



Determining the Standard Rural Population and Household Sample Sizes for Questionnaire Survey Focusing Transport Accessibility

M. A. H. TALPUR⁺⁺, I. A. CHANDIO, M. S. JUMANI*, M. NAPIAH^{**}

Department of City and Regional Planning, Mehran University of Engineering and Technology, Jamshoro

Received 11th May 2015 and Revised 6th November 2015

Abstract: This study was conducted to procure authentic data, required to resolve transport accessibility problems of rural households of the Badin sub-region, Pakistan. Over the years, transport accessibility issues, i.e. travel delays and impedance hindered local residents in performing their daily routine activities and caused socioeconomic deprivation. To clarify the research problems and to resolve them, standardized data needed to be collected from the study area. Thus, the aim of this study was to find out the standard household and population sample sizes for the questionnaire survey, centering the household's transport accessibility information, i.e. travel mode, distance, time, household income and size, etc. The United Nations standards, i.e. 10% of the total rural population, and mathematical illustrations were followed to determine population and household sample sizes. The sample proportions were determined effectively, i.e. 760 households and 88,350 inhabitants, which backed in conducting a questionnaire survey and interviews from household heads. Henceforth, modified probability sampling technique was adopted for the survey proceedings, and the number of 100 questionnaires was filled from the study area along with interviews from household heads.

Keywords: Sample Size, Household Interview, Modified Probability Sampling, Questionnaire Survey, Transport Accessibility, Rural Sub-Regions.

1. INTRODUCTION

Transport accessibility may be defined as, “*ease or comfort in reaching desired destinations* (Hansen, 1959).” Transport inaccessibility can be considered as a pressing problem of the rural population, which troubled rural inhabitants in completion of their routine activities efficiently. The chronic socioeconomic problems of rural sub-regions, including inaccessibility erupted, because of the absence of transport services and suitable infrastructure (Olsson, 2006). There could be many reasons for pre-defined issues, but the prominent are the absence of data and policy guidelines. In this regard, to know the ground realities, questionnaire survey can really assist researchers to know the hidden truths about the study area. Thus, before the commencement of survey proceedings, it is mandatory to identify the number of targeted households and inhabitants. Therefore, the aim of this study is to determine the standard population and household sample sizes for questionnaire survey.

This research was conducted with the conviction and belief that pre-defined study goal could play a central role in survey proceedings, planned for the collection of data from the study area. To obey the study objective, varieties of data-collection and survey techniques were reviewed (Talpur, Napih, Chandio, & Khahro, 2012). Consequently, depending on the research objective, subject matters and research

limitations; the central idea of this study involved the review of various data collection or survey techniques applicable in rural sub-regions.

Researchers always collected data for their research purposes, and this study served the purpose of providing socioeconomic and travel data of rural households for transport accessibility measures. It is a grave reality that data collection is complex exercise, when it would be collected from the scattered rural population, which spread over thousands of square-kilometers. *Sampling* is a valuable data collection tool, which assisted researchers over the years to collect authentic data in tough situations. The sampling can be further classified into two major categories; explicitly, “*probability*” and “*non-probability sampling*.” Therefore, the proper technique was selected for this study comprised of these both aforementioned sampling methods.

2. MATERIAL AND METHODS

The sub-region of Badin was considered as one of the most deprived sub-regions of Pakistan (District Vision Badin-A Framework for Sustainable Development, 2006; (Jamal and Lohano, 2008). The sub-region originated its title from paramount township, Badin. It was positioned between 24° 13' and 25° 12' north latitudes, and 68° 21' and 69° 20' east longitudes. The Badin sub-region included five Talukas; namely, *Badin, Matli, Golarchi (Shaheed Fazil Rahu), Talhar*

⁺⁺Corresponding Author: Mir Aftab Hussain Talpur, Email: aftab.talpur@faculty.muett.edu.pk H/P: +92-333-2735205

*Department of Industrial Engineering & Management Mehran University of Engineering and Technology, Jamshoro, Sindh, Pakistan.

**Department of Civil & Environmental Engineering, Universiti Teknologi PETRONAS, Malaysia

and *Tando Bago*. There were six settlements selected from the sub-region of Badin; specifically, *Matli*, *Tando Ghulam Ali*, *Talhar*, *Badin*, *Tando Bago* and *Golarchi* for the questionnaire survey. The hundred questionnaires were filled from these six settlements, which can be seen in (Fig. 1), highlighted with sky blue rectangles.

