Sampling Techniques for fisheries data collection

Dr V. Geethalakshmi
Principal Scientist
ICAR-CIFT, Cochin
geethasankar@gmail.com

INTRODUCTION

Intuitive application of the principles of sampling in science has been taking place for a long time. However, it was not called sampling but inductive reasoning. Many scientific results are based on observations in just a few experiments. Apparently, it was possible to generalize these experimental results. Although inductive reasoning has been commonly applied both in everyday life and in science for a long time, sampling as a well-defined statistical method is fairly young. Its history started just more than a century ago, in the year 1895.

Anders Kiaer, the founder and first director of Statistics Norway, was the founder and advocate of the survey method that is now widely applied in official statistics and social research. With the first publication of his ideas in 1895 he started the process that ended in the development of modern survey sampling theory and methods.

The classical theory of survey sampling was more or less completed in 1952. Horvitz and Thompson (1952) developed a general theory for constructing unbiased estimates. Whatever the selection probabilities are, as long as they are known and positive, it is always possible to construct a useful estimate. Horvitz and Thompson completed the classical theory, and the random sampling approach was almost unanimously accepted. Most of the classical books about sampling were also published by then (Cochran, 1953; Deming, 1950, Hansen, Hurwitz and Madow 1953, Yates 1949).

The primary objective of a sample survey is to estimate the characteristic(s) under study using a representative sample, which is a subset drawn from population that accurately reflects the members of the entire population. A representative sample should be an unbiased indication of what the population is like. The representative sample is drawn using a sampling method which is a scientific and objective procedure of selecting units from a population and provides a sample that is expected to be representative of the population as a whole.

Even though the sample is representative of the population, and data is reliable, the sample can never reproduce the result a population will give. Therefore, an error gets introduced due to sampling. The discrepancy between the sample estimate and the population value that would be obtained by enumerating all the units in the population in the same manner in which the sample is enumerated are termed as sampling error.

Some situations arise where a probability sampling is not possible. For example, in case of a survey where the respondents are to face unpleasant questions, to ensure sufficient number of responses, volunteers are selected. Also in cases where convenience is the priority units are selected accordingly. Such a sample is called purposive sample.

SAMPLING DESIGN

Let a finite population U consists of N units labelled {1,2,....N}. A sample s* from U is an ordered sequence of n units from U which may be represented as s*={i_1, i_2,.....i_n}. Here i_1, i_2,.....i_n represent the labels of 'n' units drawn from U and 'n' is the sample size.

There may be many such sets of samples of size 'n' which can be drawn from the population. Also, while drawing the units from the population, we can perform the selection with or without replacement. For example while drawing five cards from a pack of 52 playing cards, we can select the first card, again place it in pack, and draw the second card and so on. Here there is a chance that same card gets selected again. This type of sampling is called with replacement sampling.

Sampling in which the units are selected without replacing them back or where the units once got selected does not have a chance of getting selected in the subsequent selections is called sampling without replacement.

Let S* ={s*} i.e. the set of all possible samples from population U. Let p(s*) denote the probability of drawing the sample s* from S* and let p(s*)≥ 0 so that $\sum_{s^* \epsilon S^*} p(s^*) = 1$.
Let π_i denote the probability that ith unit is included in a sample. Then using the addition law of probability, π_i= P(one of the samples containing the ith unit is selected)= $\sum_{i \epsilon s^*} p(s^*)$ where the summation is taken over all the samples containing the ith unit. Assume that $\pi_i > 0, i = 1,2, \dots . N$.

An ordered sampling design is defined as the collection S*={s*} together with the probability measure P*={p(s*)} defined on S* such that p(s*)$\geq$ 0 and $\sum_{s^* \epsilon S^*} p(s^*) = 1$ and is denoted by $D(S^*), P^*)$

PROBABILITY SAMPLING

Any procedure of selecting a sample s* with probability p(s*) for all s*∈ S* is called a probability sampling procedure and a sample selected through such a procedure is called a probability sample. When the probability of selecting a unit from a population is equal for all units in the population then p(s*)=1/total number of possible samples. When all the units in the population have the equal chance of getting selected in a sample we call the procedure as equal probability sampling.

Suppose there are 5 aqua farms in a village and the annual fish production is to be studied on the basis of a sample of size 2.

