THE DESIGN OF
FISHERIES STATISTICAL SURVEYS
- Inland Waters -

by
G.P. Bazigos
FAO-Fishery Statistical Consultant
Fishery Statistics Unit
Policy and Planning Service

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, May 1974

PREPARATION OF THIS DOCUMENT

This report has been prepared in the Policy and Planning Service, Department of Fisheries, FAO, for training purposes in the field of fishery statistics (inland waters), in particular as a guide for the design of large-scale fishery statistical surveys.

Distribution

FAO Department of Fisheries
FAO Statistics Division
FAO Regional Fishery Officers
FAO African Inland Waters Projects
Other FAO Fishery Projects
Specialised Institutes and
  Individual Scientists

Bibliographic entry

Bazigos, G.P. (1974)
FAO Fish.Tech.Pap., (133):122 p.
The design of fisheries statistical
surveys - inland waters

Fisheries statistics surveys - inland waters. Training courses covering: Concepts, methods, techniques, applications. Fishermen, fishing gear, fishing craft. Frame surveys. Catch assessment surveys. Fishing industry - economics, labour. Sampling and sampling designs. Data processing, field operations control. Working sheets and instructions, examples. Selected bibliography - technical reports.

CONTENTS

CONTENTS

CONTENTS

CONTENTS

CONTENTS

PREFACE

With science becoming more and more important in African inland waters, there is a greatly increased need for the Fishery Statisticians to be sure that they are using scientific methods for obtaining the statistical data so indispensable for sound management of a fishery and for planning, marketing and other aspects of fishery development.

It is beyond the resources of the African countries assisted by FAO Projects to collect facts year by year from each fishing economic unit at the inland water places in the country. Fortunately, a carefully designed sample survey can provide the necessary information for guidelines that a country needs, at a cost the country may well afford. This manual serves this purpose. The manual deals with the application of sampling techniques at elementary (Part I), intermediate (Part II) and advance level (Part III). Specifically, Part I and Part II are a summary of the lectures given in the training courses in African countries on the survey system of large-scale fisheries statistical sample surveys1/. In Part III, the application of sampling theory in fisheries statistical surveys is presented.

1/ Part I and Part II of the manual have been issued as an independent report: Training Courses on Fishery Statistical Surveys (Inland Waters), by G.P. Bazigos, UNDP/SF/ZAM.11 March 1973

ACKNOWLEDGEMENTS

The author of a manual must be grateful to almost everyone who has touched the field and I take pleasure in acknowledging my debt to all persons whose results I have drawn upon during the writing of this manual which is the product of five years experience at large African lakes where FAO Projects are being undertaken.

My first acknowledgement must go to Mr. C.H. Clay, FAO Co-ordinator of African Lakes Projects, for the unstinting assistance given in connection with my assignments. Appreciation must also go to the Managers of FAO Projects, to my African counterparts in the field and other field staff who made the implementation of the sample surveys, I hope, successful. My thanks also go to my colleague, Mr. L. Butler, FAO, Rome, with whom I have had the good fortune to work and critically discuss various technical problems.

I would also like to record my thanks to Mr. S.R. Coppola for his assistance in the statistical calculations required for the preparation of the manual, and last but not least my thanks are due to Miss Sheila Campbell for the patience with which she typed the drafts and master copy.

PART I: ELEMENTARY COURSE

1. INTRODUCTION

1.1 The purpose of statistical surveys

In its broadest sense the purpose of a "statistical survey" is the collection of information (data) to satisfy a definite need. The need to collect data arises in every conceivable sphere of human activity. A few examples are given below from selected fields.

1. Human population: Most governments nowadays collect information regularly about:

   a) Total number of persons
   b) Where these persons live (e.g. towns, rural areas, etc)
   c) Sex and age of these persons
   d) Their education level etc.

   These data are needed by the respective countries in order to determine their future needs in terms of food, clothing, schools, recreation facilities, etc.

2. Labour: Since labour is a key source of production, data are collected on the number of persons engaged in economic activity, the number of hours work, etc. It is also noted that the wages and salaries paid to labour determine the level of living and the demand for goods and services.

3. Industry: Collection of data in the industrial sector is no less important. The number of industrial undertakings and their kind, the number of persons engaged in them, the amount of raw materials consumed, the extent of production of goods are some of the data needed.

4. Agriculture: With rising population, it is becoming more and more important to assess the agricultural resources of a country. Some of the data needed for any planned programme of national development are: land under agriculture, areas under different crops, areas under pasture and forest, production of food-grains, fruits, etc., and the number and quality of livestock.

5. Fishing: "... the fishery is the channel between the fish in the inland waters (sea) and the inland market place, and it reacts sensitively to stimuli from both sides; to changes in conditions of supply in the inland waters (sea), and to changes in conditions of demand in the shops ..."

For sound management of a fishery and for planning, marketing and other aspects of fishery development, data of fish catch and of fishing effort involved to obtain the catch are required.

It should be noted that as far as management of fisheries is concerned, there are usually two problems:

1. There are some fisheries which are being intensely exploited and therefore need some management or regulation so that maximum yield from these fisheries could be obtained on a continuing basis without depletion of these resources taking place.

2. There may be fish stocks, especially beyond the range of current fishing operations, which are either under-exploited or unexploited at present and fisheries may be developed based on them.

2. BASIC STATISTICAL CONCEPTS

2.1 Population of units

An aggregate of well defined objects is called "population of units". Examples:

1. Total number of persons in a classroom (unit of the population = person in the classroom).
2. Total number of fishermen at Lake Victoria (unit of the population = fisherman at Lake Victoria).
3. Total number of firms in E.A.C. (unit of the population = firm in E.A.C.).

In statistical surveys the definition of a "population of units" under study involves:

1. The definition of the unit of the population.
2. The geographical limitation of the population.
3. The fixing of limits other than merely geographical e.g. in a socio-economic survey at Nairobi area we may decide that people living in institutions like prisons, mental homes and hospitals should be excluded from the survey.

2.2 Population of characteristics

Every unit in a "population of units" carries with it a number of characteristics Examples:

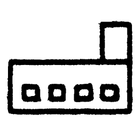
1. In the case of a human population every unit in the population carries with it a great number of characteristics, i.e.

Population of characteristics

- → Sex
- → Age
- → Weight
- → Income
- → Religion

etc. etc.

2. In the case of a population of industrial firms each unit in the population carries with it a good number of characteristics, i.e.



- → Number of employees
- → Machinery used
- → Raw material used
- → Production of goods

etc. etc.

3. In the case of a population of fish each unit in the population carries with it a good number of characteristics, i.e.

- Sex
- Weight
- Length
- Age

etc. etc.

It should be noted that in a given "population of characteristics" some of the characteristics are "quantitative" and others "qualitative".

Quantitative characteristics are measureable (e.g. age is a measurable characteristic - How old are you? 35 years, etc.).

Qualitative characteristics are not measurable (e.g. language is a non-measurable characteristic - What language do you speak? English, etc.).

3. PLANNING A FRAME SURVEY

3.1 Why a frame survey?

A frame survey is a sort of inventory survey. By this survey items of informatio (data), are collected on a number of basic characteristics needed to assess the size and structure of a fishing industry. Usually the following items of information are covered by the survey:

1. Size and area distribution of fishing sites.
2. Number of fishing boats (by type).
3. Size and composition of fishing labour force (migratory pattern of the fishermen).
4. Fishing gear owned (by type).
5. Fishermen's capital goods supply centres.

Also at the same time some information can be collected on fish processing and marketing.

The results of a frame survey can be used, among other things, to set up the "sampling frame" of the population we are dealing with. The established sampling frame is mainly used for the selection of the samples of various surveys covering the same population.

3.2 Planning a frame survey (FS)

The steps that should be taken at the process of planning a frame survey in inland waters can be broadly classified as follows:

3.2.1 The length of the shoreline of a lake

The length of the shoreline of a lake under study can be considered as one of the characteristics which must be measured at the design process of the survey. This information can be obtained by making use of the existing "topographical maps". A topographical map is a map which shows the details of the countryside.

Topographical maps may be divided into two groups:

1. Small scale maps
2. Large scale maps

Small scale maps are those that have a scale of miles to the inch, e.g. 1 inch = 2 miles or 1 inch = 8 miles. Large scale maps on the other hand, are those which have a scale of inches to a mile, e.g. 6 inch = 1 mile. Note: to measure the direct distance between two points is relatively simple. Mark off the distance on the map with a ruler and then measure this distance against the linear scale at the base of the map.

It should be noted that in certain lakes (man-made lakes) the length of the shoreline of the lake is, among other things, a function of the lake level. As the water level is not constant through time neither is the length of the shoreline - thus, the statement about the length of the shoreline should indicate at what water level the measurement was made to reveal more fully their meaning.

3.2.2 Area stratification

It is common practice at the design process of a FS to divide the area under consideration into a number of smaller areas, here called "strata". Field operations of the survey are taking place within the established strata. Past experience has proved that this method simplifies the work of the field personnel and at the same time improves the quality of the obtained results.

For stratification purposes various criteria can be used e.g.,

1. One can divide a lake arbitrarily into three parts:

   Stratum 1: southern part of the lake
   Stratum 2: middle part
   Stratum 3: northern part

2. A lake can be divided into a number of portions (strata) which should be of equal size as far as the length of shoreline is concerned.

3. The cases in which supplementary information is available from other related studies (limnological, biological), these data can be used for a proper stratification of a lake.

3.2.3 The fishing sites (Primary Sampling Units)

At inland waters where fishing is taking place the population is concentrated in "fishing sites", here called Primary Sampling Units (PSU's). Specifically, a fishing site can be considered to consist of two "statistical units":

1. "Residential area" where the fishermen are living.

2. "Landing place" or "home beach" where the fishermen keep their fishing craft (local canoes).

Further, by taking into account the "mobility" of the fishermen, fishing sites can be classified as follows:

1. Permanent fishing sites.

2. Fishing camps.

From a statistical point of view a fishing site can be considered as "permanent" if fishermen live there for a long period of time (e.g. more than one year) and have no intention to move to another place, at least in the near future.

A fishing camp is a place where fishermen live for a relatively short period of time and intend to leave the place in the near future, either to go back to their home fishing site or to another place.

Information about the state of permanency of a fishing site can easily be collected by using special questions in the questionnaire of a FS. Usually the required information is obtained from the chief of the fishermen:

1. Since when have you been in this fishing site?

2. Are you planning to move from this fishing site?

   Yes [ ] 1 No [ ] 2

   2.1 If YES,

   a) when..................................

   b) where are you going..................

Besides the problem of classifying a fishing site as a permanent fishing site or a fishing camp, there is the decision whether or not to divide a multi-grouped settlement into independent places or treat it as a unit.

To deal with the problem the following criteria can be used:

1. Physical distance between the places.
2. Recognition of one or more chiefs.
3. Difference of name or not.

Depending on the nature of these factors a decision should be made to treat the multi-grouped settlement as one unit or as a number of independent units.

A fishing site of any type may have one or more landing places. Usually small fishing sites have one small beach serving the entire unit, whereas large fishing sites might have several distinct beaches with different functions, e.g. landing of local canoes, loading canoes going to market or transport launches.

3.2.4 <u>The Fishing Economic Unit (FEU)</u>

The total fish production of a fishing industry (from the view of an economic study) is the result of the operations of the fishing economic units (FEU's) within a given period of time. Specifically, in our case FEU's fall into two categories:

1. Usual Fishing Unit (UFU), which consists of fishing craft, fishing gear and fisherman(men) to carry out fishing operations.
2. Minor Fishing Unit (MFU), which is an integral unit composed of fishing gear and fisherman (without fishing craft) to carry out fishing operations.

It should be noted that by taking as criterion of stratification the ownership of the unit, the FEU's can be divided as follows:

1. Private ownership units.
2. Agreed partnership units.
3. Cooperative units.

3.2.5 <u>The components of a FEU</u>

3.2.5.1 <u>Fishermen</u>

A *fisherman is a person who engages in actual operation of capture or culture of aquatic resources.* Therefore, family members or others who assist the work relating to the fishing operation such as unloading fish, net repairing, processing, etc., who do not participate in fishing operations are not considered as fishermen.

Fishermen may be classified concerning "employment status" as:

1. Fishermen boat owners.
2. Fishermen with gears only.
3. Assistants

Another classification of fishermen is according to the "time spent for fishing" within a year:

1. Full-time fishermen.
2. Part-time fishermen.
3. Occasional fishermen.

3.2.5.2 Fishing gear

A principal requirement of observations of fishing is that there should be correct identification and description of the gears owned by a FEU, and the methods they are using.

The gear and methods of fishing in use in African countries (inland fishery) can be divided into two main groups:

1. Small-scale fishing methods, e.g. traps, baskets, hooks, etc.
2. Commercial fishing methods, e.g. gillnets, seine nets, hand nets, etc.

It should be noted that the users of the results of fisheries statistical surveys should have in mind the existing differences among the following four distinct magnitudes of fishing gears:

1. Number of fishing gears owned by a FEU.
2. Number of fishing gears set (in a fishing operation).
3. Number of fishing gears of the number set, which were inspected.
4. Number of fishing gears, out of the inspected ones, which caught fish.

3.2.5.3 Fishing craft

From an economic point of view, the size and type of fishing craft owned by a FEU corresponds to the amount of capital invested by the unit. In our case the traditional fishing craft are (example):

Sesse-canoe

Dugout-canoe

Tomarica-canoe

Karua-canoe

etc.

