TRAINING MANUAL

On

Survey Sampling Methods

Krishi Vigyan Kendra: Nagaland University
Lumami-798627
District-Zunheboto : Nagaland

CONTENTS

INTRODUCTION

Sampling is a technique by which a part of the population is selected and results from this fraction are generalized on the whole population from which the part or sample was selected. A sampling method is a scientific and objective procedure of selecting units from the population and provides a sample that is expected to be representative of the population as a whole. Some of the advantages of sample surveys as compared to complete enumeration are **reduction in cost, greater speed, wider scope and higher accuracy.** After deciding to go for sample survey, the next question is how to draw the sample from the population.

PROCEDURE OF SELECTING A RANDOM SAMPLE:

Since probability sampling theory is based on the assumption of random sampling, the technique of random sampling is of basic significance. Some of the procedures used for selecting a random sample are as follows:

(i) Lottery Method
(ii) Use of Random Number Tables
(iii) Remainder Approach
(iv) Quotient Approach

Lottery Method: Each unit in the population may be associated with a chit/ticket such that each sampling unit has its identification mark from 1 to N. All the chits/tickets are placed in a container, drum or metallic spherical device, in which a thorough mixing is possible before each draw. Chits/tickets may be drawn one by one and may be continued until a sample of the required size is obtained. When the size of population is large, this procedure of numbering units on chits/tickets and selecting one after reshuffling becomes cumbersome. In practice, it may be too difficult to achieve a thorough shuffling. Human bias and prejudice may also creep in this method.

Use of Random Number Tables: A random number Table is an arrangement of digits 0 to 9, in either a linear or rectangular pattern, where each position is filled with one of these digits. A Table of random numbers is so constructed that all numbers 0, 1, 2, ..., 9 appear independent of each other. Some random number Tables in common use are

- ✓ Tippett's random number Tables
- ✓ Fisher and Yates Tables
- ✓ Kendall and Smith Tables
- ✓ A million random digits Table

A practical method of selecting a random sample is to choose units one-by-one with the help of a Table of random numbers. By considering two-digit numbers, we can obtain numbers from 00 to 99, all having the same frequency. Similarly, three or more digit numbers may be obtained by combining three or more rows or columns of these Tables. The simplest way of selecting a sample of the required size is by selecting a random number from 1 to N and then taking the unit bearing that number. This procedure involves a number of rejections since all numbers greater than N appearing in the Table is not considered for selection. Suppose N= 128 and we want n= 10, then we have to select three columns from the random number table and go down to columns and use to select 10 distinct number between 001 and 128. This procedure involves a number of rejections since all numbers greater than 128 appearing in the Table are not considered for selection. A second method that involves less rejection and easily applies is as follows: in a series of three digit numbers, subtract 200 from all numbers between 201 and 400, 400 from all numbers between 401 and 600, 600 from all numbers between 601 and 800, 800 from all numbers between 801 and 999 and, of course, 000 from all numbers between 000 and 200. All remainders greater than 129 and the numbers 000, 200, and so forth, are rejected.

Remainder Approach: Let N be a r-digit number and let its r-digit highest multiple be N'. A random number k is chosen from 1 to N' and the unit with the serial number equal to the remainder obtained on dividing k by N is selected. If the remainder is zero, the last unit is selected. As an illustration, let N = 123, the highest three-digit multiple of 123 is 984. For selecting a unit, one random number from 001 to 984 has to be selected. Let the random number selected be 287. Dividing 287 by 123, the remainder is 41. Hence, the unit with serial number 41 is selected in the sample.

Quotient Approach: Let N be a r-digit number and let its r-digit highest multiple be N' such that N` / N = d. A random number k is chosen from 0 to (N'-1). Dividing k by d, the quotient q is obtained and the unit bearing the serial number (q - 1) is selected in the sample. As an illustration, let N = 16 and hence N' = 96 and d = 96 / 16 = 6. Let the two-digit random number chosen be 65 which lies between 0 and 95. Dividing 65 by 6, the quotient is 10 and hence the unit bearing serial number (10- 1) = 9 is selected in the sample.

Hence, a commonly used device i.e. **remainder approach,** is employed to avoid the rejection of such large numbers. For example, select a random sample of 11 households from a list of 112 households in a village, the greatest three-digit multiple of 112 is 896. By using three digit random numbers, selected numbers are 310, 145, 497, 769, 387, 164, 514, 218, 799, 554, and 580 then the sample will comprise of households with serial numbers 086, 033, 049, 097, 051, 052, 066, 107, 015, 106 and 020. In case quotient approach is applied, the 3-digit multiple of 112 is 896 and 896/112 = 8. Using the same random number and dividing them by 8, we have the sample of households with list numbers 037, 017, 061,095, 047, 019, 063, 026, 098, 068 and 071.

NEED FOR A SAMPLE:

Collection of information on every unit in the population for the characteristics of interest is known as **complete enumeration or census.** The money and time required for carrying out a census will generally be large, and there are many situations where with limited means complete enumeration is not possible. There are also instances where it is not feasible to enumerate all units due to their perishable nature. In all such cases, the Investigator has no alternative except resorting to a sample survey. The number of units (not necessarily distinct) included in the sample is known as the **sample size** and is usually denoted by **'n'**, whereas the number of units in the population is called population size and is denoted by **'N'**. The ratio **n/N** is termed as sampling fraction.