Let aqua farm units be numbered as {1,2,3,4,5}. Then the possible samples of size 2 without replacement is as follows :
s_1* ={1,2}, s_2^*={1,3}, s_3^*= {1,4}, s_4^*={1,5}, s_5^*={2,3},s_6^*={2,4},
s_7^* {2,5}, s_8^* = {3,4}, s_9^* ={3,5}, s_{10}^* ={4,5}.
Also we have p(s_i^*)=1/10, i=1,2,...5 which is equal for all the samples listed.

Equal probability sampling procedure is called simple random sampling. Simple random sampling or srs in short form, can be performed with or without replacement.

A method of sampling such that every one of the ${}^N C_n$ possible samples of size n from N has the same probability namely $\frac{1}{{}^N C_n}$ of being selected is called simple random sampling without replacement. In the above mentioned example, probability of selecting a sample s_i*, from the above set of samples is same, in case of simple random sampling without replacement i.e. p(s_i*)=1/10, i=1,2,…10

Let Y denote the characteristic under study. In the above example it is the fish production. Denote Y_i the value of the characteristic associated with unit U_i, i=1,2,….N

Further let $\bar{Y}$ =$\frac{1}{N}\sum_{i=1}^{N} Y_i$ be the mean per unit of the population. This term is generally referred to as the Population mean. Using a sample we have to estimate this term and the estimator is commonly known as the estimator of the population mean and denoted as $\bar{y}$.

Similarly $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ is called the population variance which always associated with the population mean which also has to be estimated along with the population mean. It gives a measure of precision of our estimate. The estimator of the population variance is called as the sampling variance is generally represented as s^2.

Let y_i, i=1,2,3,….n be the values of the characteristic under study Y from the sample of size n selected from the population. Then the sample mean and variance is given by

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ are unbiased estimators of the population mean and variance respectively.

PROBABILITY PROPORTIONAL TO SIZE SAMPLING

In the previous section, we discussed the selection of sample from a population by assigning equal probability to the units to be included in the sample. In certain practical situations, some units have to given more weightage because of their contribution to the characteristic under study. For example, we want to estimate the total fish production based on a day's landing. There will be fishing boats which have gone for single day fishing, some of them for 2-5 days and a few boats might have landed fish after fishing for a week. The quantity of catch may also vary based on the number of fishing days. Therefore, more weightage should be given for the fishing vessels whose fishing duration is more compared to boats which go for single day fishing. Likewise, suppose the aquafarms discussed in the previous example are of varying sizes and because of this variation their fish production also varies. Probability proportional to size sampling or pps sampling as it is called is a sampling procedure where the sampling units are assigned probabilities for selection based on size criteria.

Suppose there are 5 aqua farms in a village and the annual fish production is to be studied on the basis of a sample of size 2. Let aqua farm units be numbered as {1,2,3,4,5}.

Let us assume that depending on their size the following probabilities can be assigned to the individual units of the 5 aquafarms : p_1=0.2, p_2=0.1,p_3=0.2, p_4=0.4, p_5=0.1. Note that $\sum p_i$=1.

Here, if the scheme is without replacement for the following set of possible samples,

s_1* ={1,2}, s_2^*={1,3}, s_3^*= {1,4}, s_4^*={1,5}, s_5^*={2,3},s_6^*={2,4},s_7^* {2,5}, s_8^* = {3,4}, s_9^* ={3,5}, s_{10}^* ={4,5}, the probability is calculated as follows :

p(s_1^*) $= p_1p_2$=0.02; p(s_2^*) $= p_1p_3$=0.04; p(s_3^*) $= p_1p_4$=0.08; p(s_4^*) $= p_1p_5$=0.05;
p(s_5^*) $= p_2p_3$=0.02; p(s_6^*) $= p_2p_4$=0.04;p(s_7^*) $= p_2p_5$=0.01 p(s_8^*) $= p_3p_4$=0.08;
p(s_9^*) $= p_3p_5$=0.02; p(s_{10}^*) $= p_4p_5$=0.04.

Given a sampling procedure D(S*,P*), a straightforward procedure for selecting a probability sample is given below :

(i) Identify all possible samples s*, say, M in number and denote the serial number from 1 to M. So here we have s_1^* ={1,2}, s_2^*={1,3}, s_3^*= {1,4}, s_4^*={1,5}, s_5^*={2,3},s_6^*={2,4},s_7^* {2,5}, s_8^* = {3,4}, s_9^* ={3,5}, s_{10}^* ={4,5}, if the scheme is without replacement and M=10.