3.2.6 The field operations of a Frame Survey (FS)

Generally speaking the main purpose of the "field operations" of a FS survey is to obtain the required items of information by using the methods which were developed at the design process of the survey.

For the field operations of a frame survey in inland waters the following three methods have been developed:

1. Method 1 - Water approach.
2. Method 2 - Aerial approach
3. A combination of the above two methods.

3.2.6.1 The water approach

In the "water approach" the main task of the "working group" (observer + assistant + speed boat) can be summarized as follows:

1. To carry out a mile by mile field survey (within each established stratum).
2. During the trip the observer has to look at all the shoreline following the topographical map. By making use of binoculars he has exactly to note all the fishing sites which he comes across and locate them carefully on the topographical map.
3. The group has to land at every detected fishing site and by following the instructions, to collect the required items of information.

Usually for the collection of the required information one of the following three methods can be used:

1. Items of information are collected by interviewing the chief of the fishermen and other competent persons of the fishing sites covered by the survey.
2. Items of information are collected by conducting a compound by compound survey in the residential area of fishing sites, covered by the FS.
3. A combination of the above two methods.

It should be noted that the effectiveness of a Frame Survey based on the water approach depends on the time needed to close the field operations. We must remember that, in inland waters, fishing economic units have a high level of mobility.

3.2.6.2 The questionnaire of a FS (water approach)

The principles used at the design process of the questionnaire of a FS (water approach) can be summarized as follows:

1. The questionnaire should be divided into two parts: a) heading of the questionnaire and b) body of the questionnaire.
2. Questions in the body of the questionnaire should be grouped into meaningful categories.
3. All questions have to be numbered by using a proper "numbering system".
4. A "code system" has to be used for a numerical identification of the fisheries sites.

Form F1 of Appendix Ia is an example of the format of the questionnaire used (water approach, method 1) in Frame Surveys at a big lake.

3.2.7 The aerial approach

This method calls for a bird's eye view from the air along the shoreline of the lake under study. The objects of the survey are:

1. To detect all the fishing sites along the shoreline of the lake.
2. To make the proper location of the fishing sites on the existing topographical maps.
3. To obtain an estimate of the size of each fishing site in terms of the number of "boats seen" and "houses seen".

Past experience has proved that:

1. The view from the air reveals the distribution of the fishing sites on the lake periphery.
2. There are no serious problems for the proper location of the fishing sites on the topographical map.
3. The counting of the fishing boats/houses is not a very difficult task.

This method has a number of advantages as well as some disadvantages:

Advantages:

1. An area frame of the fishing sites can be established within a short period of time (number of fishing sites, area distribution, size (in terms of fishing boats seen)).
2. Provides an indication of the importance of the various regions of the lake under study in terms of concentration of fishing boats.
3. The established area frame can be used for the selection of the samples of other surveys covering the same population.
4. These data can also be used to make the field operations of a frame survey based on the "water approach" covering the same population, more efficient in terms of time and effort.

Disadvantages:

1. This method does not reveal the type of the fishing sites, e.g. permanent fishing site, fishing camp, etc.
2. There is no chance of obtaining information on other basic characteristics featuring the fishing industry.
3. It is always desirable that the quality of this method be checked by using proper techniques.

3.2.7.1 <u>The questionnaire of the survey (aerial approach)</u>

For the collection of the information needed a relatively very simple questionnaire is used for the survey. Form F2 of Appendix Ia is a format of the questionnaire used.

4. THE MEANING OF COVERAGE ERRORS

4.1 <u>Sources of error (water approach)</u>

From a sampling point of view the following types of coverage errors might occur during the field operations of a Frame Survey (water approach):

1. Omissions of fishing sites: Omissions of fishing sites is one of the most serious errors in FS. This type of error affects both the total number of the existing fishing sites in the area under study, as well as the total number of the existing fishing economic units and their components (fishing boats, fishermen, fishing gear).

   Past experience has proved that the observers of a FS do not attach any weight to the size of unit covered by the survey. Thus, the probability of an omission of a small fishing site is equal to the probability of omission of a medium or large fishing site. This source of errors leads to a serious underestimation of the survey characteristics.

   The main sources of omission of fishing sites can be attributed to:

a) The carelessness of the observer.

b) Natural causes (e.g. camouflage of a fishing site by a thick fringe of trees etc.).

2. Counting errors: Incomplete coverage of the FEU's within the fishing sites covered by the survey. These errors affect the size of the fishing sites (under-estimation).

   This source of error is mainly attributed to the carelessness of the observer.

3. Erroneous inclusions: Another type of error altogether is the inclusion in the survey of purely agricultural villages or other kinds of units which are not covered by the FS. These kinds of errors increase artificially the size of population under study.

   This source of error is mainly attributed to the inexperience and carelessness of the observer.

4.2 Sources of error (aerial approach)

The following types of errors occur at the measurement process of an aerial flight:

1. Omissions of fishing sites: Specifically of small size.
2. Counting errors: Errors in the process of counting fishing canoes within the fishing sites covered by the survey. Specifically, counting errors in an aerial flight survey consist of two components:

   a) Omissions of fishing canoes which cannot be seen from the air.

   b) Inclusion of non-fishing canoes and abandoned canoes.

5. PLANNING A COVERAGE CHECK SURVEY

5.1 Introduction

Coverage Check Surveys (CCS's) have in recent years become a regular feature of the Frame Surveys in Inland Waters. For the design of a CCS modern techniques are used.

5.2 The purpose of a CCS

The main purpose of a CCS is to detect and estimate the magnitude of the coverage errors e.g. omissions and erroneous inclusions of fishing sites, omissions and erroneous inclusions of fishing economic units, which occured during the field operations of a FS.

Generally speaking, the total number of survey units covered by a Frame Survey is composed of two parts:

$$P = P_1 + P_2$$

where:

P = Overall number of units covered by the survey

P_1 = Erroneous inclusions

$P_2 = P - P_1$ = Correct inclusions

Further, the total number of survey units in the population under study can be written:

$$P = P_2 + P_3$$

where P_3 is the number of units omitted at the measuring process of the FS. Under these circumstances, the main goal of a CCS is to estimate the size of the magnitudes P_1 and P_3.

5.3 Designing a CCS

The CCS are intensive studies based on relatively small samples and every effort is made in them to attain the highest level of efficiency possible. In our case the main ingredients of a CCS can be summarized as follows:

1. The field work of the survey is carried out by the best personnel of the FS.
2. The "observation group" is instructed to carry out a mile by mile field survey in the selected area zones.
3. Items of information about the survey characteristics are obtained by using the intensive interview approach.
4. Every assistance is given to the "observation group" in order to ensure efficient field work.

For these and other reasons the results of the CCS are more reliable than the initial one (FS). We can assume that errors in the CCS are kept to the minimum.

The following estimates are provided by a CCS:

1. Estimation of the number of omitted fishing sites (and by major size groups).
2. Estimation of the total number of omitted fishing economic units (and their components).
3. Estimation of the number of erroneously covered fishing sites.
4. Estimation of erroneously covered units.

For the evaluation of the omissions, erroneous inclusions of fishing sites in a FS the following pattern is used (example):

FS	CCS		
	YES	NO	TOTAL
YES	A	B	A + B
NO	C	(D)	C + (D)
TOTAL	A + C	B + (D)	A + B + C + (D)

where:

A = fishing sites which have been covered by both the FS and the CCS

B = fishing sites which have been covered by the FS but not by the CCS. Further, B can be broken into two parts:

$$B = B_1 + B_2$$

where:

B_1 = correct inclusions in FS (omissions in CCS)

B_2 = erroneous inclusions in FS

C = fishing sites which have been covered by the CCS but not by the FS (omissions in the FS)

D = fishing sites which have been omitted in both surveys (D is an unknown magnitude unless a further CCS can be conducted)

Counting errors within the fishing sites covered by the FS are discovered by matching the data of the fishing sites identified in both the FS and CCS.

6. CATCH ASSESSMENT SURVEYS

6.1 Introduction

Statistics of fish catch is a valuable tool for sound management of a fishery and for planning, marketing and other aspects of fishery development.

The gathering of these statistics from African freshwater fisheries is complex. The fishermen are frequently widely distributed along thousands of miles of lakeshore and often land their catches at hundreds of sites. Further, various tribes are engaged in fisheries on a lake and since their fishing techniques and fishing capacities vary considerably, there is lack of uniformity in the economy. There is a good deal of mobility of the fishermen around the coastline of the lake during the year. Finally, the production of the lake is subject to regional variations as well as seasonal variations within the year. Since the population is subject to change, a survey carried out on a single occasion or on a static basis cannot of itself give any reliable estimate of fish production on a yearly basis, nor any information about the nature and rate of such a change.

6.2 Designing a Catch Assessment Survey (CAS)

From the above analysis it is obvious that a number of factors must be taken into account at the design process of a CAS.

6.2.1 Area stratification

It is well known that the productivity of the lake is a function of the prevailing limnological conditions and that these factors are not the same throughout the lake. If the surveyor had wished to ensure that different types of limnological zones (LZ's) were adequately represented in the sample he could have stratified the area under study by using as criteria of stratification limnological factors. Past experience has proved that this kind of stratification increases the representativeness and the precision of the sample.

6.2.2 Stratification of the fishing sites within the established LZ's

In order to take full advantage of possible gains from stratification the sample design calls for a proper stratification of the fishing sites within the established LZ's. The LZ variation (regarding any variable) can be reduced considerably if fishing sites within each LZ are divided into groups (minor strata) taking their size (e.g. number of boats) as the criterion of stratification. For sampling purposes, a number of fishing sites should be selected within each established minor stratum.

6.2.3 Sampling within the selected fishing sites

Within each selected fishing site one may decide to collect information from all the fishing economic units of the fishing site or to select a small sample of fishing economic units for further investigation. Modern sampling techniques suggest that the latter alternative is the best one from cost and precision points of view.

6.3 Types of CAS

By taking as criterion of classification the measurement procedure used for the collection of the information, CAS's can be grouped into three types:

1. CAS's based on the interview approach.
2. CAS's based on the real measurement approach.
3. CAS's based on a mixed approach.

In type 1 surveys the required items of information are obtained by interviewing the respective fishermen e.g., a set of questions are asked and the answers are recorded in a standardized form. It is obvious that the reliability of the results of the survey depends on the memory of the respondents and whether they are willing to provide the true information.

In type 2 surveys items of information are obtained by actual measurement of landings. In this type of survey the reliability of the results are not affected by the respondent (memory errors, etc.).

6.4 The reference period of the survey characteristics of CAS

It was the erroneous belief in the past that information for the items covered by a CAS must be collected all the year round. However, this approach hardly can be justified by the modern sampling techniques.

Highly reliable results with a minimum cost and effort can be obtained if we are in a position to determine the "optimum" reference period of the characteristics under study. It has been proved that within the various seasons of a fishing year the optimum reference period of the survey characteristics had a cycle of three to seven days. In such a case the extension of the reference period for example from three days to a month or to a quarter would result in a mere repetition of the cycle, which means that the sampling error would not be significantly affected, whereas the cost and the quality of the survey will be affected badly.

6.5 Listing

Listing of the survey units within the sample fishing sites is an important function in cases in which modern sampling techniques are going to be used for our surveys. The obtained information can be used:

1. For the selection of our samples within the sample fishing sites.
2. To provide an estimate about the mobility pattern of the fishermen.

6.5.1 What to list

Within each sample fishing site lists should be used to list:

1. Fishermen boat owners (by type).
2. Fishermen with fishing gears only.

6.5.2 How to list (general principles)

The main parts of a statistical list are:

1. The heading of the list.
2. The body of the list.

By listing we mean writing down on prepared forms (lists) the identification particulars of all the units of investigation within the sample fishing site.

It is important that the listing be:

1. Complete: in the sense that it covers the whole of the units to be surveyed.
2. Accurate: that is, the information for each unit listed on the form should be free of errors.

Rules for listing:

1. List each unit on a separate line.
2. Do not skip any lines within the body of the list.
3. List each unit only once.
4. If there are doubts use the column of remarks.
5. Give page numbers to the additional sheets and insert the code number of the fishing site.

6.6 Interviewing

6.6.1 Introduction

An interview (CAS type 1) involves a meeting between two persons; the recorder and the respondent (fisherman). Further, the respective questionnaire(s) is (are) to be used for obtaining the required items of information from the respondent.

6.6.2 How to ask the questions

The way in which you ask the questions is a matter of great importance because it is essential that we collect data from all respondents in a uniform manner. Therefore, the recorders must ask the questions in the same way and in the prescribed order.

6.6.3 Some rules for interviewing

In all African countries a lot of importance is placed on greetings. You must therefore remember to greet the people you meet in the place you have to interview. This might mean the difference between success and failure.

You must know what your goals are and be prepared to lay your cards face upwards when you are confronted with rumours i.e. explain your programme as precisely as you can.

You must know the reason for asking certain questions on your sheets, e.g. the fishermen do not want to tell you their fish catches for fear of taxation. Explain why you want it.

Listen well. This is evidence of your respect for the individual and your interest in him.

The manner and tone of your voice should convey friendliness and willingness to understand.

Avoid during the interview comments that have a negative flavour, e.g. "So you are a lazy man".

Wives are often unwilling to give information when their husbands are not present. Do not force them. Call on the house again.

When you sense that the man is drunk, carefully put off the interview. Call in another time.

You must have a sense of humour in everything you do. This releases steam. Say what you have to say with a smile.