There are certain advantages of a sample survey over complete enumeration, which are as follows:

Greater Speed

The time taken for collecting and analyzing the data for a sample is much less than that for a complete enumeration. Often, we come across situations where the information is to be collected within a specified period. In such cases, where time available is short or the population is large, sampling is the only alternative.

Greater Accuracy

A census usually involves a huge and unwieldy organization and, therefore, many types of errors may creep in. Sometimes, it may not be possible to control these errors adequately. In sample surveys, the volume of work is considerably reduced. On account of this, the services of better trained and efficient staff can be obtained without much difficulty. This will help in producing more accurate results than those for complete enumeration.

More detailed Information

As the number of units in a sample are much less than those in census, it is, therefore, can be obtained on more number of variables. However, in complete enumeration, such an effort becomes comparatively difficult.

Reduced Cost

Because of lesser number of units in the sample in comparison to the population, considerable time, money, and energy are saved in observing the sample units in relation to the situation where all units in the population are to be covered. From the above, it may be seen that the sample survey is more economical, provides more accurate information, and has greater scope in subject coverage as compared to a complete enumeration. It may, however, be pointed out that sampling errors are present in the results of the sample surveys. This is due to the fact that only a part of the whole population is surveyed. On the other hand, non sampling errors are likely to be more in case of a census study than these are in a sample survey.

DEFINITION AND CONCEPTS USED IN SAMPLING:

Elements: An element is a unit of a population for which information is sought. For example, in a household fertility survey, women in the reproductive ages are usually the ultimate elements. To facilitate data collection in a survey, it is absolutely essential that elements be well defined and physically easy to identify.

Population or universe: The population or universe is the aggregate of elements. Elements are therefore the basic units that make up and define the population. It is essential to define the population in terms of:

- ✓ Content, which calls for the definition of the type and characteristics of the elements that make up population

- ✓ Extent, which refers to the geographical boundaries as they relate to coverage
- ✓ Time, which would refer to the time period for which the population exists.

Observational units: These are units from which the observations are obtained. In interview surveys, they are called respondents.

Reporting units: are elements that report the solicited information in a survey. Note that in some cases observational and reporting units may be different. For example, in a survey of children under age 5, parents will normally give, as proxies, information pertaining to their children. In such cases, selected children, in the sample, are observational units, while parents are reporting units.

Sampling units: Sampling units are used for selecting elements for inclusion in the sample. In element sampling, each sampling unit contains one element, while in cluster sampling, for instance, a sampling unit comprises a group of elements called a cluster. For example, an enumeration area (EA) would, as a first stage sampling unit, contain a cluster of households. It is possible for the same survey to use different sampling units.

Sample units: Selected sampling units may be termed sample units and the values of the characteristics under study for the sample units are known as sample observations.

Unit of analysis: This is a unit used at the stage of tabulation and analysis. Such a unit may be an elementary unit or group of elementary units. It should be noted that the unit of analysis and the reporting unit need not necessarily be identical.

Sampling frame: The sampling frame is used to identify and select sampling units into the sample and is also used as a basis for making estimates based on sample data. This implies that the population

from which the sample has to be selected must be represented in a physical form. The frame ideally should have all sampling units belonging to the population under study with proper identification particulars. Frames should be exhaustive and preferably mutually exclusive

The commonly used types of frames in surveys are list, area and multiple frames.

A List frame: A list frame contains a list of sampling units from which a sample can be directly selected. It is preferable that the frame should have relevant and accurate information on each sampling unit such as size and other characteristics. The additional information helps in designing and/ or selecting efficient samples.

Area frames: Area frames are multistage frames that are, in general, commonly used in household surveys. In this connection, the frame consists of one or more stages of area units. In a two stage sample design, for example, the frame will consist of clusters, which can be called primary sampling units (PSUs); in selected PSUs, a list of households becomes the second stage frame. In general, frames are needed for each stage of selection. The durability of the frame declines as one move down the hierarchy of the units.

Area units: Area units cover specified land areas with clearly defined boundaries, which can be physical features such as roads, streets, rivers, rail lines, or imaginary lines representing the official boundaries between administrative divisions. Census enumeration areas are usually established within the smaller administrative units that exist in a country.

The frame or frames: Used for a household survey should be able to provide access to all the sampling units in the survey population

so that every unit has a known and non zero probability of selection in the sample. Access can be achieved by sampling from the frames, usually through two or more stages of selection. The frame for the first stage of sampling must include all the designated sampling units. At subsequent stages of sample selection frames are needed only for the sample units selected at the preceding stage.

Random Sample: One or more sampling units selected from a population according to some specified procedures are said to constitute a sample. The sample will be considered as random or probability sample, if its selection is governed by ascertainable laws of chance. In other words, a random or probability sample is a sample drawn in such a manner that each unit in the population has a predetermined probability of selection. For example, if a population consists of the N sampling units U1, U2,…,Ui,…,UN then we may select a sample of n units by selecting them unit by unit with equal probability for every unit at each draw with or without replacing the sampling units selected in the previous draws.