(ii) Form successive cumulative totals

$T_i = \sum_{j=1}^{i} p(s_j^*), i = 1,2,\ldots,M$

Choose a random number R such that $0 \leq R \leq 1$ and select the sample s_i^* with serial number i if $T_{i-1} \leq R \leq T_i$

Now T_1=0.02, T_2=0.06; T_3=0.14; T_4=0.19; T_5=0.21; T_6=0.30; T_7 =0.31; T_8=0.39; T_9=0.41; T_{10}=0.45.

Suppose R=0.43. Then sample number 10 will be selected which is {4,5}.

When the number of all possible samples are manageable and can be written down easily as in the above example of aquafarms, then it is possible to select a sample from the population using probability proportional to size sampling using the above procedure. In general, the procedure for selecting a sample with varying probability is given below:

Let X_i denote an integer which is proportional to size of the ith unit i=1,2,....,N. Form the successive cumulative totals X_1, X_1+X_2,.... $\sum_i^n X_i$ and draw a random number R not exceeding $\sum_i^N X_i$ using either the table of random numbers or the random number generator function in Excel. If $\sum_i X_i \leq R \leq \sum_i X_i$, the ith unit is selected. The procedure is repeated till 'n' units get selected.

Let Y_i, i=1,2,...,N denote the value of the characteristic under study Y for the ith unit of the population. Let P_i be the probability of selecting the ith unit in the population. Obviously, $\sum_{i=1}^{N} P_i = 1$. We shall now consider the problem of estimating the population mean $\bar{Y}$ based on the sample of n units with replacement. If the sample of size n is selected using a probability proportional to size sampling method, then denote

$Z_i = \frac{Y_i}{NP_i}$, i=1,2,...N. An estimator of Population mean is $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$.

The variance of $\bar{z}$ is given by $\hat{V}(\bar{z}) = \frac{s_z^2}{n}$ where $s_z^2 = \frac{1}{n-1} \sum_i^n (z_i - \bar{z})^2$. It is proven that both $\bar{z}$ and $\frac{s_z^2}{n}$ are unbiased estimators of the population mean and variance.

STRATIFIED SAMPLING

This type of sampling mechanism is frequently used in sample surveys where we need estimate the population parameter for a population which can be divided as groups or strata. For example a market researcher has to conduct a consumer preference study for a convenience product from fish which is planned to capture the super markets. Then his population will consist of households from an urban area and from varying levels of income groups. In order to have an reasonable representation from all sections of the population, the households should be divided into low, middle, high income groups or strata. Then a suitable sample from each group can be drawn using either simple random sampling or any procedure and the parameter(consumer preference) studied.

Another example is in agriculture where total yield of a crop is to be estimated from a state. Stratification of the farms will be done district wise and the total crop production from each district can be estimated. The groups into which the population is divided is called strata and whole procedure of drawing samples from each stratum is known as stratified random sampling. When simple random sampling is used to select samples from each stratum, then the procedure is called a stratified random sample.

We shall assume that the population of size N is divided into L strata and that sampling within each stratum is simple random sampling without replacement. Further for the hth stratum h=1,2,...L the following notations apply :

N_h the number of units

n_h the sample size

$f_h = \frac{n_h}{N_h}$, sampling fraction

Y_{hi} is value of the characteristic under study Y for the ith unit, i=1,2,....,N_h

$W_h = \frac{N_h}{N}$

Let Y_h denote the total of Y-values for units belonging to stratum h. Then the mean of stratum h is given by $\boldsymbol{\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi}}$.

Let y_{hi} denote the value of the characteristic under study Y pertaining to the ith unit in the sample from hth stratum. The the mean of sample from hth stratum is given by $\bar{y}_h = \frac{1}{n_h}\boldsymbol{\sum_{i=1}^{n_h} y_{hi}}$.