6.6.4 How to close the interview

The way in which you close the interview can have a definite effect on your future relationship with the respondent. Always try to leave him with a friendly feeling toward both you and your office so that you will be welcomed on subsequent visits to the fishing site.

6.7 Real measurement

In CAS type 2 items of information are obtained through actual measurements made by recorders on selected landings within the sample beaches. Some of the principles employed for the measuring process are as follows:

1. Before the fishing canoes return make certain you are ready to record.
2. You ought to have a seat of some kind so you can write easily on the clip-board. A rag is useful for wiping your hands after handling fish or gear. Keep your pencil sharpened so you can write neat figures in the column of the questionnaire. Always use a pencil (ball point and water equal no results). Your assistant should carry his scales and other equipment in a plastic pail so you can easily move from one selected canoe to the other. The pail can serve as a seat.
3. When a selected canoe lands ask the fisherman to show you what he has caught. Big fish will be lying in the bottom of the canoe and you can usually collect each species together and weigh them before someone takes them away. Small fish will probably still be in the net and you can weigh them as they are picked out of the nets.

6.8 How to select a sample of second stage units

In our case the selection of a sample within a fishing site is a mechanical operation. For example, a list contains N=9 units and a sample of n=3 units are going to be selected. The selection is carried out as follows:

i) select the table with heading N=9 (tables are provided);

ii) read the second column of the table where there are numbers in brackets;

iii) ring the serial numbers on the list. These are the same units of the sample; in this case numbers 3, 6 and 9 on the list are the selected units;

iv) interviews (real measurements) should be completed for the selected units only and the collected information should be recorded on the respective questionnaires.

Example: the format of a table (N=9, n=3)

Number of units in the list	Selected units	Sample units
N=9	(3) (6) (9)	1 2 3

6.9 Source documents

For the collection of the required information two kinds of source documents (questionnaires) are used (see Appendix Ib).

Form: A1-3: This questionnaire is used to obtain the required items of information on the static characteristics of the selected fishing economic units. The selection of the fishing economic units is made at the residential area of the fishing site. For the collection of the information the method of "personal interview" is used.

Form: A2-3: In this form items of information are sought on the dynamic characteristics of the sample fishing economic units (fishing effort, fish catch). The selection of the units is made at the beach (landing place) of the fishing site. For the collection of the information the method of "real measurement" is used.

APPENDIX Ia - THE SOURCE DOCUMENTS OF A FRAME SURVEY (EXAMPLE)

Form: F1

FRAME SURVEY

Name of the Recorder ____________________

Survey date [ ] [ ] [ ]

C.No. of Fishing Site [ ] [ ]

ITEMS OF INFORMATION	
1. Identification particulars of the fishing site	1. Name(s) of the fishing site: 2. Tribe(s) of fishermen: [ ]
2. Organic structure of the fishing site	(ASK): 1. Is the occupation of the fishing site by the fishermen: continuous [ ] 1 sporadic [ ] 2 2. Are the fishermen of the fishing site: permanent [ ] 1 transient [ ] 2
3. Migration history	(ASK): 1. When did they first commence fishing on the lake? ________ [ ] 2. Before the formation of the Lake were they: 1. Fishermen [ ] Number ______ where ______ ______ [ ] remarks: 2. Farmers [ ] Number ______ where ______ ______ [ ] remarks: 3. Other [ ] Number ______ where ______ ______ [ ] remarks: (ASK): 3. When did they come to the present place? ________ [ ] 4. Where were they staying before they arrived in this place? 4.1 Name of the place ________ 4.2 What kind of place was it? ________ [ ] 4.3 Where was it located? ________ [ ] 5. How long did they stay in the previous place? ________ [ ] 6. Do they intend to move again? Yes [ ] 1 No [ ] 2 6.1 If YES: 1. why? ________ [ ] 2. when? ________ [ ] 3. where? ________ [ ] General remarks on page 1 of the Form:

P2 **Form: F1**

Survey date [ ] [ ] [ ] C.No. [ ] [ ]

4.Fishing periods: (experience of the last year)	(ASK): 1.Do they fish all the year round? Yes [ ] 1 No [ ] 2 1.1 If NO, during which period(s) do they fish? 1.period: from ______ to ______ [ ] [ ] 2.period: from ______ to ______ [ ] [ ]
5.Fishing gear used: (experience of the last year)	(ASK): What kind of gear do they use for fishing (complete the following table):

Gear C.No.	Gear Name (1)	Period used (2)	Remarks (3)
01		from ______ to ______	
02		from ______ to ______	
03		from ______ to ______	
04		from ______ to ______	
05		from ______ to ______	

6.Fishing boats, Fishermen [local fishing boats which were absent on the survey day (e.g. fishing, at market) must be included in the table]

(ASK): Complete the following table:

C.No.	1.Fishing boats by kind: Name (1)	Number (2)	2.Number of fishermen: Owners (3)	Assistants (4)	Remarks
01	Canoes				
02	Plank boats (without engine)				
03					
04					
05					

09 Number of fishermen without fishing boat working for themselves ______

General remarks on page 2 of the Form:

Form: F1

P3

Survey date [ ] [ ] [ ] C.No. [ ] [ ]

7. Fish catch: (insert a ✓ in the proper box(es))

(ASK): Fish mainly caught:

1. Tilapia [ ]
2. Alestes [ ]
3. [ ]
4. [ ]

8. Disposition of fish catch: (insert a ✓ in the proper box(es))

(ASK): 1. Complete the following table:

Marketed state of fish:		Disposition of fish catch:			
C.No.	Name (1)	Total (2)	Almost total (3)	About half (4)	Less than half (5)
1	Fresh				
2	Smoked				
3					
4					
5					

(ASK): 2. Do fish traders come to the place? Yes [ ] 1 No [ ] 2

2.1 If YES:

1. How ________ [ ]
2. How often ________ [ ]
3. From where ________ [ ]

3. Do fishermen (or wives etc) go to market in order to dispose of their catches? Yes [ ] 1 No [ ] 2

3.1 If YES:

1. How ________ [ ]
2. Where ________ [ ]

9. Capital goods supply centers

(ASK): 1. From where do they buy their canoes ________ [ ]

1.1 Usual purchased value of a new canoe ________

2. From where do they buy their gears ________ [ ]

2.1 Complete the following table:

Gear		Unit	Price per unit	Remarks
C.No.	Name (1)	(2)	(3)	(4)
01				
02				
03				
04				
05				

General remarks on page 3 of Form:

Form: F2

FRAME SURVEY
(Aerial Approach)

Flight number.........................

Name of observer......................

Survey date...........................

Survey time...........................

Sample code No. of the survey area [ ]

Fishing sites seen		No. of boats* seen			No. of houses seen	Remarks
Serial No.	Topographical map particulars	Total	On the beach	On the water		
(1)	(2)	(3)	(4)	(5)	(6)	

* A separation should be made between plank boats and canoes

APPENDIX Ib - THE SOURCE DOCUMENTS OF A CAS (EXAMPLE)

CATCH ASSESSMENT SURVEY OF

Name of the recorder..................

Date..................................

FORM: A1-3

PS1

Fishing Site ( ) PSU

Sample C.No [ ] [ ]

Name......................

Location..................

ITEM	INFORMATION
1. Identification particulars of the fisherman	(ASK): Fisherman's last name 1. Marital status: single [ ] 1 married [ ] 2 other [ ] 3 2. Tribe [ ] 3. Home town [ ]
2. Boats, fishermen.	(ASK): Complete the following table for the selected fisherman/boats

Serial number	Boats		Crew			Remarks
	Type	Age	Owner	Adult assistants	Boys	
1	(1)	(2)	(3)	(4)	(5)	

3. Fishing area

(ASK): Where do you fish?

1. Depth of water		2. How far away (hours of paddle)	
Close inshore	[ ] 1	Less than 1 hour	[ ] 1
Within trees	[ ] 2	1-2 hours	[ ] 2
Beyond trees	[ ] 3	2-3 hours	[ ] 3
Open water	[ ] 4	3-4 hours	[ ] 4
		More than 4 hours	[ ] 5

4. Fishing nets

(ASK): What nets does he use in fishing?

Mesh size	1. Gill nets		2. Cast nets		Remarks
	Owned	Fished to-day	Owned	Fished to-day	
	(1)	(2)	(3)	(4)	
TOTAL					
2 Inches					
2½ "					
3 "					
3½ "					
4 "					
5 "					
6 "					
7 "					
8 "					
9 "					

5. Other fishing gears

(ASK): What other fishing gears does he use in fishing?

Type of gear	Code no	Owned (1)	Fishing to-day (2)	Remarks
1. Long lines	3			
1a. Number of hooks per line				
2. Hand lines	4			
3. Traps	5			
4. Other (specify)	6			

CATCH ASSESSMENT SURVEY OF | **PS 2 Survey Form A2-3** | **Name of the fishing site:** ( )

Name of recorder........................ Sample C.No. [ ] [ ] [*] Sample date [ ]

Sample period.......................... Form A2-3 [ ] Survey period [ ]

			* DAY 1									* DAY 2									* DAY 3								
Punch card 1	2. Fisherman's name																												
	3. Boats:	Type																											
	4. Size of crew																												
Punch card 2	5. Fishing gear used:		GN									GN									GN								
	a. Code number																												
	b. Number checked																												
	c. Mesh size																												
	6. FRESH FISH Code No.	Species	No	G	Weight	No	G	Weight	No	G	Weight	No	G	Weight	No	G	Weight	No	G	Weight	No	G	Weight	No	G	Weight	No	G	Weight
	01	CITHARINUS																											
	02	ALESTES																											
	03	LABEO																											
	04	LATES																											
	05	SYNODONTIS																											
	7. Total time spent in fishing (hrs)																												

G=Non-Gutted=0 G=Gutted=9

APPENDIX Ic - EXERCISES

Exercise No.1

A fish pond contains 10 fish. Their weight (lbs) are given below (Table 1). (a) What is the population total weight? (b) What is the number of fish with weight: (i) Up to 1 lb? (ii) More than 1 lb?

Table 1

Serial No. of fish	Weight (lbs)
1	1.5
2	0.5
3	0.5
4	1.0
5	1.0
6	2.5
7	1.5
8	0.5
9	0.5
10	1.0

Exercise No.2

A fishery place is divided into two zones (south and north). In the south zone there are four fishing sites and in the north six fishing sites. The table below (Table 2) provides the number of fishing canoes by fishing sites. What is the total number of fishing canoes by zone and for the fishery place as a whole?

Table 2

Zone	Ser. No. of fishing sites	No. of fishing canoes
I	1	20
	2	5
	3	15
	4	40
II	1	5
	2	6
	3	10
	4	15
	5	20
	6	50

Exercise No.3

A FS was conducted at a given lake. According to the design of the survey the area was divided into five strata of equal size based on the length of map shoreline. Further, the "water approach" was used to collect items of information. Prepare a figure showing an ordinary allocation of the fishing sites at the lake. Allocate an arbitrary number of canoes to each fishing site and calculate the total number of canoes by stratum and for the lake as a whole.

Exercise No.4

What is the difference between the following two methods of collecting items of information: (i) Interviewing approach (ii) Real measurement approach.

Exercise No.5

Explain the meaning of the coverage errors in a FS (water approach).

PART II: INTERMEDIATE COURSE

7. INTRODUCTION

7.1 The structure of a fishing industry

Generally speaking a fishery in an advanced area is a great complex of inter-related assemblance of fishing vessels, fishing equipment, harbour workers and machinery for fish handling, preservation and processing. There are market and storage buildings, railway sidings, plants producing fertilizer, fishmeal and oil, factories providing nets, ropes and ice, specialized shipbuilding yards and finally the houses and other social capital of the communities needed to operate all these aspects of the fishery.

In small fishing unit economies the state of the fishing industry is very simple and its organic whole can be considered rather as primitive. The main characteristics of the industry are:

1. The fisheries are diffuse, disordered and in the first stage of evolution.
2. Fishing activities are mainly carried out on a private (family) basis and the predominant type of fishing craft is the fishing canoe.
3. There is a lack of storage facilities and the handling, preservation and processing of fish is the responsibility of the fishermen (family members).
4. The marketing system is not very well organized and there are some problems for a proper disposition of the fish catch.

7.2 The division of a fishery

It is convenient in economic and statistical studies to divide a fishery into three distinct phases:

1. Primary phase: fish production.
2. Secondary phase: fish processing. This section covers the conversion of the primary product (catch or yield) into preserved commodities.
3. Tertiary phase: fish marketing. This section covers the distribution of fish and fishery products from producer to consumer.

7.3 The fishing economic unit (FEU)

The total fish production (fish catch, yield) of a fishing industry, from the viewpoint of an economic study, is the result of the operations of the fishing economic units (FEU's) within a given period of time. Specifically, in our case, FEU's fall into two categories:

1. Usual Fishing Unit (UFU) which is an integrated unit composed of fishing craft, fishing gear and fishermen to carry out fishing operations.
2. Minor Fishing Unit (MFU) which is the integral unit composed of fishing gear and fisherman (without fishing craft) to carry out fishing operations.