Non-random Sample: A sample selected by a non-random process is termed as non-random sample. A Non-random sample, which is drawn using certain amount of judgment with a view to getting a representative sample, is termed as judgment or purposive sample. In purposive sampling, units are selected by considering the available auxiliary information more or less subjectively with a view to ensuring a reflection of the population in the sample. This type of sampling is seldom used in large-scale surveys mainly because it is not generally possible to get strictly valid estimates of the population parameters under consideration and of their sampling errors due to the risk of bias in subjective selection and the lack of information on the probabilities of selection of the units.

Population Parameters: Suppose a finite population consists of the N units U1, U2,…,UN and let Yi be the value of the variable y, the characteristic under study, for the i-th unit Ui, (i=1,2,…,N). For instance, the unit may be a farm and the characteristic under study may be the area under a particular crop. Any function of the values of all the population units (or of all the observations constituting a population) is known as a population parameter or simply a parameter.

Statistic, Estimator and Estimate: Suppose a sample of n units is selected from a population of N units according to some probability scheme and let the sample observations be denoted by y1,y2,… ,yn. Any function of these values which is free from unknown population parameters is called a statistic. An estimator is a statistic obtained by a specified procedure for estimating a population parameter. The estimator is a random variable and its value differs from sample to sample and the samples are selected with specified probabilities. The particular value, which the estimator takes for a given sample, is known as an estimate.

Sampling and Non-sampling Error: The error arising due to drawing inferences about the population on the basis of observations on a part (sample) of it, is termed sampling error. The sampling error is non-existent in a complete enumeration survey since the whole population is surveyed. The errors other than sampling errors such as those arising through non-response, incompleteness and inaccuracy of response are termed non-sampling errors and are likely to be more wide-spread and important in a complete enumeration survey than in a sample survey. Non-sampling errors arise due to various causes right from the beginning stage when the survey is planned and designed to the final stage when the data are

processed and analyzed. The sampling error usually decreases with increase in sample size (number of units selected in the sample) while the non-sampling error is likely to increase with increase in sample size. As regards the non-sampling error, it is likely to be more in the case of a complete enumeration survey than in the case of a sample survey since it is possible to reduce the non-sampling error to a great extent by using better organization and suitably trained personnel at the field and tabulation stages in the latter than in the former.

SAMPLING PROCEDURES:

(i) Probability Sampling
(ii) Non Probability Sampling

Probability Sampling: If the unit in the sample selected using some probability mechanism, such a procedure is called Probability Sampling. There are following type of Probability Sampling:

- ✓ Simple Random Sampling
- ✓ Stratified Random Sampling
- ✓ Cluster Sampling
- ✓ Systematic Sampling

Simple Random Sampling (SRS): Simple random sampling (SRS) is a probability sample selection method where each element of the population has an equal chance/probability of selection. Selection of the sample can be with or without replacement. It can be regarded as the basic form of probability sampling applicable to situations where there is no previous information available on the population structure. SRS is attractive by virtue of its being simple in terms of selection and estimation procedures (for example, of sampling errors). To select n units out of N such that each of ${}^{N}C_{n}$ has an equal chance of being selected.

It is easily verified that that all ${}^{N}C_{n}$ distinct samples have an equal chance being selected by this method. Consider one distinct sample, that is, one set of n specified units. At the first draw the probability that some one of the n specified units is selected is n/N. At the second draw the probability that some one of the remaining (n-1) specified units is drawn is (n-1) / (N-1), and so on. Hence the probability that all n specified units are selected in n draws is

$$n/N \times (n-1)/(N-1) \times (n-2)/(N-2) \times \ldots \times 1/(N-n+1)$$
$$= n'(N-n)'/(N)' = 1 / NCn$$

Since a number that has been drawn is removed from the population for all subsequent draws, this method is also called as Simple Random Sampling Without Replacement.

The examples below illustrate the calculation of the probability of selection under SRS:

1. First we consider a finite population of 100 households H_1, H_2, … … H_i, … … H_{100} with income values X_1, X_2, … … X_i, … … X_{100}.

   In this example, the probability of any particular unit's being selected is 1/100

2. As a second example, we note that in order to draw a sample of households, the target households can be numbered serially in a frame/list. Using random numbers, a sample of, say, size 25 can be selected. For the equal probability selection method (EPSEM) *f* is the overall sampling fraction for the elements.

   Thus, f = n/N

   If n =25, the sample size, and N=100, the total number of households, then the sampling fraction, which is the probability of selection, is 25/100 = 1/4

Types of sample selection under simple random sampling :

There are two common methods of sample selection under simple random sampling, namely:

(a) Simple random sampling with replacement (SRSWR)
(b) Simple random sampling without replacement (SRSWOR)

(a) Simple random sampling with replacement: Simple random sampling with replacement is based on random selection from a population carried out by replacing the chosen element after each draw that means returning of selected unit to the population before next draw in the population. The probability of selection of an element remains unchanged after each draw, and any selected independent samples are independent of each other. This procedure gives rise to N^n possible samples when order of selection of units in the sample taken into account, where N and n denote the population and sample respectively. SRSWR consists of N^n possible sample, probability of selecting a sample of n size under SRSWR is $p(s) = 1/N^n$.