$\boldsymbol{S_h^2 = \frac{1}{(N_h-1)}\sum_{j=1}^{N_h}(Y_{hj} - \bar{Y}_h)^2}$, the mean square based on N_h units

$\boldsymbol{s_h^2 = \frac{1}{(n_h-1)}\sum_{j=1}^{n_h}(y_{hj} - \bar{y}_h)^2}$, the sample mean square based on n_h units

Unbiased estimator of the population mean $\bar{Y}$ is given as

$$\boldsymbol{\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h}$$

Here it is assumed that the sampling is carried out independently in each stratum. The variance of the stratified sampling estimator $\boldsymbol{\bar{y}_{st}}$ is given by

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{(N_h - n_h)}{N_h}\frac{S_h^2}{n_h}$$

An unbiased estimator of the variance of $\bar{y}_{st}$ is given by

$\boldsymbol{\widehat{V}(\bar{y}_{st})} = \sum_{h=1}^{L} W_h^2 \frac{(N_h-n_h)}{N_h}\frac{s_h^2}{n_h}$ .The expression for variance of the stratified sampling estimator shows that the precision of the estimator is based on the n_h i.e. the stratum sample sizes. Once we decide to conduct any survey for estimating characteristic under study pertaining to a population we will be given a cost within which the survey should be conducted. Therefore we have the liberty only to decide the sample size within cost limit. But the precision will be more if the variance is less or in other words, when the sample size is more. Practically, when we desire that the sample size should be increased, cost of coverage will also increase. Since the sample size *n* is fixed in advance, the problem at hand usually is the allocation of sample sizes n_h within each stratum.

Optimum allocation : The guiding principle is that decide n_h in such a manner to estimate population mean $\bar{Y}$ with desired precision for a minimum cost or with a maximum precision for a given cost. The allocation of the sample in accordance with this principle is called the optimum allocation.

Suppose c_h denotes the cost of the survey in stratum h. Then total cost of survey can be represented as

$$C = \sum_{h=1}^{L} c_h n_h$$

Then the variance of the estimator $\bar{y}_{st}$is minimum for given cost C_0 or the cost of the survey is minimum for a given variance V_0 when n_h is proportional to $\frac{W_h S_h}{\sqrt{c_h}}$. Therefore in optimum allocation the sample sizes are allotted to stratum according to the following formula $n_h = \frac{W_h S_h}{\sqrt{\mu_0 c_h}}$.

Neyman allocation : When c_h is sample for all the strata, then the sample sizes are allotted to stratum according to the following formula $n_h = \frac{W_h S_h}{\sqrt{\mu_N}}$ and this type of allocation is called the Neyman allocation. $1/\sqrt{\mu_N}$ is called the constant of proportionality.

Proportional allocation : When n_h is proportional to W_h then the sample size can be allocated according to the formula $n_h = \frac{W_h}{\sqrt{\mu_P}}$ where $1/\sqrt{\mu_P}$ is called the constant of proportionality.

CLUSTER SAMPLING

Before applying any sampling procedure, the population is divided into finite number of distinct identifiable units called the sampling units or elements. Groups of elements can be called clusters. In some practical situations it is more convenient to sample clusters from a population than selecting the individual sampling units. In crop estimation surveys, when the total yield of a crop is to be determined, the sampling frame or list of farms may not be readily available from all the villages. But the list of villages will be available. Here, cluster sampling can be employed by considering the villages as cluster of farms.

When the sampling unit is a cluster then the sampling is called cluster sampling. In cluster sampling all the elements in the selected cluster will be enumerated. A necessary condition for employing a cluster sampling procedure is that every element or smallest unit in the population will correspond to one and only one unit of the cluster so that the total number of sampling units in the frame will cover all the units of population under study with no omission or duplication.

When the entire area containing the population under study is subdivided into area segments, and each element of a population is associated with one and only one area segment then the procedure is called area sampling. It is not necessary that all the elements associated with an area segment be located physically within its boundaries. For example in case of aquaculture, different ponds belonging to the same farm household or farmer will not necessary be in the same location or adjacent. Such a segment is called a open segment.

Let the population consist of N clusters of M elements each. Using simple random sampling without replacement n clusters are selected from the N clusters. We use the following notations :

Y_{ij} denotes the value of the characteristic under study for the jth element of the ith cluster; j=1,2,….M; i=1,2,…..N.