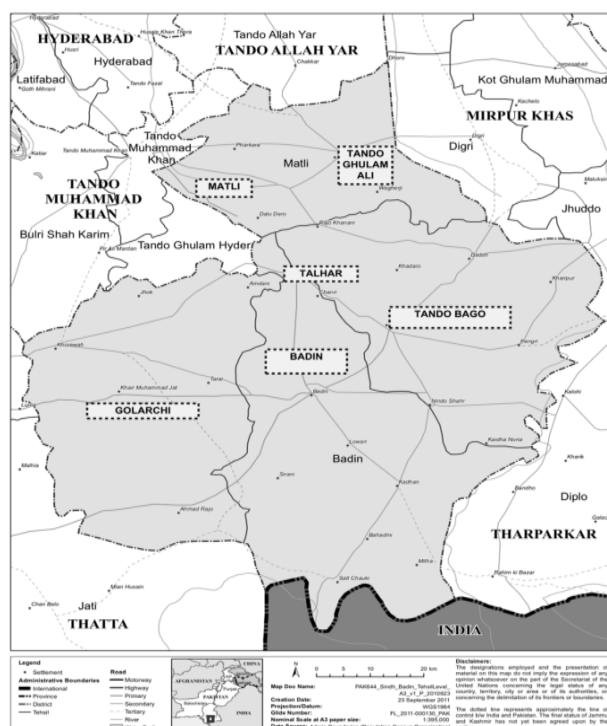


Fig. 1. Badin sub-region location map with survey areas

Different survey sites were exemplified in (Fig. 1), where the questionnaire survey was conducted for the primary data-collection.

2.1 Sampling Methods

By keeping in view the research constraints, limitations and objectives, the following methods were selected for the purpose of reliable data retrieval:

2.1.1 Modified Probability Sampling

The modified probability sampling method can be used to make the sampling process convenient and efficient (Sangngam & Suwattee, 2010). This technique actually has provided liberty to use both appropriate methods from the probability and non-probability sampling. Therefore, cluster sampling from the probability and quota sampling of the non-probability sampling categories were selected for this study.

2.1.1.1 Cluster Sampling

In this typology of sampling, different groups (*clusters*) can be selected rather than a single unit (Picquelle and Mier, 2011). The cluster sampling may

help researchers from various aspects, as if cost reduction, especially when the location of the population is scattered and personal interviews will have to be conducted (Horney, *et al.*, 2012). This sampling type assisted in a household sampling process based on income groups. The cluster sampling just requires the list of all-units exist in sampled clusters (Anderson, *et al.*, 2011).

The outstanding example of a cluster sampling of any geographic area based on frames or neighborhoods that can be viewed in (Fig. 2).

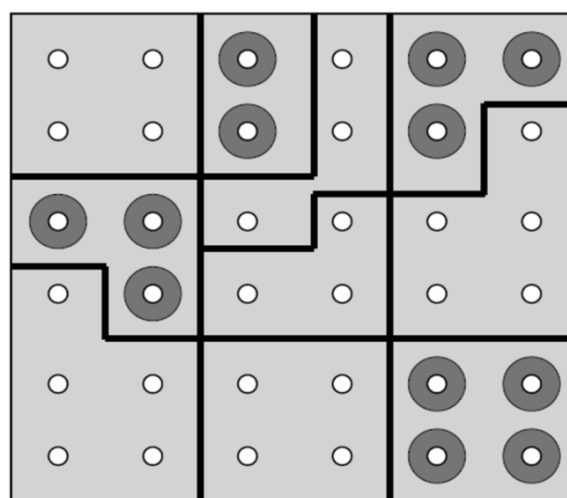


Fig. 2. Cluster sampling (Fellegi, 2010)

Before going to survey, it was mandatory to know the sampled population and household number, which is the prime objective of this research. Therefore, the sampled population and households were determined to manage the survey exercise proficiently.

3. RESULTS AND DISCUSSION

Results were determined with respect to standard sample population and household sizes. These sample sizes are based on *United Nations (UN)* standards, which she follows for the steering survey studies from rural sub-regions of developing countries.