$p(s) = 1/N \times 1/N \times 1/N \times .... \times 1/N = 1/ N^n$

One of the major drawbacks of SRSWR is that there are chances of repetition of unit in the selected sample.

Example:

Given below are the weights (in Kgs) of 4 participants of training Programme:

Participants	: A	B	C	D
Weights	: 50	75	60	65

Enumerate all possible WR Samples of size 2. Also write values of the study variables (Weight) for the sample units.

Solution

Here, N = 4 and n = 2. There will, therefore, be $4^2 = 16$ possible samples. There are enumerated below along with the weight values for the units included in the sample.

Table: Possible samples along with their variable value

Sample	Participants in the sample	Weight for sampled Participants	Sample	Participants in the sample	Weight for sampled Participants
1	A, A	50, 50	9	C, A	60, 50
2	A, B	50, 75	10	C, B	60, 75
3	A, C	50, 60	11	C, C	60, 60
4	A, D	50, 65	12	C, D	60, 65
5	B, A	75, 50	13	D, A	65, 50
6	B, B	75, 75	14	D, B	65, 75
7	B, C	75, 60	15	D, C	65, 60
8	B, D	75, 65	16	D, D	65, 65

(b) Simple random sampling without replacement (SRSWOR): SRSWR has the drawback that one or more sampling units occur more than once in a sample. In SRSWOR sampling, the units are selected one by one from the population, and the unit selected at any particular draw is not replaced back to the population before selecting a unit at the next draw. The simple random sampling without replacement strategy is therefore the most frequently used simple random sampling procedure. In this procedure, the selection process is continued until n distinct units are selected and all repetitions are ignored. The possible sample in the SRSWOR is ${}^{N}C_{n}$. And the probability of selecting a sample of n size under SRSWOR is $p(s) = 1/{}^{N}C_{n}$.

$p(s) = n/N \times (n-1)/(N-1) \times (n-2)/(N-2) \times \ldots \times 1/(N-n+1)$
$= n'(N-n)'/(N)' = 1 / {}^{N}C_{n}$

The following are some of the properties of simple random sampling without replacement:

- It gives a fixed sample size
- It results in equal probability of selection for every element/ unit (EPSEM)

Example:

Given below are the weights (in Kgs) of 4 participants of training Programme:

Participants	: A	B	C	D
Weights	: 50	75	60	65

Enumerate all possible WOR Samples of size 2. Also write values of the study variables (Weight) for the sample units.

Solution

Here, N = 4 and n = 2. There will, therefore, be ${}^{4}C_{2} = 6$ possible samples. There are enumerated below along with the weight values for the units included in the sample.

Table: Possible samples along with their variable value

Sample	Participants in the sample	Weight for sampled Participants	Sample	Participants in the sample	Weight for sampled Participants
1	A, A	50, 50	9	C, A	60, 50
2	A, B	50, 75	10	C, B	60, 75
3	A, C	50, 60	11	C, C	60, 60
4	A, D	50, 65	12	C, D	60, 65
5	B, A	75, 50	13	D, A	65, 50
6	B, B	75, 75	14	D, B	65, 75

7	B, C	75, 60	15	D, C	65, 60
8	B, D	75, 65	16	D, D	65, 65

Note: Red coloured samples will not be included in the sample because it has got repeated.

Stratified Random Sampling: In the Stratified Random Sampling method, the sampling units in the population are divided into groups called strata. Stratification is usually carried out so that the population is subdivided into heterogeneous groups that are internally homogeneous such as sex-yielding two strata: male and female, or upon a combination of two or more criteria such as sex and level of education and so on. Stratification is undertaken to ensure proper representation of important subpopulation groups without biasing the selection operation. Heterogeneity among strata and homogeneity within strata should thus constitute the primary feature that should guide the establishment of strata. It can therefore easily be seen feature why urban and rural areas are often established as two of the strata for a household survey, urban and rural populations are different from each other in many ways (type of employment, source and amount of income, average household size, fertility rates, etc.) while the members of the population within each subgroup are similar.

The heterogeneity feature is a useful guide in determining how many strata should be created. There should be no more strata than there are identifiable subpopulations for the particular criterion being used to define strata. For example, if a country is divided into eight geographical regions for administrative purposes and two of the regions are very much alike with respect to the subject matter of a proposed survey, an appropriate sample design could be accomplished by creating seven strata combining the two similar regions).

STRATIFIED RANDOM SAMPLING CAN BE DIVIDED INTO TWO TYPES:

(a) Proportionate Stratified Random Sampling
(b) Disproportionate Stratified Random Sampling

Proportionate Stratified Random Sampling: Proportionate allocation in stratified random sampling involves the use of a uniform sampling fraction in all strata. This implies that the same proportion of units is selected in each stratum. For example, if we decide to select a total sample of 10 per cent, this means that we shall select 10 per cent of units from each stratum. Since the sampling rates in all strata are the same, the sample elements selected in the sample will vary from stratum to stratum. Within each stratum, the sample size will be proportionate to the number of elements in the stratum.

In this case the sampling fraction is given by fh = nh / Nh = n / N implying an EPSEM design.