By taking as criterion of classification the state of the ownership, the FEU's can be divided into three categories:

1. Private ownership.
2. Agreed ownership.
3. Cooperative.

8. SAMPLING SURVEYS IN FISHERIES STATISTICS

8.1 Why Sampling Surveys?

In its broadest sense the purpose of a Sample Survey is the collection of information to satisfy a definite need. A Sample Survey has now come to be considered as an organized fact-finding instrument. Its importance nowadays lies in the fact that it can be used to summarize, for the guidance of administration, management, etc., facts which would otherwise be inaccessible owing to the remoteness and obscurity of the persons or other units concerned, or their numerousness. As a fact-finding agency a Sample Survey is primarily concerned with the accurate ascertainment of the individual facts recorded and with their compilation and summarization.

8.2 Type of fisheries statistical surveys needed

For the collection of the items of information covered by fisheries statistics the following types of statistical surveys are needed.

8.2.1 Primary phase (fish production)

8.2.1.1 Frame Survey (FS)

A Frame Survey is a sort of inventory survey. Its main aim is to provide accurate information about the size and structure of the fishing industry under study.

The methods used for the collection of the information in a FS are either "aerial approach" or "water approach" or a combination of both of these.

The survey items covered by a FS are:

1. Number, size (in terms of fishing boats) and area distribution of the fishing sites.
2. Number of fishing economic units (FEU's) of the industry and of their components (fishing craft, fishermen, fishing gear).
3. Migration of the FEU's.
4. General information about the methods used by the fishermen for processing and marketing their catches.

From a sampling point of view, a frame survey is a census survey based on the method of complete enumeration of the survey units.

Since a frame survey secures a complete coverage of the population under study, it provides an ideal "sampling frame" for the selection of the samples of other surveys covering the same population.

8.2.1.2 Catch Assessment Survey (CAS)

The primary objectives of a CAS are to obtain estimates on a current basis of total catch in the lake (river), and on a regional basis. Secondary objectives include the estimation of the species composition of the catch and the fishing effort involved in obtaining the catch.

The methods used for the collection of the items of information in a CAS are either "objective measurements" or "subjective measurements" or a combination of both methods.

For sampling purposes the method of sampling used for the survey is "sampling in space and time".

It has been proved that there is a quantitative relationship between the results of a CAS and the results of a Stock Assessment Survey (SAS). The established mathematical models are of great value for management purposes.

8.2.1.3 Cost and Earning Survey (CES)

CES is a fishery economic survey. One of its main objectives is to study the size of profit derived from individual fisheries undertakings. Also, the results of the survey can be used to estimate the cost of production of different types of fishing economic units in order to evaluate the existing fish price system.

The sampling method used for the survey is a "two-phase stratified sample". In this case a sub-sample is selected from the main sample of the Catch Assessment Survey. Before any selection is made the sample units are grouped into strata, by taking as criterion of stratification the type of the fishing economic unit and a few units are selected, with equal probabilities, from each established stratum.

For profitability studies the items covered by the survey are:

1. Current expenses which are directly spent for each fishing operation.
2. Wages paid to employed labourers (estimated cost of unpaid family members), repairing expenses for hull (engine), depreciation cost for fixed assets etc. The profit gained by the individual fisheries undertaking equals:

   (Total fish sale)-(Above items, 1+2)

8.2.2 Secondary phase (processing)

8.2.2.1 Fish Processing Survey (FPS)

In a large fishing unit economy a Fish Processing Survey can be considered to consist of two parts:

1. An inventory survey where the main objective is to determine the structure, capacity and organization of the processing industry.
2. A production survey where the main objective is the collection of information about the volume and value of processed aquatic animals and of the products produced.

In small fishing unit economies curing of fish is done mainly by the fishermen (relatives of the fishermen) themselves or by minor fish processors who live in the fishing villages. The processing takes place soon after the fish are landed at the producing fishing sites. Further, some fish which have already been processed are reprocessed mainly at market landing places to get higher grade commodities. It is obvious that higher grade commodities are of a different nature than the usual ones, especially from the viewpoint of measures for fish price maintenance. Reliable estimates of the survey characteristics with a minimum cost can be obtained by integrating the survey design of a Fish Processing Survey with the designs of the Frame Survey and Catch Assessment Survey. This principle was successfully employed in a number of inland water statistical surveys in Africa.

8.2.3 Tertiary phase (marketing)

8.2.3.1 Fish Marketing Statistical Survey (FMSS)

The main objectives of the survey are the collection of information on the quantity of fish transacted and the corresponding price of fish at the wholesaler stage. Other objectives of the survey are to trace the marketing routes of fish transacted and to study the structure of retail fish markets and the price of fish purchased by the consumer.

A marketing statistical survey for wholesale transactions can be planned either as a PSS - at producing fishing sites, or a MSS - at market landing places. The design of a PSS - at producing fishing sites, is usually integrated with the design of the Catch Assessment Survey covering the same population. In such a case the survey unit is the fishing economic unit. The MSS - at market landing places, is conducted independently from the CAS and the required items of information are obtained from the fishmongers.

Table 8.2.3.1.1 gives an idea of the existing relationship between survey items and type of surveys in fisheries statistics.

Table 8.2.3.1.1 Relation between survey items and type of surveys (Fisheries Statistical Surveys)

Phase	Type of survey	Survey unit	Survey items (groups)
A.Primary phase	1.Frame Survey	Fishing site/Fishing Economic Unit (FEU)	1.Number of FEU's 2.Area distribution of FEU's 3.Ingredients of the FEU's: 3.1 No.of fishing craft 3.2 No.of fishermen 3.3 No. of fishing gear 4.Mobility of the FEU's 5.General information on processing and marketing habits of the FEU's
	2.Catch Assessment Survey (CAS)	Fishing Economic Unit	1.Fish catch (total, regional basis) 2.Species composition of catch 3.Fishing effort items
	3.Cost Earning Survey (CES)	Fishing Economic Unit (by type)	1.Total revenue 2.Cost: 2.1 Running cost 2.2 Wages, etc. 2.3 Maintenance and repair boat/engine 2.4 Repair or renewal fishing gear 2.5 Other charges 2.6 Depreciation cost (hull, engine) 3.Estimated amount of capital invested and interest of the amount of capital invested 4.Amount of taxes (levies) paid
B.Secondary Phase	1.Fish Processing Survey (FPS)	Processing unit	1.Number of processing units by type 2.Processing capacity 3.Processed products
C.Tertiary Phase	1.Fish marketing statistical surveys	Marketing unit	1.Number of marketing units 2.Quantity of fish transacted 3.Price of fish at wholesaler 4.Price of fish paid by consumer

9. THE LABOUR FORCE OF A FISHING INDUSTRY

9.1 Determining the labour force

Manpower statistics consist mainly of separating population data into the categories given, and comparing them. However, the users of labour force data should have in mind the existing differences among the following three distinct magnitudes of labour force statistics.

9.1.1 Population of working classes

It is convenient to treat the number of people from 15 to 64 years of age as the group supplying the bulk of the economically active, calling it the "population of working classes". In our case, the population under study can be characterised as one with a relatively high fertility. It gains a large number of younger people each year and the age structure is weighed with a large proportion of children. According to the local standards the "population of working classes" includes children with ages of less than 15 years.

9.1.2 Economically active population

Having defined the population of working classes the next step is to set up a scheme for determining which people are "economically active" and which people are not. This implies a standard for judging what activities constitute "productive work" and some criteria to judge what degree of performance is sufficient to class a person as "active".

9.1.3 Economically active population by industry

We have defined above the economically active population as persons reporting a productive work (from an economic point of view). These persons can now be classified into groups by taking as a criterion of classification "where the person is employed" i.e. in which sector of the economy (industry) the person is employed (fishing, agriculture, building construction, etc.).

In our case we are going to deal with only the labour force of the fishing industry.

9.2 Labour force by sub-sector of the industry

It has been mentioned that the fishery is divided into three phases (sections):

1. Primary phase: fish production.
2. Secondary phase: fish processing.
3. Tertiary phase: fish marketing.

In such a case the total labour force of the industry (L) can be reclassified into groups by taking as a criterion of classification the structure of the industry:

$$L = L_1 + L_2 + L_3$$

where:

L_1 : labour force of the primary phase (LF - P.Ph)

L_2 : labour force of the secondary phase (LF - S.Ph)

L_3 : labour force of the tertiary phase (LF - T.Ph)

9.3 Labour force by type of activity

Within each sub-sector of the industry a further grouping of the labour force can be introduced by taking as a criterion of stratification the actual type of work (occupation) of each individual. Table 9.3.1 gives an idea of the occupational structure of the fishing industry on a sub-sector basis.

Table 9.3.1 Distribution of the Economically Active Population by occupation within each sub-sector of the fishing industry

Sub-sector	Occupation
1. Primary sector	1.1 Fisherman 1.2 Boy assistant fisherman of age less than ... years 1.3 Unpaid family member who assists the work relating to the fishing operation 1.3.1 Loading fishing material 1.3.2 Unloading fish 1.3.3 Net repairing 1.3.4 Boat repairing 1.3.9 Other (specify) 1.4 Net maker 1.5 Boat maker 1.6 Porter 1.9 Other (specify)
2. Secondary sector	2.1 Processing unit 2.1.1 Fisherman 2.1.2 Wife (relatives) of the fisherman 2.1.3 Other (specify)
3. Tertiary sector	3.1 Fish trader (wholesale) 3.1.1 Trader-wife (or relative) of the fisherman 3.1.2 Local trader 3.1.3 Non-local trader 3.1.4 Other (specify) 3.2 Fish trader (retail) 3.2.1 Itinerant trader 3.2.2 Permanent trader

9.4 Fishing labour force

From the above analysis it is obvious that the fishing labour force constitutes only a part of the labour force of the primary phase (LF - P.Ph) of a fishing industry. Specifically, fishing labour force can be defined as the number of persons who engage in actual operation of capture or culture of aquatic resources. Therefore persons who participate in the work relating to fishing operations, (unloading fish, net repairing, etc.) who never go on the water are units of another component of the LF - P.Ph, and not of the fishing manpower.

9.4.1 Measuring the fishing labour force

For measuring the fishing labour force one of the following two approaches can be used:

1. Census approach.
2. Labour force approach.

According to the "census approach" fishermen are simply those who report participation in fishing operations during the previous year.

According to the "labour force approach" fishermen are those who actually participate in fishing operations during the survey period (week).

It is obvious that the above two approaches reflect different conceptions of the nature of the economic activity. Since a fishery is highly seasonal the labour force approach can be used to measure fluctuations in the number of fishermen over time.

9.4.2 Classification of fishermen

Fishermen may be classified according to the "time spent for fishing" within a year:

1. Full-time fishermen.
2. Part-time fishermen.
3. Occasional fishermen.

Another classification of fishermen is according to "employment status".

1. Fishermen canoe owners.
2. Fishermen with gear only working for themselves.
3. Assistants.

10. RESPONSE ERRORS AND SUPERVISION

10.1 Response errors

10.1.1 What is meant by response errors

It is reasonable to make the assumption that for the "survey unit" covered by a survey there is always a true value (Individual True Value ITV) for the characteristic under study. In other words, it is reasonable to assume that there is a true value y_i attached to the unit U_i in the population.

A recorder assigned to collect information on unit U_i plays a role of a person who is trying to shoot at a target. In only some cases will he succeed and the number of "successes" in a survey will depend on:

1. The nature of the question.
2. The way the question is put.
3. The experimental conditions of the survey e.g. how much precaution has been taken at the process of designing the survey to minimize the chance of measurement errors.

10.1.2 The meaning of IRE and TRE

In any case the difference between an ITV and the value "recorded" on the questionnaire is the Individual Response Error (IRE).

If, for example, the ITV of nets owned by a fisherman is 1000 yards and the value "recorded" is 600 yards there is a response error.

The aggregate of IRE's we term Total Response Error (TRE).

It is obvious that the seriousness of the TRE in a particular survey will depend on the extent to which the IRE's cancel each other out.

10.1.3 Main sources of response errors

10.1.3.1 Asking the questions

Recorders are usually instructed on:

1. How to greet the respondent.
2. How to ask the questions e.g. keep to the wording and order on the questionnaire and to ask the questions in a stated manner (in case of "real measurements" there are special instructions which the recorder has to follow up).

If the recorder does not follow the instructions it tends to produce a source of response errors.

10.1.3.2 Probing

In spite of instructions recorders may differ in the extent to which they probe in order to arrive at what they consider to be accurate response. Differential probing undoubtedly gives scope for the operation of response errors.

10.1.3.3 Recording the answer

Carelessness in recording is another potential source of error. Potential errors are:

1. Recorders may omit to record answers.
2. Recorders may record answers incorrectly.

10.1.3.4 Cheating

An altogether different cause of response error is conscious distortion or cheating.

It should be stressed that survey designers have their own methods for detecting cheating and this is the case in our surveys.

10.2 Supervision

The success of a method using the interview method (real measurement) depends largely on the ability of the recorders to elicit acceptable responses. Their selection and training is very important.

Observations by supervisors during the course of the survey operations is valuable for maintaining standards and keeping the quality of the recorders at an acceptable level.

10.2.1 The supervisor's task

The supervisor is responsible for:

1. Arranging payment of recorders for their monthly pay and their per diem claim.
2. Simplifying the field movements of the recorders e.g. providing transport etc.
3. Solving queries etc., which are referred to him.
4. Checking the completeness and consistency of the completed questionnaires. Every three months the supervisor should prepare a quarterly report for each recorder. It describes sources of errors detected in the course of reviewing their work in the office. As required, the report should specify what special steps should be taken to avoid making similar errors in the future.

10.2.2 Field supervision

It has been said that recorders are human beings and therefore liable to make mistakes. It is therefore advisable to keep the quality of the field work constantly under review and to investigate any case where a recorder appears to be doing unsatisfactory work.