$\bar{Y}_{i.} = \frac{1}{M}\sum_{j=1}^{M} Y_{ij}$, denote the mean per element of ith cluster

$\bar{\bar{Y}} = \frac{1}{N}\sum_{i=1}^{N} \bar{Y}_{i.}$, the mean of cluster means

$\bar{Y} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} Y_{ij}$, the population mean.

Then an unbiased estimator of the population mean is given by

$\bar{\bar{y}} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_{i.}$, which is actually mean of cluster means based on the sample observations from the selected n clusters.

The mean square between elements in the ith cluster is $S_i^2 = \frac{1}{M-1}\sum_{j=1}^{M}(Y_{ij} - \bar{Y}_{i.})^2$.

The mean square between cluster means is given as $S_b^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{Y}_{i.} - \bar{Y})^2$.

The variance of the estimator $\bar{\bar{y}}$ is V($\bar{\bar{y}}$)=$\left(\frac{1}{n} - \frac{1}{N}\right) S_b^2$.

An unbiased estimator of V($\bar{\bar{y}}$) is given by

$\hat{V}(\bar{\bar{y}}) = \left(\frac{1}{n} - \frac{1}{N}\right) s_b^2$ where is the sample mean square between the cluster means.

For example, in order to estimate the fish production from aqua farms of a particular district, clusters of aquafarms can be formed and a sample of few clusters selected and completely enumerated.

SYSTEMATIC SAMPLING

A method of sampling in which only the first unit is selected at random and the rest being selected automatically according to a pre-determined pattern is called systematic sampling. Examples where this kind of sampling is often employed is forest survey. To estimate the number of trees or timber in a forest where the units are innumerable systematic sampling is used. Another example is application in mangrove forestation where the parameter of interest to find out the density.

Assume that the population consists of N units serially numbered from 1,2,…N.. Assume further that N is expressible as a product of two integers k and n, so that N=kn. Draw a random number less than or equal to k, say i, and select the unit with the corresponding serial number and every k-th unit in the population thereafter. The sample will contains the units with serial numbers, i, i+k, i+2k,….i+(n-1)k.

Selection of every kth time interval to observe fishing crafts for estimation of fish production is an example where systematic sampling can be used. The advantages of systematic sampling is it involves low cost and is simple to follow.

Systematic sampling resembles stratified sample in the sense that one unit is selected from each stratum containing k consecutive units. However this resemblance is only casual. In stratified sampling the unit to be drawn from each stratum is randomly selected and in systematic sampling the position of the unit is predetermined relative to the first units selected. Unless the units in each stratum are arranged at random, systematic sampling can never be equivalent to stratified random sampling.

Systematic sampling strictly resembles cluster sampling. A systematic sample is equivalent to one cluster of elemets selected from k clusters of n units each, . Since the first number less than or equal to k is chosen at random, each one of the k clusters get an equal chance of getting drawn as a sample.

Let Y_{ij} denote the value of the characteristic under study for the jth unit of the ith cluster bearing the serial number i+(j-1)k, i=1,2…,k, j=1,2,….,n. Further let

$\bar{Y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$, $\bar{Y}_{.j} = \frac{1}{k}\sum_{i=1}^{k} Y_{ij}$, $\bar{Y} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij} = \frac{1}{n}\sum_{j=1}^{n} \bar{Y}_{.j} = \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{i.}$

The sample mean $\bar{y}_{sys} = \bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$ is an unbiased estimator of $\bar{Y}$ with variance given by $\mathrm{V}(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{Y}_{i.} - \bar{Y})^2$

SUB-SAMPLING OR TWO-STAGE SAMPLING

In the cluster sampling, all the units of the selected clusters are measured completely. If the units within the same cluster give more or less the same value, then it is less costlier to observe a sample of units from it. A common practice is to select first the clusters which are called the first stage or primary units. Units which are chosen from the cluster are called second stage units. This is known as two-stage sampling or sub-sampling. An application of two-stage sampling in fisheries is for estimation of marine fish landings from the country. Here selected landing centres are the first stage units and the second stage units are the selected boats landing at these centres for recording the data on fish catch. When the number of stages is more than two from which a sample is selected, then it is called multi-stage sampling.

Consider two-stage sampling when the first-stage units are of equal size and simple random sampling without replacement is employed at each stage. Let the population consist of N first stage units with M second stage units in each of the first stage unit. Let $\bar{Y}_{i.} = \frac{1}{M}\sum_{i=1}^{M} Y_{ij}$ be the mean of observations in the ith first stage unit.