Example: There are 1000 students in a university, comprising 600 B.Sc, 300 M.Sc and 100 Ph.D students. Now investigator wants to draw a sample of 16 % students from these tree strata in similar proportion as they appear in the population.

Solution:

Strata	Number of Students	Proportion with population	Sampling fraction	Sample
B.Sc	600	0.6	0.16	96
M.Sc	300	0.3	0.16	48
Ph.D	100	0.1	0.16	16
Total	1000	1.00	0.16	160

Disproportionate Stratified Random Sampling: The method of disproportionate sampling involves the use of different sampling

rates in various strata. In Disproportionate Stratified Random Sampling, the samples drawn from each stratum are not necessarily distributed according to their proportion in the population from which they are randomly selected. In other words, some of the stratum of the population may be overrepresented and some underrepresented.

Cluster Sampling: This is random sampling in which sampling units are not individual elements of the population or universe, but groups of elements or cluster. In cluster sampling, we follow these steps:

- ✓ Divide population into clusters (usually along geographic boundaries)
- ✓ Select clusters Randomly
- ✓ Measure all units within sampled clusters

Example: A researcher wants to study the problems of rural school going children and wants a sample of 10 % children from the rural schools in the district. The researcher may select at random 10 % of rural school (Clusters) in the district and sample all the children attending those schools.

Systematic Sampling: Systematic sampling is a probability sample selection method in which the sample is obtained by selecting every *kth* element of the population where *k* is an integer greater than 1. The first number of the sample must be selected randomly from within the first *k* elements. The selection is made from an ordered list. This is a popular method of selection especially when units are many and are serially numbered from 1 to *N*. Suppose that *N*, the total number of units, is an integral multiple of the required sample size *n* and that *k* is an integer, such that $N = nk$ *or* $k = N/n$. A random number is then selected between 1 and *k*.

Here are the steps we need to follow in order to achieve a systematic sample:

- ✓ Number the units in the population from 1 to N
- ✓ Decide on the n (sample size) that you want or need
- ✓ k = N/n = the interval size
- ✓ Select an integer between 1 to k

Let us suppose 4 is the random start, and then the sample will be of size n with units serially numbered as follows:

4, 4 + k, 4 + 2k, ..., 4 + (n-1)k

N=100

want n = 20

N/n = 5

select a random number from 1-5: choose 4

start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

One of the major disadvantages of systematic sampling is that if once the first unit is selected randomly then for the next unit there are 100 % probability to be selected in the sample and for the unit falling between first to next, there is 0 % probability to get selected in the sample.

Non Probability Sampling: The procedure of selecting Sample without using any probability mechanism is termed as the Non Probability Sampling.

There are following important type of Non Probability Sampling:

- ✓ Quota Sampling
- ✓ Purposive Sampling
- ✓ Double Sampling
- ✓ Sampling for Hidden Population (Chain referral Sampling)
  - (i) Snowball sampling
  - (ii) Key Informants Sampling
  - (iii) Targeted Sampling
  - (iv) RD Sampling(Respondent Driven Sampling)

Quota Sampling: In quota sampling, we select units nonrandomly according to some fixed quota. There are two types of quota sampling: proportional and non proportional.

Proportional quota sampling: We want to represent the major characteristics of the population by sampling a proportional amount of each. For instance, if we know the population has 40% women and 60% men, and that we want a total sample size of 100, we will continue sampling until we get those percentages and then we will stop. So, if we've already got the 40 women for sample, but not the sixty men, we will continue to sample men but even if legitimate women respondents come along, we will not sample them because we have already "met our quota." The problem here is that we have to decide the specific characteristics on which we will base the quota. Will it be by gender, age, education race, religion, etc.

Non proportional quota sampling is a bit less restrictive. In this method, we specify the minimum number of sampled units we want in each category or some time we ignore the restriction on minimum number of sampled units we want in each category

rather will try to get required number of sample. Here, we are not concerned with having numbers that match the proportions in the population. This method is the non probabilistic analogue of stratified random sampling in that it is typically used to assure that smaller groups are adequately represented in your sample.

Purposive Sampling: This is also a non random sampling in which the sample is arbitrarily selected because the characteristics which they posses are deemed important for the research. In the Purposive Sampling, the investigator has some belief that the sample being selected as typical of the population or is a very good representative of the population. This is also known as judgment sampling. For studying attitude of the people towards the national issue, a sample of journalists, teachers and legislators may be chosen, which is example of purposive sampling. This method ensures that those individuals will be included in the samples that are relevant to the research design. Because of limitation of time and money, purposive sampling is the popular method for student research project.

Double Sampling: Double sampling is defined as drawing a small sample of individuals from a bigger sample of them. For example, a researcher wants to study the knowledge of newly married couple about the family planning through mailed questionnaire. For this purpose a questionnaire is mailed to 1000 couple residing in different localities. Only 50 percent, that is 500 questionnaires is returned. From these 500 questionnaires, the researcher draws a random sample of 100 and mails another set of questionnaire to get in-depth knowledge about family planning. This method is known as double sampling. This method has disadvantage of taking much time and labour of researcher.