The main objects of field work checks are:

1. To check the way the recorders select the samples of "survey units".
2. To test whether a recorder in fact made all interviews claimed.

3. Whether his response rate is satisfactory.
4. Whether he is asking the questions and interpreting and recording the answers in accordance with instructions.

11. JUDGEMENT AND RANDOM SAMPLE

11.1 Sample versus complete enumeration

From a sampling point of view information on a population may be collected in two ways:

1. Complete enumeration or census. In a census every unit in the population under study is enumerated.
2. Sample enumeration or sample survey. In a sample survey enumeration is limited to only a part or a sample selected from the population.

The main advantages of a sample survey over census are:

1. A sample survey is less costly than a census.
2. It takes less time to collect and process data from a sample survey than a census.
3. The results from a well planned and well executed sample survey are more accurate than those from a complete census that can be taken.

11.2 Judgement and random sample

For the collection of information on a sampling basis the following two methods can be used.

11.2.1 Judgement sample

This method of sampling is mainly used when the purpose of the survey is to obtain some indications of the survey characteristics in a relatively short period of time. To give some examples of judgement sample surveys:

1. Information can be collected almost inexpensively by asking persons known as experts in the subject.
2. Information can be obtained from a few units of the population that appear to be representative of the universe under consideration.

The above procedures are rejected outright by the survey statisticians because we do not know any objective method of measuring the confidence to be placed in the results obtained when the sample is selected by judgement. No amount of adjustment and refinement of data from a convenient sample can obscure the two basic weaknesses of the estimates i.e. that they are subject to unknown systematic errors, just that the sample does not provide any basis for objective evaluation of their precision.

11.2.2 Probability sample

The picture completely changes as soon as we begin using a sampling procedure in which "every unit belonging to the population has a known and non-zero probability of being selected in the sample". This method of selection is called "probability sampling" or "random sampling".

The importance of randomness in the selection cannot be overemphasized. It is an essential part of protection against systematic errors and the whole theoretical framework of probability theory rests on it.

To ensure randomness the method of selection must be independent of human judgement. There are two basic procedures:

1. The lottery method: Each member in the population is represented by a disc, the discs are placed in an urn and well mixed and a sample of the required size is selected (either with or without replacement).
2. The use of random numbers: The members of the population are numbered from 1 to N and n (n<N) are selected from one of the tables in any convenient and systematic way. These become the sample.

Both procedures are independent of human judgement and ensure randomness; the first because the units of the population can be regarded as arranged randomly, the second because the numbers used for making the selection have been generated by a random procedure.

11.3 Sampling in space and time

From a sampling point of view sampling surveys are of two types:

1. Static sample surveys or sample surveys carried out on a single occasion: with the objective of determining the characteristics of the surveyed population at or about a given point in time.
2. Dynamic sample surveys or sample surveys over time: these surveys are mainly used when the population is subject to change and information should be collected on the nature or rate of such change.

The sample method used for dynamic sample surveys in fisheries statistics is that of sample in space and time. A sample of area units first are selected on a random basis. The sample area units are randomly allocated to a number of time periods. Items of information are selected only from the selected area/time units.

12. PROCESSING THE RESULTS OF A SAMPLING SURVEY

12.1 Introduction

With the field part of the survey completed, the processing of the material and the highly skilled task of analyzing begins. The main ingredients of the processing of sample data in inland fisheries statistical surveys are:

1. Editing.
2. Coding.
3. Estimation.
4. Tabulation.
5. Presentation and interpretation of results.

12.2 Editing

Before the completed questionnaires can be regarded ready for further processing they should be checked for completeness, accuracy and uniformity.

1. Completeness: the first point to check is that there is an answer for every question. The main sources of incompleteness of a questionnaire are:

   1.1 The respondent refused to give an answer.
   1.2 The recorder forgot to ask the question or to record the answer.

2. Accuracy: it is not enough to check that all the questions are answered; one must try to check whether the answers are accurate. Inaccuracy may be due to carelessness or to a conscious attempt to give misleading answers, and it may arise from either respondent or recorder. It should be noted that answers needing arithmetic, even of the simplest kind, often cause trouble.

12.3 Coding

The source documents used for our surveys are pre-coded questionnaires and only a few questions have to be coded after the collection of the respective information. Post-coding is usually done at the editing time of the source documents.

12.4 Estimation

Estimation involves the use of the sample data in order to get estimates for the population characteristics. The estimation procedure (manual estimates) involves:

1. Transfer the items of information from the source documents to the Working Sheets (WS).
2. Working Sheets have been designed in such a way that to get an estimate very simple calculations should be conducted.

12.5 Tabulation

From the Working Sheets the obtained estimates are transferred to the Summary Working Sheets (SWS) which are the detailed tables of the survey. From the detailed tables the summary tables are constructed. These tables are usually relatively small in size and are designed to set forth one finding or a few related findings as effectively as possible.

12.6 Preparation of reports

When the analysis of a survey has been completed it is usually necessary to embody the results in a report. The layout of a report has been covered by a memorandum prepared by the United Nations Sub-commission on statistical sampling entitled Recommendations Concerning the Preparation of Reports on Sampling Surveys. The main points of recommendations are as follows:

1. Purposes of the survey.
2. Coverage.
3. Items of information.
4. Design of the survey.
5. Survey operations.
6. Cost.
7. Personnel and equipment.
8. Accuracy of the survey.

13. QUANTITATIVE RELATIONSHIP BETWEEN BASIC VARIABLES

13.1 Introduction

One of the chief objectives of science is to estimate the values of one factor by reference to the values of an associated factor. When the relationship between two factors is of quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in brief formula is known as "correlation".

It may surprise some of us to know that there is very close relationship between a good number of basic variables in the field of fisheries statistics.

13.2 Some important statistical variables

The user of the results of inland fisheries statistical surveys in African countries should have in mind the existing difference among the following magnitudes:

1. Fish production (FP) (variable "X"): The total fish production (X) of a fishing industry is the result of the operations of the fishing economic units (FEU's) within a given period of time.
2. Commercial fish production (CFP) (variable "Y"): The total commercial fish production (Y) is a part of the total fish production which enters the distribution channel within a given period of time. Further, CFP can be broken down as follows:

   $$Y = Y_1 + Y_2 + Y_3$$

   where:

   Y_1 : CFP which enters the distribution channel through the existing lakeside markets.

   Y_2 : CFP which enters the distribution channel through footpaths leading to the major road system and inland markets.

   Y_3 : CFP which is disposed locally without passing through the lakeside markets.

From the above analysis one can easily establish the following unequalities:

1. $X \geq Y$ or $D_1 = X - Y$
2. $Y \geq Y_1$ or $D_2 = Y - Y_1$

The secondary magnitudes D_1 and D_2 are different in nature. Specifically, D_1 expresses mainly (self-consumption)+(wages and services paid in fish)+(fish given to relatives)+(losses of fish). D_2 expresses the amount of the CFP which does not reach the existing lakeside markets. The size of D_2 is a function of the area distribution of the fishing industry, the existing lake transportation system, the location of the lakeside markets and inland markets, the existing road system and the purchasing power of the local consumers. For a family of African lakes it was estimated that the value of D_2 ranges between:

D_2/min. 0.15Y ———————— D_2/max. 0.60Y

13.3 Quantitative relationship between the variables "X" and "Y"

The following table shows the yearly fish production and the yearly commercial fish production of ten fishing sites at a given small lake (live weight). Prepare the "scatter diagram" illustrating the existing relationship between these two variables (example):

Serial number of fishing site	Fish production (m. tons) -x-	Commercial fish production (m. tons) -y-
1	120	96
2	150	120
3	80	64
4	200	160
5	120	96
6	60	48
7	40	32
8	80	64
9	250	200
10	300	240

13.3.1 <u>How to prepare a scatter diagram</u>

1. Draw two straight lines OX and OY at right angles (the lines are called axis).
2. Fish production is plotted along the X-axis while commercial fish production is plotted along the Y-axis.
3. Find the points which show the commercial fish production corresponding to the fish production given.

13.3.2 <u>The regression equation y=f(x)</u>

Judging from the chart of the above example we see that the relationship between the two variables is linear, and that the straight line appears to be a good fit to the empirical data. The established general equation can be given by:

$$y = a + bx$$

13.3.3 <u>Estimating the regression coefficient (a simple method)</u>

Two pairs of corresponding values of x and y can be obtained from the values in the above table. For example:

when, $x = 60 \quad y = 48$

$x = 80 \quad y = 64$

Substituting the above values in:

$$y = a + bx$$

$$48 = a + 60b \qquad (1)$$

$$64 = a + 80b \qquad (2)$$

Subtracting (2) from (1):

$$-16 = -20b$$

$$b = 0.8$$

$$a = 0$$

The equation is:

$$y = 0 + 0.8x$$

$$\text{or } y = 0.8x \qquad (3)$$

From equation (3) estimates of commercial fish production can be made for any desired level of fish production within the limits of the observations shown on the chart.

13.3.4 The meaning of the estimated regression coefficient "b"

In the above example the value of the regression coefficient "b" was found equal to 0.8. This means that out of one unit of fish production 0.8 goes to the commercial fish production.

13.4 Quantitative relationship between the variables "U" and "W"

Experimental studies at a lake "L" have proved that there is quantitative relationship between the following two variables:

Variable "U" : Number of existing fishing craft per fishing site (water approach)

Variable "W" : Number of fishing craft seen (aerial approach)

13.4.1 Formulation of the problem

During the year 1969 a number of statistical studies were conducted at lake "L" aiming to discover proper "control characteristics" which are of value at the design process of large scale sample surveys. In the surveys, among other things, the existing quantitative relationship between the above two variables, "U" and "W", was studied.

For survey operations the shoreline of the lake was divided into a number of zones of equal size and one of them was selected as "control zone" - within the selected zone two frame surveys (FS's) were conducted. FS_1 based on the "water approach" and FS_2 based on the "aerial approach". The table below gives the obtained results (example):

Lake "L", control zone, 1969

Serial no. of fishing sites found	No. of fishing craft seen1/ -w-	No. of existing fishing craft -u-
1	6	10
2	7	10
3	12	20
4	5	5
5	25	40
6	20	30
7	6	10
8	4	6
9	5	8
10	14	20
11	7	11
12	20	30
13	15	20
14	6	10
15	21	30
16	6	10
17	10	15
18	4	5
19	4	10
20	3	10

1/ No. of fishing craft seen on the beach plus on water

13.4.2 The scatter diagram

The scatter diagram based on the data in the above table shows that there is a quantitative relationship between the two variables "W", "U" and the straight line appears to be a good fit to the empirical data.

13.4.3 Linear regression equation u=f(v)

The established general equation is given by:

$$u = a + bv$$

Two pairs of corresponding values of "u" and "v" can be obtained from the values given in the table in section 13.4.1. For example:

when,

$v = 20 \quad u = 30$

$v = 10 \quad u = 15$

Substituting these in:

$$u = a + bv$$

$$30 = a + 20b \qquad (1)$$

$$15 = a + 10b \qquad (2)$$

Subtracting (2) from (1):

$$15 = 10b$$

$$b = 1.5$$

$$a = 0$$

The equation is:

$$u = 1.5v \qquad (3)$$

From the equation (3) estimates of the existing number of fishing craft per fishing site can be made for any number of fishing craft seen within the limits of observations shown on the chart.

13.4.4 The meaning of the estimated regression coefficient "b"

In the above example the value of the regression coefficient "b" was found equal to 1.5. This means that one unit of "fishing craft seen" corresponds to 1.5 of "existing fishing craft".

PART III: TECHNIQUES OF SAMPLING (ADVANCED COURSE)

14. BASIC IDEAS OF SAMPLING

14.1 Accuracy and precision

One of the stages in the "survey system" of a sample survey is the selection of the sample of the survey. As a result of a sample survey estimates are calculated on the characteristics of the surveyed population. An estimate calculated from the sample is said to be precise if it is near the "expected value", that is, census count taken under identical experimental conditions. For example, if $\bar{y}$ expresses the average sample value of a given characteristic and m the corresponding expected value, the absolute precision of the estimate is given by:

$$|d| = |\bar{y}-m|$$

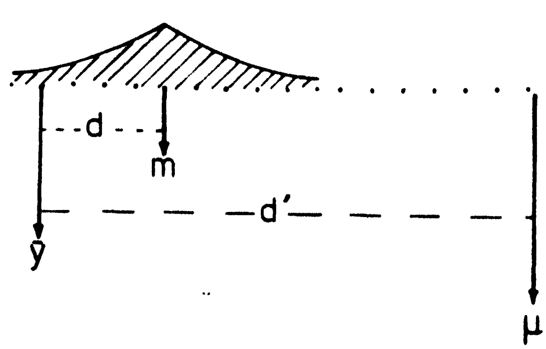
The calculated estimate may not necessarily be near the true value aimed at; that is, it need not be accurate. In our notation, if μ is the true value, the absolute accuracy of the estimate is given by:

$$|d'| = |\bar{y}-\mu|$$

From the above discussion it is obvious that precision refers to closeness of a sample estimate to the expected value while accuracy refers to the closeness to the true value. Deviation between the expected value and true value occurs when the errors present in the data do not average to zero.

It should be noted that the expected value can be obtained from repeated applications of the given sampling procedure[1/]. In the figure below (Figure 1) there is a graphical explanation of the meaning of precision and accuracy of the sample estimate $\bar{y}$.