3.1 Standard Population Sample

The standard level is (5 – 10%) of the total population that can be taken as a sample of the total population (Turner *et al.*, 2008). Therefore, let us calculate the target population that can be sampled from the total number of inhabitants. The entire population of the Badin sub-region was about 1,177,997 inhabitants in the year, 2012. It was expected that 75% of the whole lived-in rural areas, and depending mostly on the agriculture sector. Consequently, 75% of 1,177,997 = 883,500 is the target population; hence, by taking maximum level 10% of 883,500 = 88,350. This is the

standard number of the sampled population that must be taken into account while conducting the survey.

3.2 Household Sample Calculation

The household sampling is an important exercise and must be conducted to cope-up with the ground realities (Bostoen & Chalabi, 2006). Therefore, it is compulsory for interviewers to estimate the expected number of households and population samples for smooth survey proceedings. The equation 1 considers the limits of household sample calculations.

$$n_h = \frac{(z^2)(r)(1-r)(f)(k)}{(p)(\bar{n})(e^2)(1)} \quad (1)$$

Where,

n_h = Household Sample Size

z = Static Confidence level

r = Key Indicator Estimator

f = Sample Design Effect, $deff = 2$, Default Value

k = Nonresponse Rate Multiplier

p = Proportion of the Total Population Accounted by Target Population on which (r) is Based

$\bar{n}$ = Average Household Size

e = Margin of Error

To facilitate the calculation process, the general format of the equation 1 was referred, which can be reviewed as follows:

$$n_h = \frac{(3.84)(1-r)(1.2)(1.1)}{(r)(p)(6)(0.01)} \quad (2)$$

Hence, the equation 2 can be further expressed as follows:

$$n_h = \frac{(84.5)(1-r)}{(r)(p)} \quad (3)$$

Therefore, (n_h) is the parameter to be calculated, which was computed henceforth with the help of equation 3.

3.2.1 Household Sample Calculation

The main indicator of a questionnaire survey was to know the proportion of inhabitants from the entire rural population having socioeconomic and accessibility problems.

In such case the highest proportion was taken, i.e. 10% of the total rural population. Therefore,
 $r = 10\% = 0.1$

If the target population would be the whole population, the value of (p) would be taken as one, and sometimes this can be ignored.

The sample estimation was based on the 10% margin of error; therefore,

$$e = 0.10 \quad r = 0.10 \times 0.10 = 0.01$$

$$e = 0.01$$

So, applying these values in the equation 3 as follows:

$$n_h = \frac{(84.5)(1-r)}{(r)(p)}$$

$$= \frac{(84.5)(1-r)}{(r)(p)}$$

$$= \frac{(84.5)(1-0.1)}{(0.1)(1)}$$

$$= \frac{(84.5)(0.9)}{(0.1)}$$

$$= \frac{(84.5)(0.9)}{(0.1)}$$

$$= \frac{(76.05)}{(0.1)}$$

$$n_h = 760.5 \text{ (Household Sample)}$$

Thus, approximately the number of 760 households was required for the questionnaire survey. The sample population was calculated according to available standards.

Table 1 shows the definite number of target inhabitants, including the sum of questionnaires filled from the study area.

Table 1. Sampled Population (Estimated)

Taluka	Total Population on 2012	75 % Of Total Lives In Rural Areas	10% Sample Standard Selection	Number Of Sampled Households (Questionnaires)
Badin	300,707	225,530	22,553	23
Talhar	145,868	109,400	10,940	20
Matli	255,990	191,992	19,200	20
Tando Bago	283,424	212,570	21,257	15
Golarahi	192,008	144,005	14,400	22
Total	1,177,997	883,497	88,350	100

To comply with the determined standard household and population sample sizes, the modified probability method was adopted to take advantage of cluster and quota sampling techniques. Clusters from the study area were chosen, which have at least 100 residential units and possessed similar socioeconomic characteristics. As, one questionnaire would be able to represent 100 households and approximately 1,000 inhabitants. Similarly, 100 questionnaires would exemplify 10,000 households and 100,000 persons of the study area. With the help of quota sampling, this proportion actually did take into account to satisfy with the standard levels of

population and sampled households. Thus, the proportions were made and cluster-sampling methodologies were implemented.