Sampling for Hidden Population (Chain referral Sampling)

A population is "hidden" when no sampling frame exists and public acknowledgement of membership in the population is potentially threatening. Accessing such populations is difficult because standard probability sampling methods produces low response rate and responses that lack candor (authenticity). Hidden population have two characteristics: first, no sampling frame exists, so the size and the boundaries of the population are unknown: and second they exist in strong privacy concerns, because membership involves stigmatized or illegal behaviour, leading individuals to refuse to cooperate, or give unreliable answers to protect their privacy. Traditional methods, such as household survey, cannot produce reliable samples, and they are insufficient, because most hidden population are rare.

There are four methods for sampling hidden population:

1. Snowball sampling
2. Key informant sampling
3. Targeted sampling
4. RD Sampling.

Snowball Sampling: Snowball sampling or chain referral sampling is a method that has been widely used in qualitative sociological research. The method yields a study sample through referrals made among the people who share or know of others who possess some characteristics that are of research interest. This method is well suited for a number of research purposes and is particularly applicable when the focus of study in on sensitive issue, possibly concerning a relatively private matter, and thus requires the knowledge of insider to locate people for the study. In this, inferences about individuals must rely mainly on the initial sample, since additional individuals found by tracing chains are never found randomly or even with known biases. Referrals occur through network links, so subjects with larger personal network will be oversampled and relative isolates will be excluded.

Key Informant sampling: Key Informant sampling is designed to overcome response biases by selecting especially knowledgeable respondents and asking them about other behaviour rather than their own. Key informant do not interact with a random group of potentials biases. Key informant approach has limitations: it cannot be used to access highly detailed and personal information; and the sampling may also have an institutional bias if key informant is from the professional section.

Targeted Sampling: it involves two basic steps: first, field researchers map a target population (to the extent that may succeed in penetrating the local networks linking potential respondents, this prevents the under sampling that traditional approaches would produce); and second, field researchers recruit a pre specified number of subjects at sites identified by the ethnographic mapping, ensuring that subjects from different areas and sub- group will appear in the final stage. The adequacy of targeted sampling depends on the accuracy and comprehensiveness of the ethnographic mapping.

Respondent- Driven Sampling: RDS assumes that those best able to access members of hidden populations are their own peers. It differ from traditional snowball sampling in two respects; first, whereas snowball sampling typically involves an incentive for participation, RDS involve a dual incentives system- the reward for being interviewed (a primary reward) plus reward for recruiting others into the study (a secondary reward). This study also uses mix of material (monetary) and symbolic (the opportunity to help protect oneself and one's peers from a deadly epidemic) rewards. A second difference between RDS and Typical snowball sampling is that subjects are not asked to identify their peers to the investigator, but to recruit them into the study. This distinction is crucial when dealing with the hidden populations that are subjected to considerable repression.

DETERMINATION OF SAMPLE SIZE:

Some of the concept used in determination of sample size

Level of precision (A or e): The level of precision is the closeness with which the sample predicts where the true values in the population lie. The difference between the sample and the real population is called the sampling error. If the sampling error is ±3%, this means we add or subtract 3 percentage points from the value in the survey to find out the actual value in the population. For example, if the value in a survey says that 65% of farmers use a particular pesticide, and the sampling error is ±3%, we know that in the real-world population, between 62% and 68% are likely to use this pesticide. This range is also commonly referred to as the margin of error.

The level of precision you accept depends on balancing accuracy and resources. High levels of precision require larger sample sizes and higher costs to achieve those samples, but high margins of error can leave you with results that aren't a whole lot more meaningful than human estimation.

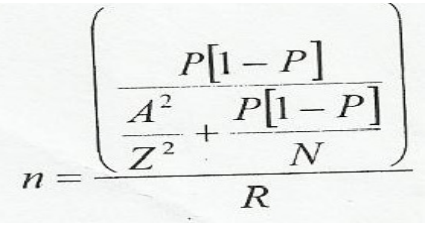
Confidence Level (Z): The confidence level involves the risk we're willing to accept that our sample is within the average or "bell curve" of the population. A confidence level of 90% means that, 90 of these samples would have the true population value within the range of precision, and 10 would be unrepresentative samples. Higher confidence levels require larger sample sizes. 95% confidence level is standard for most social-science applications, though higher levels can be used. If the confidence level that is chosen is too low, results will be "statistically insignificant".

Degree of Variability (P): Variability is the degree to which the attributes or concepts being measured in the questions are distributed throughout the population. A heterogeneous population divided more or less 50%-50% on an attribute or a concept, will be harder to measure precisely than a homogeneous population, divided say 80%-20%. Therefore, the higher the degree of variability we expect the distribution of a concept to be in your target audience, the larger the sample size must be to obtain the same level of precision. If we estimate that 25% of the population in your country farms organically and 75% does not, then your variability would be .25. If variability is too difficult to estimate, it is best to use the conservative figure of 50%.

Note: when the population is extremely heterogeneous (i.e., greater than 90-10), a larger sample may be needed for an accurate result, because the population with the minority attribute is so low.