Figure 1. Sampling distribution of the estimates



1/ The expected value is the mean of the frequency distribution of the estimates of all possible samples derived from repeated applications of the given survey method. The frequency distribution of the estimates is also called the sampling distribution of the estimates.

From the above discussion it is obvious that the level of precision of a sample estimate depends on the spread of the sampling distribution, and this is conveniently measured by its own standard deviation, i.e. the standard error of the estimate. If the sample statistic in question is the total value of a characteristic, then the relevant distribution is the "sampling distribution of the total", the relevant standard error the "standard error of the total"; if it is a proportion, the terms are "sampling distribution of proportion" and "standard error of proportion".

14.2 Sources of errors in sample surveys

It is important from a theoretical as well as a practical viewpoint to classify the various kinds of errors that arise in the survey system of a sample survey into the following two categories:

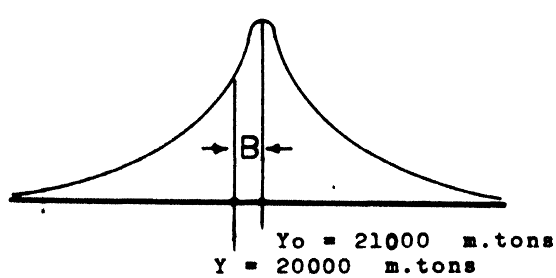
1. Sampling errors.
2. Non-sampling errors.

Specifically, in the sampling theory, the idea of sampling errors are introduced under the assumption that any measurement of the j^{th} unit is correct for that unit (errors of measurement are ignored). Further, there is no dearth of examples to show that errors of measurement or observation, or errors of response, are present when a survey is carried out (or a census is taken). In the domain of non-sampling errors are included response errors, coverage errors, processing errors, etc. In the following sections the meaning of sampling errors (sections 2a-2d) and non-sampling errors (sections 2e-2g) is discussed.

14.2a Bias of estimation

Text books and manuals of mathematical statistics discuss the bias which is attributed to the estimator of a sample survey, i.e. the difference between the expected value and the true value. The ratio estimate1/ is a good example of biased estimate. In the ratio method the average (or expected value) of all the sample estimates is not necessarily equal to the parameter under estimation. For example, if the true quantity of total fish catch in a given year at lake A is Y = 20000 metric tons and the relevant expected value based on ratio estimator is Yo = 21000 metric tons, the ratio estimate is subject to a bias of a level B = Yo-Y = 21000 metric tons - 20000 metric tons = 1000 metric tons. In the figure below (Figure 2) there is a graphical presentation of the meaning of bias of ratio estimate based on large samples:

Figure 2.



14.2b Selection by means of incomplete sampling frames

Area sampling on the basis of maps is often used in the fisheries statistical surveys at inland waters in Africa. This procedure can give an unbiased sample if the sampling frames used are complete. The same situation exists when sampling units are selected on the basis of different lists and registers of the units. The use of incomplete lists often gives a strongly biased sample.

1/ See section 15.5

14.2c Non-response

Like an incomplete sampling frame, there is a danger of bias in data selection in which, from units chosen in the sample, it is not possible to get information on the surveyed characteristics. It is a risk of bias to base the results of the survey on the respondents alone, since the non-respondents may be different from the respondents. Sampling theory points the way to methods for dealing with non-response.

14.2d Other sources of sampling bias

(i) In systematic sampling[1/] if the population consists of a periodic trend, e.g. a simple sine curve, the effectiveness of the sample depends on the value of the spacing interval. Specifically, systematic selection can lead to a bias if the spacing interval is an even multiple of the half-period.

(ii) In the case of current sample surveys there is a risk of bias if the sampling frames used are not kept up-to-date.

(iii) The use of purposive samples or quota samples might lead to bias results.

14.2e Response errors

In the previous sections it has been indicated that, in getting at the optimum sampling design, it has been assumed that the job is to estimate from a sample what would have been obtained from a complete census conducted under identical conditions. This statement of the problem avoids dealing with response errors, i.e. observation errors that are present in a census taken with equal care. The real problem in this regard is the measurement of response errors. Specifically, three points need consideration: (i) How response errors arise; (ii) Detection of response errors; and (iii) Estimation of response variance.

(i) How response errors arise Suppose a fisherman is asked about the total quantity of fish caught in a given day and that the response obtained is x_{jt}[2/]. By using the "real measurement approach" the true quantity of fish caught is given by y_j. From a sampling point of view, the response x_{jt} obtained is a random variable[3/] with mean $\bar{X}_j$ (average individual response) and variance σ_j^2. If there are N fishermen at the given inland water place, we have the values:

$$\bar{X}_{j=1},\ \bar{X}_{j=2},\ \ldots\ ,\ \bar{X}_{j=N}$$

which are the average individual responses to the question "quantity of fish caught in a given day". The average of $\bar{X}_j$ i.e.:

$$\bar{X} = \frac{1}{N}\sum_{j=1}^{N}\bar{X}_j$$

is called the "expected survey value" obtained under given experimental conditions of the survey. Further, the average of the true values y_j is given by:

$$\bar{Y} = \frac{1}{N}\sum_{j=1}^{N}y_j$$

By using our notation, the difference:

$$B' = \bar{X}-\bar{Y}$$

1/ See section 15.4

2/ Suffix t stands for the trials and suffix j for the units in the population.

3/ The difference $x_{jt}-\bar{X}_j$ is called the "individual response deviation" and the difference $\bar{X}_j-y_j$ is called the "individual response bias".

is called the "bias of the survey" when the purpose is to estimate $\bar{Y}$, the average quantity of fish caught per fisherman/given day.

(ii) Detection of response errors Past experience has proved that the magnitude and direction of the difference between the true value and the value obtained from the survey depends upon the method used in the survey, the design of the source documents and the human elements, i.e. the recorder and the respondent.

In the field of large scale statistical surveys detection of response errors is obtained through the Quality Check Surveys (QCS's). The QCS's are intensive studies of relatively small samples and every effort is made in them to attain the highest level of efficiency possible. Through the QCS's the magnitude and direction of the "gross errors" are assessed i.e. the algebrical differences which are calculated by comparing the individual answers between the main survey and QCS, and the magnitude of "net error" is estimated i.e. the overall error remaining after any cancellation of gross errors has taken place. It should be noted that many of the individual response deviations may be very large. But this does not mean that the response bias of the survey will necessarily be large. Some of the individual errors may be positive, some may be negative, and the response bias may be small.

(iii) Estimation of response variance Response variance, like sampling variance, can be estimated from the obtained sample data. This can be illustrated with the following example: At Lake Volta (Ghana) a large fishing site harboured N=315 fishing economic units. To get an estimate of the level of response variance of average catches per unit the following survey system was employed:

1. A sample of n=30 fishing economic units was selected on to-fishing day and fish catch information was taken (y) by using the "real measurement approach".
2. On day to+1 a sample of n'=100 fishing economic units was selected (in the n' units the sample of n units was included) and information was collected of fish catch on to-fishing day (x) by using the "interview approach".

For the above example the estimators of total variance, sampling variance and response variance are given in the table below:

Source of variance	Estimators
1. Response variance	$s_r^2 = (\frac{1}{n}-\frac{1}{n'})s_d^2$, where, $s_d^2 = \frac{1}{n-1}\left[\sum_i^n d_i^2 - \frac{(\sum_i^n di)^2}{n}\right]$ and, di = yi-xi
2. Sampling variance	$s_s^2 = (\frac{1}{n'}-\frac{1}{N})s_y^2$, where, $s_y^2 = \frac{1}{n-1}\left[\sum_i^n y_i^2 - \frac{(\sum yi)^2}{n}\right]$
3. Total variance	$s_T^2 = s_r^2+s_s^2 = (\frac{1}{n}-\frac{1}{n'})s_d^2+(\frac{1}{n'}-\frac{1}{N})s_y^2$

It should be noted that in the field of large scale surveys control of response errors is achieved through a proper selection, training and supervision of the field personnel of the surveys, consistency checks of the completed questionnaires, re-interviews, etc.

14.2f Coverage, content errors

Frame Survey (FS) taking in inland waters is a big operation requiring the use of enumerators and other persons at various levels, transport boats, cars, etc. Due to the sheer size of the several operations involved and the peculiarities of the surveyed population errors of different types are likely to creep in. In FS's there are two kinds of errors to be checked:

1. Coverage errors i.e. 1) omissions or erroneous inclusions of fishing sites, and 2) omissions or erroneous inclusions of fishing economic units within the fishing sites covered by the Frame Survey.
2. Content errors i.e. errors on the reports of the number of gear owned by the fishermen, number of gear fished, etc. (see section 14.2e).

An estimate of the sources of coverage errors and their magnitudes are calculated through the Coverage Check Surveys1/.

14.2g Other sources of non-sampling bias

1. Editing errors: With the field part of the survey completed, the various stages of processing the material and the highly skilled task of analyzing it begin. It cannot be taken for granted that the data coming from the field are free of all errors. Incompleteness, inconsistency and inaccuracy in the completed questionnaires will strongly affect the reliability of the obtained results if they are not checked at the editing stage of the survey system of the survey.
2. Coding errors: Another source of bias is the errors arising in the codes of the surveyed characteristics. To avoid coding errors the coders must be practised on a sample of data and the problems that arise are discussed to bring about uniformity and consistency in the procedure.
3. Recording and arithmetical errors: Arising at the stage when estimates of the surveyed magnitude are calculated and in the tabulation process, these errors strongly affect the level of reliability of the obtained results.

14.3 Mean Square Error (MSE)

The users of statistics are always interested in the question of the level of reliability of the results obtained from a given sample survey. The index of the joint effect of bias and of random error is the Mean Square Error. Specifically, in the sampling theory estimators of MSE have been produced for the following two cases:

1. MSE of statistics when response errors are ignored.
2. MSE of statistics when response errors are taken into account.

To define the MSE in the above case (1) we use the notation already established in the previous sections. For a given statistic z in the sample (i.e. total, mean, proportion, etc.), let the expected value be $m(m = E(z)$, where the operator E stands for the expected value), the standard error of the estimate is σ_z and the variance of the statistic is:

$$V(z) = \sigma_z^2$$

1/ See: 1) Section 5.3, 2) Frame Surveys at Volta Lake, St.S./2, FIO:SF/GHA/10, March 1970 by G.P. Bazigos.

If μ is the true value of the parameter under estimation, the bias is given by:

$$B = m-\mu$$

The formula for the MSE is given by:

$$\begin{aligned} MSE(z) &= E(z-\mu)^2 \\ &= E\{(z-m)+(m-\mu)\}^2 \\ &= V(z)+B^2 \\ &= V(z)\{1+(\frac{B}{\sigma_z})^2\} \end{aligned}$$

From the above analysis it is obvious that if z is unbiased the variance and the Mean Square Error of the statistic would coincide. Of two estimators z_1 and z_2 the one giving the smaller Mean Square Error around the parameter to be estimated will be preferred.

In the case (2) the Mean Square Error of the statistic $\bar{x} = \frac{1}{n}Sx_{jt}$ (see section 14.2e) is given by:

$$E(\bar{x}-\bar{Y})^2 = A+B+C+D$$

Where:

$$A = V\left[\frac{1}{n}\sum(x_{jt}-\bar{x}_j)\right]$$, is the response variance

$$B = V\left[\frac{1}{n}\sum(\bar{X}_j-\bar{X})\right]$$, is the sampling variance

$$C = 2Cov.\left[\frac{1}{n}\sum(x_{jt}-\bar{X}_j),\frac{1}{n}\sum(\bar{X}_j-\bar{X})\right]$$, reflects the correlations between response and sampling deviations, and

$D = (\bar{X}-\bar{Y})^2$, is the square of the response bias.

14.4 <u>The application of confidence intervals to detect the bias (errors of measurement are ignored)</u>

It has been pointed out that the bias caused by the sampling can be detected through the examination of the organic structure of the sample. It is sometimes possible to study the final effect of the bias by comparing the sample estimates with the results of a census survey, provided that such information is available, or with relevant supplementary information.

Let us use the following notation: z is the sample statistic of n elements, the standard error of statistic is σ_z and the value of the population parameter is Z. The t-fold value of the standard error is the margin of error:

$$\delta = t\sigma_z$$

The probability that the interval $z\pm\delta$ covers the population parameter is:

$$P(z-\delta<Z<z+\delta)$$

In the case of bias the confidence interval does not cover the population parameter at the given probability level. An indication of the magnitude of the bias of the statistic can be obtained by comparing the value of Z with the established limits (lower, upper) of the confidence interval.

14.5 Methods of de-biasing

In many cases fishery statisticians are asked to assess the efficiency of the sampling procedures of various sample surveys undertaken with the absence of an experienced designer. In most of the cases the selected samples are biased. How can the results of the sample surveys be improved by eliminating or at least reducing the bias? One solution to the problem is to design a new survey based on healthy statistical principles. This solution is, however, very costly and hardly provides a practical solution to the problem. It is therefore advisable not to change the structure of the sample and to try to get better results on the basis of the given biased sample. The sample bias can be diminished by using certain methods of de-biasing:

1. Bias of estimation can be eliminated by choosing the proper estimator or increasing the size of the sample or by using the method of stratification after selection and weighting with the real proportions of strata.

2. Bias by means of incomplete sampling frames can be eliminated by introducing correcting factors in the estimation procedure of the survey.