The population mean is given by $\bar{Y} = \bar{Y}_{..} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} Y_{ij}$, and the estimate of the population mean is given by $\hat{y}_2 = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$ where $\bar{y}_i$ is the mean of the 'm' secondary units selected from the ith first stage unit.

The estimate of the variance of the sample mean $\hat{y}_2$ is given by

Var$(\hat{y}_2) = \left(\frac{1}{n} - \frac{1}{N}\right).s_b^2 + \frac{1}{N}\left(\frac{1}{m} - \frac{1}{M}\right)\bar{s}_w^2$, where $s_b^2 = \frac{\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_2)}{n-1}$, $\bar{s}_w^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2$, where $s_i^2 = \frac{\sum_{i=1}^{m}(y_{ij} - \bar{y}_i)}{m-1}$

Sukhatme et. al. (1997) gives the estimation procedure for estimating population mean when the first stage units are unequal.

ESTIMATION OF MARINE FISH LANDINGS

India has a coastline of about 8129 km and there are about 3000 marine fishing villages and about 1400 landing centres along the coastline. Fishing boats arrive at numerous locations all along the coastline during day and at times during night also for landing the fish catch. Central Marine Fisheries Research Institute, Cochin has standardized the methodology for estimation fish landings from marine sources for the entire nation.

The sampling design adopted by CMFRI to estimate resource-wise/region-wise landings is based on stratified multi-stage random sampling technique. In this, the stratification is over space and time. Over space, each maritime state is divided into suitable, non-overlapping zones on the basis of fishing intensity and geographical considerations. These zones have been further stratified into substrata, on the basis of intensity of fishing. Major fisheries harbours/ centres are classified as single centre zones for which there is an exclusive and extensive coverage. The stratification over time is a calendar month.

One zone and a calendar month is a space-time stratum and primary stage sampling units are landing centre days. Nine landing centres are selected at random from each zone for recording fish landings. For observation purpose, a month is divided into 3 groups, each of 10 days. From the first five days of a month, a day is selected at random, and the next 5 consecutive days are automatically selected. From this three clusters of two consecutive days are formed. From the 2nd and 3rd group of 10 days, 3 clusters of two days are formed systematically with a sampling interval of 10 days. Therefore, there will be a total of 18 days in a month with nine clusters of 2 days each.

The observation is made from 1200 hrs to 1800 hrs on the first day and from 0600 hrs to 1200 hrs on the second day, in a centre. The `night landing' obtained by enquiry on the second day covering the period of 1800 hrs of the first day to 0600 hrs of the next day are added to the day landings so as to arrive at the landings for one (landing centre day) day (24 hours).

In all in a month, 9 landing centre days are observed which forms the first stage units. The second stage units are the fishing boats landing on the day of observation.

When the total number of boats landed is 15 or less, on the day of observation, the landings from all the boats are enumerated for catch and other particulars. When the total number of boats exceeds 15, the following procedure is followed to sample the number of boats (Alagaraja, 1984)

Table 6.9. Enumeration of boats

Number of units landed	Fraction to be examined
Less than or equal to 15	100 %
Between 16 and 19	First 10 and the balance 50 %
Between 20 and 29	1 in 2
Between 30 and 39	1 in 3
Between 40 and 49	1 in 4
Between 50 and 59	1 in 5 etc.

Based on the data collected from the fishing units on catch, the estimate of total landings in a day, in a month and in a year can be obtained for landing centre, zone, district and state even for each species. For further study, Srinath, et. al. (2005) can be referred.

ESTIMATION OF INLAND FISH PRODUCTION IN INDIA AND PRACTICAL ISSUES

Inland fisheries enjoys prime of place in Indian economy. It provides employment and livelihood for fishers who solely depend on it. In inland fishery sector, the data collection on various important parameters such as the catch, size of fleet, level of employment, per capita yield etc. is an enormous task owing to the sporadic spatial and temporal distribution of the resources. Attempts are being made by Central Inland Fisheries Research Institute to collect data using communication devices like mobile from the fishermen operating in remote centres. Unlike marine sector, inland fisheries cannot claim a satisfactory status with regard to data collection.