Estimated the Response Rate (R): The base sample size is the number of responses you must get back when you conduct your survey. However, since not everyone will respond, you will need to increase your sample size, and perhaps the number of contacts you attempt to account for these non-responses. To estimate response rate that you are likely to get, you should take into consideration the method of your survey and the population

Formula for determination of Sample size:



$$n = \frac{\left(\frac{P[1-P]}{\frac{A^2}{Z^2} + \frac{P[1-P]}{N}} \right)}{R}$$

Where:

n = sample size required

N = number of people in the population

P = estimated variance in population, as a decimal: (0.5 for 50-50, 0.3 for 70-30)

A = Precision desired, expressed as a decimal (i.e., 0.03, 0.05, 0.1 for 3%, 5%, 10%)

Z = Based on confidence level: 1.96 for 95% confidence, 1.6449 for 90% and 2.5758 for 99%

R = Estimated Response rate, as a decimal

Yamane (1967) provides a simplified formula to calculate sample sizes.

$$n = \frac{N}{1 + N(e)^2}$$

BIBLIOGRAPHY:

Cochran, William G. (1977): Sampling Techniques. Third Edition. John Wiley and Sons.

Des Raj (1968): Sampling Theory. TATA McGRAW-HILL Publishing Co. Ltd.

Heckathorn, D.D. (1997): Respondent- Driven Sampling: A New approach to study of Hidden Population. Social problem, Vol. 44 (2). Accessed on 22.05.2013

Mukhopadhyay, Parimal (2009): Theory and Methods of Survey Sampling. Second Edition. PHI

Ray, G.L. and S. Mondal (1999): Research Methods in Social Sciences and Extension Education .Kalyani Publishers

Singh, Daroga, Singh, Padam and Pranesh Kumar (1978): Handbook of Sampling Methods. I.A.S.R.I., New Delhi.

Singh Ravindra and Mangat N.S. (1996): Elements of Survey Sampling, Kluwer Academic Publishers.

Sukhatme, P.V. and Sukhatme, B.V. (1970): Sampling Theory of Surveys with Application. Second Edition. Iowa State University Press, USA.

Trochim, W. and James P. Donnelly. Research Methods Knowledge Base. Accessed on 10.08.2011

Tyagi, K. K. (2012): Recent Advances in Sample Survey and Analysis of Survey Data Using Statistical Softwares. Reference Manual-1, IASRI, New Delhi.

UN Publication (2008): Designing Household Survey Samples: Practical Guidelines. Series F No. 98. Accessed on 09.05.2013

www.extension.psu.edu/evaluation/pdf/TS60.pdf. Accessed on 22.05.2013

www.rand.org/publications/classics/randomdigits. Accessed on 09.11.2012

TABLE 1 - RANDOM DIGITS

11164	36318	75061	37674	26320	75100	10431	20418	19228	91792
21215	91791	76831	58678	87054	31687	93205	43685	19732	08468
10438	44482	66558	37649	08882	90870	12462	41810	01806	02977
36792	26236	33266	66583	60881	97395	20461	36742	02852	50564
73944	04773	12032	51414	82384	38370	00249	80709	72605	67497
49563	12872	14063	93104	78483	72717	68714	18048	25005	04151
64208	48237	41701	73117	33242	42314	83049	21933	92813	04763
51486	72875	38605	29341	80749	80151	33835	52602	79147	08868
99756	26360	64516	17971	48478	09610	04638	17141	09227	10606
71325	55217	13015	72907	00431	45117	33827	92873	02953	85474
65285	97198	12138	53010	94601	15838	16805	61004	43516	17020
17264	57327	38224	29301	31381	38109	34976	65692	98566	29550
95639	99754	31199	92558	68368	04985	51092	37780	40261	14479
61555	76404	86210	11808	12841	45147	97438	60022	12645	62000
78137	98768	04689	87130	79225	08153	84967	64539	79493	74917
62490	99215	84987	28759	19177	14733	24550	28067	68894	38490
24216	63444	21283	07044	92729	37284	13211	37485	10415	36457
16975	95428	33226	55903	31605	43817	22250	03918	46999	98501
59138	39542	71168	57609	91510	77904	74244	50940	31553	62562
29478	59652	50414	31966	87912	87154	12944	49862	96566	48825
96155	95009	27429	72918	08457	78134	48407	26061	58754	05326
29621	66583	62966	12468	20245	14015	04014	35713	03980	03024
12639	75291	71020	17265	41598	64074	64629	63293	53307	48766
14544	37134	54714	02401	63228	26831	19386	15457	17999	18306
83403	88827	09834	11333	68431	31706	26652	04711	34593	22561
67642	05204	30697	44806	96989	68403	85621	45556	35434	09532
64041	99011	14610	40273	09482	62864	01573	82274	81446	32477
17048	94523	97444	59904	16936	39384	97551	09620	63932	03091
93039	89416	52795	10631	09728	68202	20963	02477	55494	39563
82244	34392	96607	17220	51984	10753	76272	50985	97593	34320
96990	55244	70693	25255	40029	23289	48819	07159	60172	81697
09119	74803	97303	88701	51380	73143	98251	78635	27556	20712
57666	41204	47589	78364	38266	94393	70713	53388	79865	92069
46492	61594	26729	58272	81754	14648	77210	12923	53712	87771
08433	19172	08320	20839	13715	10597	17234	39355	74816	03363
10011	75004	86054	41190	10061	19660	03500	68412	57812	57929
92420	65431	16530	05547	10683	88102	30176	84750	10115	69220
35542	55865	07304	47010	43233	57022	52161	82976	47981	46588
86595	26247	18552	29491	33712	32285	64844	69395	41387	87195
72115	34985	58036	99137	47482	06204	24138	24272	16196	04393
07428	58863	96023	88936	51343	70958	96768	74317	27176	29600
35379	27922	28906	55013	26937	48174	04197	36074	65315	12537
10982	22807	10920	26299	23593	64629	57801	10437	43965	15344
90127	33341	77806	12446	15444	49244	47277	11346	15884	28131
63002	12990	23510	68774	48983	20481	59815	67248	17076	78910
40779	86382	48454	65269	91239	45989	45389	54847	77919	41105
43216	12608	18167	84631	94058	82458	15139	76856	86019	47928
96167	64375	74108	93643	09204	98855	59051	56492	11933	64958
70975	62693	35684	72607	23026	37004	32989	24843	01128	74658
85812	61875	23570	75754	29090	40264	80399	47254	40135	69916

