. To know the behavior of aforementioned parameters, it was aimed to investigate about the normality of data. Thus, the objective of this study was set, which was successfully achieved, and normality test was categorically conducted. Only, the curve linked to parameter household size was recorded normal; whereas, the remaining curves were recorded left skewed together with higher kurtosis levels.

In addition, the mean, mode and median of the data parameters clarified travel and socioeconomic characteristics of rural households. Since, the parameters selected in this study possessed versatile characteristics, like income (local currency), travel time (minutes), traveldistance (kilometers) and household size (the number of family members). As, the values of these parameters were recorded differently, so the mean, median and mode values were also explicated variations. Therefore, as it was described earlier that most of the curves were not found normal, but the curve linked to household size. The reason is during data-collection, lower income household were selected, which possessed similar socioeconomic features, including household size. This is the reason why the curve of the parameter, i.e. household size was found normal.

Besides, the data related to other three parameters, like household income, travel distance and time showed disparities. The main reason of performing these statistical tests was to determine the behavior of the normal curve in order to check the normality of the overall data set. The goal of this study was accomplished successfully.

However, kurtosis levels were greater than standard level of zero for all mentioned figures. Meanwhile, Fig. 9 showed a bell-shaped normal curve, as the skewness value was recorded within the acceptable ranges. The values of mean, mode and median of this parameter were also recorded in close ranges (symmetrical). In addition, all the curves were drawn as *"one-humped*," which exhibited the overall normality of the data.

### 6. **RECOMMENDATIONS**

The absence of data is a one the major problems of rural planning agencies, which hindered the development process of policy proposals. In particular, the policy plans are the key ingredients for engineers and planners to mitigate the ever-increasing issues of rural communities. Because of the lack of development plans, accessibility and related issues have arisen over the years in the Badin sub-region, Pakistan. These problems made rural households vulnerable to physical, socioeconomic and environmental calamities. Hence, within the context of study area, it is recommended that the accessibility issues of rural households should be eliminated with the help of proper policy guidelines. The development plans should be developed and various professionals, like civil engineers, planners or surveyors should be given equal opportunities in the decision-making process, as they could add their valued inputs for the betterment of rural communities.

### ACKNOWLEDGEMENTS

The corresponding author is truly grateful to the reviewers for their valuable comments, co-authors and the Department of City & Regional Planning, Mehran University of Engineering & Technology, Jamshoro, Pakistan, for their support and provision of encouraging research environment.

#### REFERENCES

 Velaga, N. R., Beecroft, M., Nelson, J. D., Corsar, D., and Edwards, P., "Transport Poverty Meets the Digital Divide: Accessibility and Connectivity in Rural Communities", Journal of Transport Geography, Volume 21, pp. 102-112, 2012.

- [2] Naeem, A., "Sindh's Imbalanced Urbanization: Seeking Sustenance through Revival of Historic Urban Centres", Journal of Sustainable Development, Volume 4, 2011.
- [3] Gulati, A., Minot, N., Delgado, C., and Bora, S., "Growth in High-Value Agriculture in Asia and the Emergence of Vertical links with Farmers", Linking Small-Scale Producers to Markets: Old and New Challenges, The World Bank, 2005.
- [4] Combes, P.P., and Lafourcade, M., "Competition, Market Access and Economic Geography: Structural Estimation and Predictions for France", Regional Science and Urban Economics, Volume 41, pp. 508-524, 2011.
- [5] deEspindola, G.M., deAguiar, A.P.D., Pebesma, E., Câmara, G., and Fonseca, L., "Agricultural Land use Dynamics in the Brazilian Amazon Based on Remote Sensing and Census Data", Applied Geography, Volume 32, pp. 240-252, 2012.
- [6] Reggiani, A., Bucci, P., Russo, G., Haas, A., and Nijkamp,
   P., "Regional Labour Markets and Job Accessibility in City Network Systems in Germany", Journal of Transport Geography, Volume 19, pp. 528-536, 2011.
- [7] Talpur, M.A.H., Madzlan, N., Irfan, A., Chandio, I.A., and Hussain, S., "Time-Space Geography: A Creditable Transport Accessibility Measure for Rural Dwellers", Applied Mechanics and Materials, Volume 567, pp. 763-768, 2014.
- [8] Jaraíz, C.F.J., Mora, A.J., Gutiérrez, G.J.A., and Jeong, J.S., "Comparison of Regional Planning Strategies: Countywide General Plans in USA and Territorial Plans in Spain", Land Use Policy, Volume 30, pp. 758-773, 2013.
- [9] Talpur, M.A.H., Napiah, M., Chandio, I.A., and Khahro, S.H., "Research Framework Focusing Transportation Accessibility Planning, Computer Based Modeling and Transportation Policy Outlines for Remote Regions of Developing World", Research Journal of Environmental & Earth Sciences, Volume 5, pp. 32-40, 2013.

- [10] Jamal, H., and Lohano, H.R., "Understanding Rural Poverty Dynamics: The Case of the Poorest District of Sindh Pakistan", 2008.
- [11] Talpur, M.A.H., Napiah, M., Chandio, I.A., and Khahro, S.H., "Transportation Planning Survey Methodologies for the Proposed Study of Physical and Socio-Economic Development of Deprived Rural Regions: A Review", Modern Applied Science, Volume 06, pp. 1-16, 2012.
- [12] Turner, A., Yansaneh, I., Jambwa, M., Banda, J., Menozzi,
   C., and Kassa, B., "Designing Household Survey Samples: Practical Guidelines", United Nations, New York, 2008.
- [13] "District Vision Badin (A Framework for Sustainable Development)", International Union for Conservation of Nature and Natural Resources. The World Conservation Union, Pakistan, Sindh Programme Office,1-Bath Island Road, Karachi, pp. 134, 2006.
- [14] Mirabella, J., "Hypothesis Testing with SPSS, http:// www.drjimmirabella.com/ebook/excerpt%20from%20
   Hypothesis%20Testing%20with%20SPSS%20ebook%20
   (Jim%20Mirabella).pdf, Retrieved on 11.06.2013, 2006.
- [15] Kenyon, S., "The 'Accessibility Diary': Discussing a New Methodological Approach to Understand the Impact of Internet Use Upon Personal Travel and Activity Participation", Journal of Transport Geography, Volume 14, pp. 123-134, 2006.
- [16] Srinivasan, S., and Rogers, P., "Travel Behavior of Low-Income Residents: Studying Two Contrasting Locations in the City of Chennai, India", Journal of Transport Geography, Volume 13, pp. 265-274, 2005.
- [17] O'Neil, S., "Basic Statistics for the Utterly Confused", University of Pretoria Faculty of Economic & Management Science, 2009.
- [18] Shaikh, M.U., Shaikh, A.A., Qureshi, S., and Fatima, S.S., "Applied Statistical Methods and Estimations", First Edition, Jahangir Publications, C-76, Gulistane Sajjad, Hyderabad, Sindh, Pakistan, 2010.

Mehran University Research Journal of Engineering & Technology, Volume 35, No. 1, January, 2016 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

82