It was revealed from the questionnaire survey and interviews from the household heads that the rural population of the Badin sub-region was struggling from the transport accessibility and socioeconomic aspects. Because of the unavailability of public transport facilities and dilapidated road network infrastructure, local inhabitants were facing difficulties in the completion of their routine activities. Delayed travel activities and long commuting time enhanced the level of inaccessibility, which put negative effects on household income and the whole economy of the sub-region. Therefore, the data were collected from rural households, as an individual-based accessibility measure and rural travel demand modeling techniques could be implemented to diminish the accessibility problems of local residents.

4. **CONCLUSION**

The purpose of this study was only to determine the standard household and population sample sizes for the questionnaire survey proceedings. The objective of this research was successfully achieved, and the standard household and population sample sizes, i.e. 760 households and 88,350 inhabitants were determined. With the help of modified probability sampling, proportion was made to comply with the computed standard household and population sample number, during the questionnaire survey proceedings. The number of 100 questionnaires was filled from six different settlements of the study area, together with personal interviews from the 100 household heads of targeted residential neighborhoods. The outcome of the questionnaire survey highlighted the transport accessibility issues of rural households, i.e. delayed travel activities, trembling socioeconomic conditions and its impacts on the travel decision-making process. In addition, this should also be noted that the collected data were then utilized in the development of individual-based accessibility measure and rural travel demand modeling, which is out of the scope of this study.

ACKNOWLEDGMENT

The authors are grateful to the students of the 11-Batch, Department of City and Regional Planning, Mehran University of Engineering and Technology, Jamshoro for their helping-hands in the household questionnaire survey exercise

REFERENCES:

Anderson, D. R., D. J. Sweeney, T. A. Williams, (2011). *Essentials of Statistics for Business and Economics*: Cengage Learning.

Bostoen, K., Z. Chalabi, (2006). Optimization of Household Survey Sampling without Sample Frames. *International Journal of Epidemiology*, 35(3), 751-755. doi:10.1093/ije/dyl019.

District Vision Badin-A Framework for Sustainable Development. (2006). International Union for Conservation of Nature and Natural Resources. The World Conservation Union (IUCN), Pakistan, Sindh Programme Office, 1-Bath Island Road, Karachi-75530 Retrieved from www.sindh.iucn.org.

Fellegi, I. P. (2010). *Survey Methods and Practices*. Ottawa, Canada: Statistics Canada.

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, 35(02), 73-76.

Horney, J., M. E. Zotti, A. Williams, J. Hsia, (2012). Cluster Sampling with Referral to Improve the Efficiency of Estimating Unmet Needs among Pregnant and Postpartum Women after Disasters. *Women's Health Issues*, 22(3), e253-e257. doi:http://dx.doi.org/10.1016/j.whi.2012.01.002.

Jamal, H., H. R. Lohano, (2008). *Understanding Rural Poverty Dynamics: The Case of the Poorest District of Sindh, Pakistan*.

Olsson, J. (2006). Responses to change in accessibility : Socio-economic impacts of road investment: The distributive outcomes in two rural peripheral Philippine municipalities. (2006).

Picquelle, S. J., K. L. Mier, (2011). A Practical Guide to Statistical Methods for Comparing Means from Two-Stage Sampling. *Fisheries Research*, 107(1-3), 1-13. doi:http://dx.doi.org/10.1016/j.fishres.2010.09.009.

Sangngam, P., P. Suwattee, (2010). Modified sampling scheme in inverse sampling without replacement. Paper presented at the International Conference on Networking and Information Technology (ICNIT).

Shah, S. M. H., A. A. Muhammad, K. S. Hussain, (2012). Implementation of Devolution Plan; Variations in the Perceptions of District Education Officers of Pakistan. *Journal of Research and Reflections in Education*, 1 (2), 30-39.

Talpur, M. A. H., M. Napiah, I. A. Chandio, S. H. Khahro, (2012). Transportation Planning Survey Methodologies for the Proposed Study of Physical and Socio-economic Development of Deprived Rural Regions: A Review. *Modern Applied Science*, 06(07), 1-16. doi:10.5539/mas.v6n7p1

Turner, A., I. Yansaneh, M. Jambwa, J. Banda, C. Menozzi, B. Kassa, P. Benjamin, (2008). *Designing Household Survey Samples: Practical Guidelines*.