TABLE 2 – RANDOM DIGITS

40603	16152	83235	37361	98783	24838	39793	80954	76865	32713
40941	53585	69958	60916	71018	90561	84505	53980	64735	85140
73505	83472	55953	17957	11446	22618	34771	25777	27064	13526
39412	16013	11442	89320	11307	49396	39805	12249	57656	88686
57994	76748	54627	48511	78646	33287	35524	54522	08795	56273
61834	59199	15469	82285	84164	91333	90954	87186	31598	25942
91402	77227	79516	21007	58602	81418	87838	18443	76162	51146
58299	83880	20125	10794	37780	61705	18276	99041	78135	99661
40684	99948	33880	76413	63839	71371	32392	51812	48248	96419
75978	64298	08074	62055	73864	01926	78374	15741	74452	49954
34556	39861	88267	76068	62445	64361	78685	24246	27027	48239
65990	57048	25067	77571	77974	37634	81564	98608	37224	49848
16381	15069	25416	87875	90374	86203	29677	82543	37554	89179
52458	88880	78352	67913	09245	47773	51272	06976	99571	33365
33007	85607	92008	44897	24964	50559	79549	85658	96865	24186
38712	31512	08588	61490	72294	42862	87334	05866	66269	43158
58722	03678	19186	69602	34625	75958	56869	17907	81867	11535
26188	69497	51351	47799	20477	71786	52560	66827	79419	70886
12893	54048	07255	86149	99090	70958	50775	31768	52903	27645
33186	81346	85095	37282	85536	72661	32180	40229	19209	74939
79893	29448	88392	54211	61708	83452	61227	81690	42265	20310
48449	15102	44126	19438	23382	14985	37538	30120	82443	11152
94205	04259	68983	50561	06902	10269	22216	70210	60736	58772
38648	09278	81313	77400	41126	52614	93613	27263	99381	49500
04292	46028	75666	26954	34979	68381	45154	09314	81009	05114
17026	49737	85875	12139	59391	81830	30185	83095	78752	40899
48070	76848	02531	97737	10151	18169	31709	74842	85522	74092
30159	95450	83778	46115	99178	97718	98440	15076	21199	20492
12148	92231	31361	60650	54695	30035	22765	91386	70399	79270
73838	77067	24863	97576	01139	54219	02959	45696	98103	78867
73547	43759	95632	39555	74391	07579	69491	02647	17050	49869
07277	93217	79421	21769	83572	48019	17327	99638	87035	89300
65128	48334	07493	28098	52087	55519	83718	60904	48721	17522
38716	61380	60212	05099	21210	22052	01780	36813	19528	07727
31921	76458	73720	08657	74922	61335	41690	41967	50691	30508
57238	27464	61487	52329	26150	79991	64398	91273	26824	94827
24219	41090	08531	61578	08236	41140	76335	91189	66312	44000
31309	49387	02330	02476	96074	33256	48554	95401	02642	29119
20750	97024	72619	66628	66509	31206	55293	24249	02266	39010
28537	84395	26654	37851	80590	53446	34385	86893	87713	26842
97929	41220	86431	94485	28778	44997	38802	56594	61363	04206
40568	33222	40486	91122	43294	94541	40988	02929	83190	74247
41483	92935	17061	78252	40498	43164	68646	33023	64333	64083
93040	66476	24990	41099	65135	37641	97613	87282	63693	55299
76869	39300	84978	07504	36835	72748	47644	48542	25076	68626
02982	57991	50765	91930	21375	35604	29963	13738	03155	59914
94479	76500	39170	06629	10031	48724	49822	44021	44335	26474
52291	75822	95966	90947	65031	75913	52654	63377	70664	60082
03684	03600	52831	55381	97013	19993	41295	29118	18710	64851
58939	28366	86765	67465	45421	74228	01095	50987	83833	37216

Prepared By
Dr. Kundan Kumar
Subject Matter Specialist (Agril. Extension)
Krishi Vigyan Kendra: Nagaland University
Lumami-798627
District-Zunheboto (Nagaland)
E-mail: kundanagextension@gmail.com

Published By
Programme Co-ordinator
Krishi Vigyan Kendra: Nagaland University
Lumami-798627
District-Zunheboto (Nagaland)