India has vast potential inland resource scattered through out the country. However, their concepts and definitions vary from one region to another region. So the data collected from these resources are sometimes neither comparable nor compilable at Central Level. There is a strong need for uniform concepts, definition, collection and compilation of methodology for this sector.

The Central Inland Fisheries Research Institute, Barrackpore, made an attempt to estimate the area and catch from ponds in the district of Hoogly, West Bengal during 1962-63 but it did not lead to accomplishment of the task at hand. In 1973-75, the NSSO conducted a survey covering three districts, one each in West Bengal, Tamil Nadu and Andhra Pradeshwith the aim of obtaining estimate of catch both from impounded water and riverine resources by enquiry. The estimates worked out were not satisfactory, particularly from riverine resources.

In another pilot survey conducted by IASRI, New Delhi and CIFRI, Barrackpore in one district of West Bengal during 1978-81, the data were collected both by enquiry and by physical observation. The main objectives of the survey were (1) to evolve suitable sampling methodology for estimation of (a) inland water resources, (b) total catch for inland fisheries and (2) to study the prevailing practices of pisciculture. The study covered only ponds in the district of 24-Parganas in West Bengal. The catch estimate of other important resources like estuaries, rivers, brackish water impoundments, beels could not be attempted due to limited manpower. In spite of all these attempts, there is no scientifically designed and accepted method for collection and estimation of all types of inland fishery resources.

However, the Department of Animal Husbandry and Dairying, Ministry of Agriculture, Govt. of India, Plan, entrusted the development of uniform concepts, definitions and terminologies for various inland fishery resources and a suitable and standardized methodology for collection and estimation of inland fishery resources and catch to Central Inland Fisheries Research Institute, Barrackpore in collaboration with the states. The methodologies have been developed and tested in various states during 8th and 9th Plans. The states have been provided training and guidance for estimation of catch from various inland resources during 10th Plan and since then the estimation of inland fish catch is continuing.

SAMPLING AND NON-SAMPLING ERRORS

The errors involved in the collection, processing and analysis of data can be broadly classified as (i) Sampling and Non-sampling errors.

(i) Sampling errors : Sampling errors have their origin in sampling and arise due to the fact that only a part of the population (i.e. sample) has been used to estimate the population parameters and draw inferences about the population, As such the sampling errors are absent in a complete enumeration survey. The reasons of such errors may be due to faulty selection of sample, substitution of observation for the sampling unit which could not be covered during the survey, faulty demarcation of the sampling unit and constant error due to improper choice of the statistics for estimating the population parameters.

(ii) Non-sampling errors : These errors mainly arise at the stages of observation, ascertainment and processing of the data and are thus present in both complete enumeration survey and the sample survey. Thus the data obtained from the complete census though free from the sampling errors, would still be subject to non-sampling errors whereas data obtained in a sample survey would be subject to both sampling and non-sampling errors. Non-sampling errors may occur due to

(a)Faulty planning or definitions: After stating the objectives of the survey, definitions about the characteristics for which data to be collected should be specified . Here the non-sampling errors may occur due to data specification being inadequate and inconsistent with the objectives of the survey. At times error may be due to the location of the units and actual measurement of the characteristic, errors in recording or may be due to a ill-designed questionnaire.

(b) Response error : When the respondent misunderstood a particular question and furnish improper information. At times, the respondent deliberately gives wrong information when the questions are sensitive. Questions based on 'recall' memory of the respondent will sometimes lead to improper or incomplete information.

(c) Non-response bias :Non-response bias occurs when the full information is not got from all the sampling units. In the event of respondent not at home or even after repeated calls the respondent is not able to furnish the information fully such a bias occurs.

(d) Errors in coverage : If the objectives of the survey is not precisely stated then some units which are not to be covered will be enumerated under the survey and certain units will be excluded from the survey which are relevant and are to be covered under the survey.

(e) Compiling errors : Various operations such as data processing such as editing and coding of the responses, tabulation and summarizing the orginal observations made in the survey are a potential source of error. Compilation errors are subject to control though verification, consistency check, etc.

(f) Publication errors : The errors committed during presentation and printing of tabulated results are basically due to two sources. The first refers to the mechanics of publication – the proofing error and the like. The other, which is of more serious nature lies in the failure of the survey organization to point out the limitations of the statistics.