3. In the case of non-response, we select a small sample from the sample of non-respondents and the sub-sample data are used in order to revise the sample estimates. From a sampling point of view, the survey population composes of two sub-populations or strata: respondents (N_1) and non-respondents (N_2), ($N=N_1+N_2$). Let n_1 be the number of sample respondents, while n_2 units failed to respond. We take a sub-sample of size $r_2 = \frac{1}{k}n_2$ from non-respondents and collect information from these units. An estimate of the population mean is given by:

$$\hat{\bar{y}} = \frac{1}{n}(n_1\bar{y}_1+n_2\bar{y}_2)$$

where,

$$\bar{y}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} y_{1j}, \bar{y}_2 = \frac{1}{r_2}\sum_{j=1}^{r_2} y_{2j}$$

14.6 Costs of Fisheries Statistical Surveys

With a large and increasing number of sample designs thrown up by the development of sampling theory, the choice of near optimal design is far from easy. From a sampling point of view, that design is to be chosen which yields a required quantity of information at a specified accuracy (precision) at minimum cost; or alternatively the design providing the required information at specified cost with maximum accuracy (precision). However, other factors should also be taken into account at the process of designing a sample survey. In the field of fisheries statistics the time which passes from the design of the surveys until the publication of the results strongly influences the utility of the obtained results. In the figure below (Figure 3) there is a graphical representation of the relationship between the time required for obtaining the results and the utility of the results:

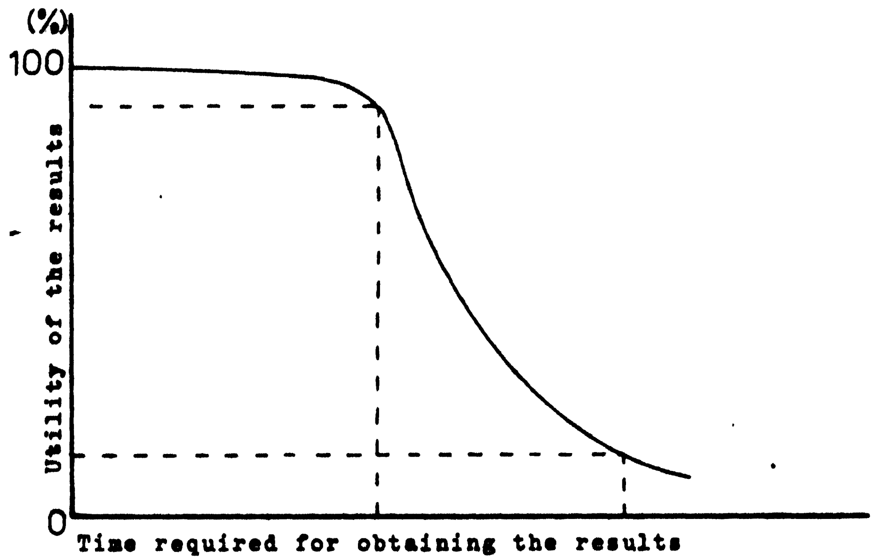


Figure 3.

Among the various arguments in favour of using sampling procedures is that the sampling method can often supply the required information with greater speed and at lower cost than a complete enumeration. For this reason, information on the costs involved in sample surveys is of particular value for the development of the sample surveys within the countries and is also of help to other member countries. Since every operation means cost, an attempt is made to use simple, straightforward procedures, procedures which can be completed within the time schedules, and which take into account all administrative requirements. In the figure below (Figure 4) there is a graphical representation of the various phases of the survey system of a current Catch Assessment Survey (CAS) at inland waters.

From a sampling point of view, two types of cost computation must be distinguished:

1. Pre-estimated cost i.e. calculation of the costs at the process of designing the survey.
2. Actual cost i.e. calculation of the costs after the completion of the survey.

Further, for optimum planning it is mostly necessary to determine the "cost-function", i.e. the function which indicates the quantitative relationship between cost and the sample size of the survey.

For a better presentation of the cost items of a CAS[1/], a two-way table is used. In the following table the first criterion of classification refers to the various phases of the survey system of a CAS and the second criterion refers to the type of the cost items, i.e. one-time cost, current fixed cost, current variable cost.

1/ CAS: Catch Assessment Survey.

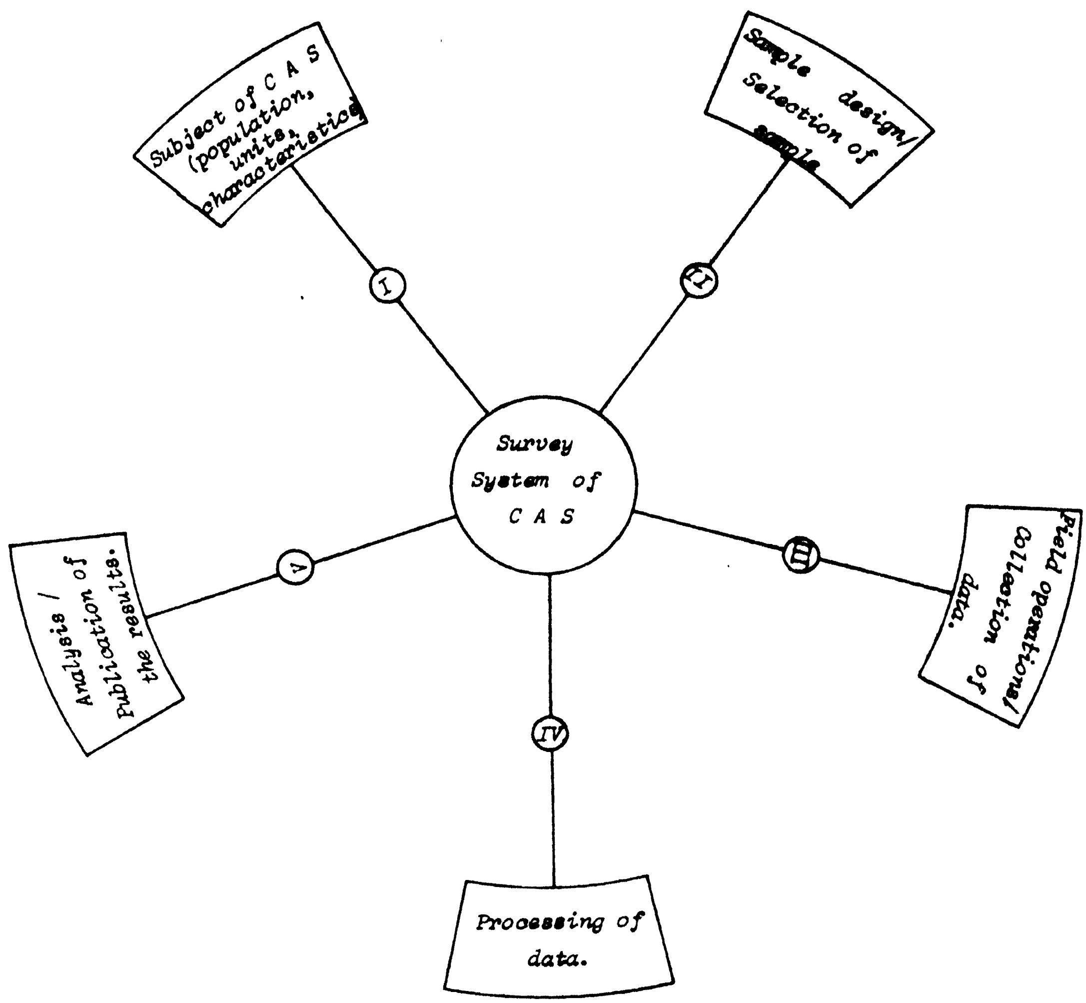

Subject of C A S (population, units, characteristics)
Sample design/ Selection of sample
Survey System of C A S
I
II
III
IV
V
Field operations/ Collection of data.
Processing of data.
Analysis / Publication of the results.

Figure 4.

The cost items of a current CAS

Phases of the survey system - CAS	One-time Costs	Current Costs Fixed Costs	Current Costs Variable Costs	Remarks
I. Subject of CAS				
1. Tentative analysis of available material	x			
2. Reconnaissance Surveys (RS's)	x			
3. Analysis of the results of RS's	x			
4. Definition of various sampling units and population of characteristics	x			
5. Preparation of codification system	x			
6. Selection of field personnel (FP)	x			
7. Opening/keeping the files of FP		x		
II. Sample design				
1. Assessment of the quality of existing Sampling Frames	x			
2. Coverage Check Surveys (CCS's) 1/	x			
3. Analysis of the results of CCS's	x			
4. Selection of the sample of the survey	x			
5. Preparation of the source documents of CAS	x			
6. Training of the field personnel	x			
7. Mini Pilot Survey	x			
8. Main Pilot Survey	x			
III. Field operations				
1. Preparation/allocation of mapping material to FP		x		
2. Preparation/allocation of source documents to FP		x		
3. Identification/Replacement of selected fishing sites			x	
4. Field work		x		
5. Field supervision		x		
6. Movements of supporting unit		x		
IV. Processing				
1. Checking the completed forms and further inquiries		x		
2. Coding and checking		x		
3. Estimation of control characteristics (manual calculations)		x		Using working sheets
4. Tabulation and adjusting (manual)		x		
5. Checks of tables		x		
6. Mechanical Tabulation		x		
V. Analysis/Publications				
1. Arithmetical analysis		x		
2. Graphic evaluation		x		
3. Preparation of reports		x		
4. Publication of results		x		
5. Answers to special questions of users			x	

1/ Coverage Check Surveys might be conducted at this stage.

14.7 Integration of sample surveys

Modern sample surveys are becoming multipurpose in character in the sense that information is collected on hundreds of items belonging to different fields of enquiry and that the results must be made available before they become out of date. In the research programme at inland waters in Africa, the demand for data is normally satisfied through a series of different surveys, i.e. Fishery Statistical Surveys, Biological Surveys, Limnological Surveys, etc. While the various research fields are covered by specific research programmes it is very important to integrate the survey system of the various surveys and the resources available. Integration implies greater efficiency and lower cost for the surveys covering the population under investigation. The main operational steps in which integration might be achieved are:

1. Integration at the sampling stage.
2. Integration of the field operations.
3. Integration of the pre-mechanical processing of data.
4. Integration of the mechanical processing of data.
5. Integration of the organizational aspects of the surveys.

15. TYPES OF SAMPLE DESIGN

15.1 Introduction

The scope of this chapter is to present in compact form a collection of the types of sample design used in fisheries statistics (inland waters) at the present time with indications involved in their application. It is hoped that the sampling methods presented in this chapter with illustrations will make sampling more efficient in the hands of the respective workers of this field.

15.2 Simple Random Sampling (SRS)

Assuming that there are N fish (species A) in a fish pond. If these units can all be distinguished from one another they can be denoted by:

Fish: 1st, 2nd, 3rd, ... , j , ... , N^{th}

or, Unit: $a_1, a_2, a_3, \ldots, a_j, \ldots, a_N$, $(j = 1, 2, \ldots, N)$

Now the following question needs consideration: If N = 4 how many distinct samples[1/] of size n = 2 can be drawn from the N units? The number of distinct samples is given by the combinatorial formula:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{4!}{2!2!} = 6$$

The six distinct samples of size n = 2 are given in the table below:

Sample	Sample units
s_1	a_1, a_2
s_2	a_1, a_3
s_3	a_1, a_4
s_4	a_2, a_3
s_5	a_2, a_4
s_6	a_3, a_4

1/ The same fish is not allowed to occur twice in the sample (sampling without replacement, wrp).

If the procedure of selection is such that each sample has an equal probability to be selected, the type of sampling is called Simple Random Sampling.

In practice, it is impossible to produce each time all the distinct samples of size n and select one of them. Usually a simple random sample is drawn unit by unit. The units in the population are numbered from 1 to N. A series of random numbers between 1 and N is then drawn by means of a table of random numbers[1/]. The units in the population which bear these numbers constitute the sample. It has been proved that this method produces simple random samples.

15.2.1 Estimation of the population mean and total (SRS)

Assuming that for each unit a_j in the population is attached a variate value y_j for the characteristic (y). By using our notation the following magnitudes can be defined:

a: Population total: $Y = \sum_{j=1}^{N} y_j = y_1+y_2+y_3+ \dots +y_N$

b: Population mean: $\bar{Y} = \frac{Y}{N} = \frac{1}{N}(y_1+y_2+y_3+ \dots +y_N)$

c: Sample total: $y = \sum_{j=1}^{n} y_j = (y_1+y_2+y_3+ \dots +y_n)$

d: Sample mean: $\bar{y} = \frac{y}{n} = \frac{1}{n}(y_1+y_2+ \dots +y_n)$

e: Estimated population total: $\hat{Y} = N\bar{y}$

It has been proved that in simple random sample the sample mean $\bar{y}$ is an unbiased estimate of the population mean $\bar{Y}$.

Example 1: A fish pond contains N = 1200 fish. A simple random sample of n = 40 fish was selected and their total weight was obtained (y = 2000 gr):

1. What is the total weight of fish in the population (estimate)?
2. What is the estimated average weight per fish?

Solution:

1. $\hat{Y} = N\bar{y} = \frac{N}{n}y = \frac{1200}{40}x2000gr. = 60kg.$
2. $\bar{y} = \frac{y}{n} = \frac{2000gr}{40} = 50gr.$

15.2.2 Sampling error of $\bar{y}$

It has been proved that in simple random sampling (wrp) the variance of the sample mean is given by[2/]:

$$(1) \quad V(\bar{y}) = \frac{1}{n}(1-\frac{n}{N})S_y^2 = (\frac{1}{n}-\frac{1}{N})S_y^2 = (\frac{N-n}{nN})S_y^2$$

where S_y^2 is the variance per unit in the population,

1/ The same number is not allowed to enter the sample more than once (sampling without replacement). Table of random numbers is given in Appendix IIIa.

