

Introduction to Sampling Theory

Mini K. G.

Basic of Sampling

A "sample" is a miniature representation of a larger group or aggregate. In other words, the sample provides a specimen picture of a larger whole. This larger whole is termed as the "population" or "universe". The procedure of collecting data for each and every unit belonging to the population, which is the aggregate of all units of a given type under consideration, is termed as census. The effort, money and time required for carrying out census will be enormously large. The census is necessary only if information is required for each and every unit in the population under study. Most of the time, it is impossible, difficult or expensive to observe all the elements of a population in order to arrive at a valid conclusion. Moreover, the sizes of populations are often so large that the study of all the units would not only be expensive but also cumbersome and time consuming, for example, the fishery resources in Indian Seas.

For fisheries research, it is impossible to collect information about the all the fishery resources in the sea. So, a researcher will have to select a representative few, i.e., a sample from the population of the survey. This process is known as sampling. In fisheries research, the sampling surveys are an essential ingredient of fish stock assessment. In sample survey, data is collected from a part or a sample of the population under study in a suitable manner and inference is drawn about the population on the basis of observations made on the selected units. If the samples are selected with an adequate criterion, it is possible to measure the precision of the conclusions or inferences about that population.

The objective of a sample survey is to estimate the unknown population parameters like total, ratio or median based on a random sample drawn by some specified rule from the population. A sample is considered as a subset of population. The principal advantages of sampling compared to census are reduction in cost, greater speed, wider scope and higher accuracy and more importantly the estimate error. Most fisheries research requires some planning and sampling designs before the research is undertaken. The objective of conducting sample surveys in fisheries is to collect data from the stocks and their exploitation level, to study the characteristics of the resources, the effects of exploitation on the abundance of these resources and to determine appropriate fishing levels to obtain

the best possible catches at present and during future years. Fisheries research is most often concerned with the estimation of population mean and totals. e.g. the total catch by different types of gears in a particular landing centre.

Sampling error and Non-sampling error

Sampling error arises from the fact that samples differ from their populations in that they are usually small sub-sets of the total population. Non-sampling errors can be defined as errors arising during the course of survey activities rather than resulting from the sampling procedure. Non-sampling errors can occur because of problems in coverage, response, non-response, data processing, estimation and analysis.

Methods of Sampling

Sampling methods can be broadly classified into two categories:

Probability Sampling

Non-Probability Sampling

Probability Sampling: Probability sampling is based on random selection of units from a population. In other words, the sampling process is not based on the discretion of the researcher but is carried out in such a way that the probability of every unit in the population is pre-defined.

Some of the characteristics of a probability sample are:

- each unit in the sample has some probability of entering the sample,
- weights appropriate to the probabilities are used in the analysis of the sample and
- the process of sampling is automatic in one or more steps of the selection of units in the sample.

Non-probability Sampling: Non-probability sampling is based on the judgement of the researcher. The guiding factors in non-probability sampling include the availability of the units, the personal experience of the researcher and his/her convenience in carrying out a survey. Since these samples are not prepared through random sampling techniques, they are known as non-probability samples. Depending on the technique used, non-probability sampling includes purposive, incidental, snowball and quota sampling.

Sampling Scheme

In survey sampling a fixed finite population is under consideration, where the population elements are labeled so that each element can be identified. Probability sampling helps to select a random sample from a fixed population and assigns a probability of selection to each element. The probability of selection need not be equal for all the elements.

A specific sampling scheme is used in drawing the sample. The collection of techniques for the selection of the sample is termed as sampling scheme.

A sample scheme can be described by two factors.

- **Sampling method.** Sampling method refers to the rules and procedures by which some elements of the population are included in the sample. The basic sampling techniques which are commonly employed are simple random sampling, stratified sampling, cluster sampling, systematic sampling and sampling with unequal probabilities of selection of units particularly with probability proportional to size.
- **Estimator.** The estimation process for calculating sample statistics is called the estimator. Different sampling methods may use different estimators. For example, the formula for computing a mean score with a simple random sample is different from the formula for computing a mean score with a stratified sample. Similarly, the formula for the standard error may vary from one sampling method to the next.

The principal steps in a sample survey are given below.

- i. **Statement of the objectives of the survey:** The objectives of the survey have to be defined clearly prior to the actual survey.
- ii. **Definition of the population to be sampled:** The population will be used to denote the aggregate from which the sample is chosen. Whenever possible, the population to be sampled should obviously coincide with the population about which information is wanted.
- iii. **Determination of the data to be collected:** It is well to verify that all the data are relevant to the purpose of the survey, and that no essential data are omitted.
- iv. **Methods of measurement:** When the kinds of data that are needed have been decided, there may be a choice as to the methods of measurement to be employed.
- v. **Choice of sampling unit:** As a preliminary to the selection of a sample, the population must be subdivided in some way into parts which will be called sampling units, or units. The sampling units must together comprise the whole of the population, and they must be non-overlapping, in the sense that every element in the population belongs to one and only one unit. The construction of a complete list of sampling units, sometimes called a frame, maybe one of the major practical problems. Sometimes the frame is impossible to construct, as with the population of fish in the sea.
- vi. **Selection of the sample:** There are varieties of procedures by which the sample may be selected. The selection also involves a decision about the size of the sample, which in turn requires a provisional estimate of the cost of the survey, to ensure that the sample will fall within the allowable budget.

- vii. Organization of the field work: In extensive surveys, many problems of administration are involved. The personnel must receive training in the purpose of the survey and in the methods of measurement to be employed and must be adequately supervised in their work. A procedure for early checking of the quality of the data may be invaluable. Plans must be made for handling non-response, that is, the failure of the enumerator to obtain information from certain of the units in the sample.
- viii. Summary and analysis of the data: The first step is to edit the completed questionnaires, in the hope of amending recording errors, or at least of deleting data that are obviously erroneous. Decisions about tabulating procedure are needed in the case where answers to certain questions were omitted by some respondents or had to be deleted in the editing process. Thereafter, the tabulations which lead to the estimates are performed. Different methods of estimation may be available for the same data.
- ix. Information gained for future surveys: The more information we have initially about a population, the easier it is to devise a sample which will give accurate estimates. Any completed sample is potentially a guide to improved future sampling, through the data which it supplies about the means, standard deviations, and nature of the variability of the principal measurements, and about the costs involved in getting the data.

References

- Cochran, W. G. 1977. *Sampling Techniques*, third edition. John Wiley & Sons, Inc., New York, 428pp.
- Hansen, M. H., W. N. Hurwitz, and W.G. Madow. 1953. *Sample Survey Methods and Theory*. John Wiley and Sons, Inc., New York.
- Lohr, S. L. 1999. *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA, 494pp.
- Sukhatme, P. V. and B. V. Sukhatme. 1970. *Sampling Theory of Surveys with Applications*. Iowa State University Press, Ames, IA. 452pp.
- Thompson, M. 1997. *Theory of Sample Surveys*. Chapman & Hall, 312pp.