2/ In (1) the factor $(1-\frac{n}{N}) = \frac{N-n}{N}$ may be called the finite population correction (fpc). If $\frac{n}{N}$ is small, this factor is close to unity.

$$S_y^2 = \frac{1}{N-1}\sum_{j=1}^{N}(y_j-\bar{Y})^2 = \frac{1}{N-1}\left(\sum_{j=1}^{N}y_j^2-\frac{\left(\sum_{j=1}^{N}y_j\right)^2}{N}\right)$$

From the above formula (1) it is obvious that the variance of the sample mean depends upon the population variance and decreases as the sample size increases.

The standard error of $\bar{y}$ equals the square root (positive) of the variance of $\bar{y}$:

(2) $S_{\bar{y}} = \sqrt{V(\bar{y})} = S_y\sqrt{\frac{N-n}{nN}}$

The meaning of $S_{\bar{y}}$: The standard error of $\bar{y}$ shows the degree of concentration of the sample means around the population mean (errors of measurement are ignored). If the value of $S_{\bar{y}}$ is small it implies that the probability of a large deviation from the population mean is small.

For n>30, the statistic $\bar{y}$ follows the normal distribution $N(\bar{Y}, S_{\bar{y}})$. In such a case there is a probability of 95 percent that the sample mean falls within the interval:

(3) $\bar{Y}-1.96S_{\bar{y}}<\bar{y}<\bar{Y}+1.96S_{\bar{y}}$

Further, from the above inequality the confidence interval for $\bar{Y}$ can be established (P = 95 percent, large samples):

(4) $\bar{y}-1.96S_{\bar{y}}<\bar{Y}<\bar{y}+1.96S_{\bar{y}}$

The standard error of the sample mean can be expressed as a fraction or percentage of the population mean. This magnitude is called the coefficient of variation of the sample mean, $CV(\bar{y})$. It expresses the relative precision of the statistic $\bar{y}$:

(5) $CV(\bar{y}) = \frac{S_{\bar{y}}}{\bar{Y}}$

$$= \frac{S_y}{\bar{Y}}\sqrt{\frac{N-n}{nN}} = CV(y)\sqrt{\frac{N-n}{nN}}$$

Where:

$CV(y) = \frac{S_y}{\bar{Y}}$: coefficient of variation in the population.

In practice S_y^2 is hardly known. An unbiased estimate of S_y^2 can be obtained by using the data of the selected sample:

(6) $\hat{S}_y^2 = s_y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j-\bar{y})^2 = \frac{1}{n-1}\left(\sum_{j=1}^{n}y_j^2-\frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}\right)$

An unbiased estimate of the variance of the sample mean is given by:

(7) $v(\bar{y}) = \frac{1}{n}(1-\frac{n}{N})s_y^2$

$$= (\frac{1}{n}-\frac{1}{N})s_y^2 = \frac{N-n}{nN}s_y^2$$

The estimated standard error of $\bar{y}$ is given by:

$$(8)\quad s_{\bar{y}} = \sqrt{v(\bar{y})} = s_y\sqrt{\frac{N-n}{nN}}$$

The estimated coefficient of variation of $\bar{y}$ is given by:

$$(9)\quad cv(y) = \frac{s_{\bar{y}}}{\bar{y}} = \frac{s_y}{\bar{y}}\sqrt{\frac{N-n}{nN}} = cv(y)\sqrt{\frac{N-n}{nN}}$$

where, $cv(y) = \frac{s_y}{\bar{y}}$ estimated coefficient of variation in the population.

Also, the estimated confidence interval of $\bar{Y}$ is given by (P = 95 percent, large samples):

$$(10)\quad \bar{y}-1.96s_{\bar{y}}<\bar{Y}<\bar{y}+1.96s_{\bar{y}}$$

Example 2: A Frame Survey (FS) was conducted at Lake A. The table below (Table 15.2.2.1) gives the total number of available gill nets per fishing site (y) covered by the survey (N = 72 fishing sites covered by the FS). Operations needed:

1. Select a sample of n = 25 fishing sites (SRS, wrp).
2. Use the sample data and calculate the magnitudes s_y^2, s_y, $cv(y)$, $v(\bar{y})$, $s_{\bar{y}}$, $cv(\bar{y})$.
3. Estimate the confidence interval of $\bar{Y}$.
4. Calculate the population mean by using the data of the Frame Survey.
5. Calculate the absolute level of precision of the sample mean.

Table 15.2.2.1 Total number of gill nets available per fishing site (y), FS-Lake A

Ser.No. of fishing site	Gill nets available	Ser.No. of fishing site	Gill nets available	Ser.No. of fishing site	Gill nets available	Ser.No. of Fishing site	Gill nets available	Ser.No. of fishing site	Gill nets available
01	5861	16	129	31	378	46	218	61	129
02	1860	17	129	32	368	47	200	62	126
03	722	18	124	33	332	48	185	63	126
04	470	19	121	34	323	49	180	64	123
05	397	20	112	35	297	50	177	65	122
06	357	21	102	36	295	51	177	66	119
07	338	22	2170	37	274	52	176	67	112
08	314	23	203	38	257	53	170	68	112
09	295	24	148	39	255	54	166	69	111
10	221	25	108	40	246	55	166	70	110
11	221	26	790	41	244	56	152	71	109
12	176	27	651	42	241	57	151	72	108
13	162	28	511	43	238	58	149		$\sum_{j=1}^{72} y_j = 25767$
14	151	29	421	44	236	59	139		
15	150	30	410	45	224	60	132		

Solution:

1. In the table below (Table 15.2.2.2) the selected sample is given.

Table 15.2.2.2 Selected sample of fishing sites

Sample Ser.No.	Gill nets available	Sample Ser.No.	Gill nets available	Sample Ser.No.	Gill nets available	Remarks
01	1860	11	790	21	152	
02	397	12	651	22	139	(SRS, wrp)
03	338	13	421	23	123	
04	221	14	378	24	119	
05	176	15	332	25	112	
06	151	16	297		y= 8132	
07	129	17	274			
08	124	18	257		$\sum_{j=1}^{25} y_i^2 = 5782660$	
09	112	19	200			
10	203	20	176		$\bar{y} = 325.28$	

2. The estimated magnitudes are:

2.1 $s_y^2 = \frac{1}{24}(5782660-\frac{(8132)^2}{25}) = 97392$ gill $nets^2$

2.2 $s_y = \sqrt{97392 \text{ gill nets}^2} = 312.08$ gill nets

2.3 $cv(y) = \frac{312.08}{325.28}x100 = 95.94$ percent

2.4 $v(\bar{y}) = \frac{72-25}{25x72}x97392 = 2543$ gill $nets^2$

2.5 $s_{\bar{y}} = \sqrt{2543 \text{ gill nets}^2} = 50.43$ gill nets

2.6 $cv(\bar{y}) = \frac{50.43}{325.28}x100 = 15.50$ percent

3. The estimated confidence interval of $\bar{Y}$ (P = 95 percent, t-student):

$325.28-2.064x50.43<\bar{Y}<325.28+2.064x50.43$

221.19 gill nets $<\bar{Y}<$ 429.37 gill nets

4. The calculated population mean is:

$$\bar{Y} = \frac{1}{N}\sum_{j=1}^{72} y_j = \frac{25767}{72} = 357.87 \text{ gill nets}$$

5. The calculated absolute level of precision of the sample mean is:

$|d_1| = |325.28-357.87| = 32.59$ gill nets

15.2.3 Sampling error of $\hat{Y}$

In section 15.2.1 it is indicated that the estimated total of a survey characteristic is given by:

$$\hat{Y} = N\bar{y}$$

Therefore the variance of $\hat{Y}$ is:

(11) $V(\hat{Y}) = N^2V(\bar{y}) = N^2\frac{(N-n)}{nN}S_y^2 = \frac{N(N-n)}{n}S_y^2$

The standard error of $\hat{Y}$ is:

(12) $S_{\hat{Y}} = \sqrt{V(\hat{Y})} = S_y\sqrt{\frac{N(N-n)}{n}}$

The coefficient of variation of $\hat{Y}$ is:

(13) $CV(\hat{Y}) = \frac{S_{\hat{Y}}}{Y} = \frac{NS_{\bar{Y}}}{N\bar{Y}} = \frac{S_{\bar{Y}}}{\bar{Y}} = CV(\bar{y})$

The confidence interval of Y is given by (P = 95 percent, large samples):

(14) $\hat{Y}-1.96S_{\hat{Y}}<Y<\hat{Y}+1.96S_{\hat{Y}}$

Unbiased estimates of the above magnitudes (11-13):

The estimated variance of $\hat{Y}$ is:

(15) $v(\hat{Y}) = N^2v(\bar{y}) = N^2\frac{(N-n)}{nN}s_y^2 = \frac{N(N-n)}{n}s_y^2$

The estimated standard error of $\hat{Y}$ is:

(16) $s_{\hat{Y}} = \sqrt{v(\hat{Y})} = s_y\sqrt{\frac{N(N-n)}{n}}$

The estimated coefficient of variation of $\hat{Y}$ is:

(17) $cv(\hat{Y}) = \frac{s_{\hat{Y}}}{\hat{Y}} = \frac{Ns_{\bar{y}}}{N\bar{y}} = \frac{s_{\bar{y}}}{\bar{y}} = cv(\bar{y})$

The estimated confidence interval of Y is (P = 95 percent, large samples):

$\hat{Y}-1.96s_{\hat{Y}}<Y<\hat{Y}+1.96s_{\hat{Y}}$

<u>Example 3</u>: By using the data of Table 15.2.2.2 calculate $v(\hat{Y})$, $s_{\hat{Y}}$, $cv(\hat{Y})$. Estimate the confidence interval of Y. Calculate the absolute level of precision of $\hat{Y}$ (Y = 25767).

Solution:

Estimated magnitudes:

1. Estimated variance of $\hat{Y}$,

   $v(\hat{Y}) = \frac{N(N-n)}{n}s_y^2 = \frac{72(72-25)}{25}x97392 = 13182981.12$ gill $nets^2$

2. Estimated standard error of $\hat{Y}$.

   $s_{\hat{Y}} = \sqrt{13182981.12 \text{ gill nets}^2} = 3630.84$ gill nets

3. Estimated coefficient of variation of $\hat{Y}$,

   $cv(\hat{Y}) = \frac{3630.84}{23420.16}x100 = 15.50$ percent
   where,
   $\hat{Y} = 72x325.28 = 23420.16$ gill nets

4. Estimated confidence interval of Y (P = 95 percent, t-student),

$23420.16-2.064 \times 3630.84 < Y < 23420.16+2.064 \times 3630.84$

15926.11 gill nets $< Y < 30914.21$ gill nets

5. Estimated absolute precision of $\bar{Y}$,

$|d_2| = |23420.16-25767| = 2346.84$ gill nets

15.2.4 Sample size

An important problem arising in a sample survey is the determination of the size of the sample. The following are the principal steps involved in the choice of the sample size:

1. We must know what is expected to be achieved through the survey.
2. Some equation must be found connecting the sample size and the requirements specified in (1).
3. This equation will contain certain unknown quantities belonging to the population. These quantities must be estimated.

From the above formula (5) it is obvious that there is a relationship between the size of the sample and the precision of $\bar{y}$. Provided that CV(y) is known (data of a previous survey) the required sample size for a given precision of $\bar{y}$ would be:

$$CV(\bar{y}) = CV(y)\sqrt{\frac{1}{n}-\frac{1}{N}}$$

or, $$\frac{CV(\bar{y})}{CV(y)} = \sqrt{\frac{1}{n}-\frac{1}{N}}$$

if, $g = \frac{CV(y)}{CV(\bar{y})}$, then

$$\frac{1}{g} = \sqrt{\frac{1}{n}-\frac{1}{N}}, \text{ and}$$

$$(18) \quad n = \frac{Ng^2}{N+g^2}$$

Example 4: A fish pond contains N = 1000 fish (species A). A survey is planned to estimate the average weight per fish. How many fish should be selected to achieve a $CV(\bar{y}) = 0.05$ (Note: From a previous survey it was found that cv(y) = 30 percent).

$N = 1000$ fish, $\quad g = \frac{0.30}{0.05} = 6$

and,

$$n = \frac{1000 \times 6^2}{1000+6^2} = 34.7 \approx 35 \text{ fish}$$

15.2.5 Estimation of proportions

We shall now consider the problem of estimating the proportion of a population belonging to a certain class A. Assuming that in a given fish pond there are two kinds of fish, species A and species B. We would like to estimate the proportion of species A in the population, through a simple random sample selected from the fish pond. If for every fish in the pond we define the variate y_j as 1 if the fish belongs to class A (species A) and 0 otherwise it is easy to see that the total number of fish belonging to class A in the pond is:

(19) $N_A = \sum_{j=1}^{N} y_j$

The proportion (P) of fish belonging to class A is:

(20) $P = \frac{1}{N}\sum_{j=1}^{N} y_j = \frac{N_A}{N}$, ($N = N_A+N_B$, total number of fish in the fish pond).

From the above discussion it is obvious that estimating P is equivalent to estimating a population mean, where the mean is defined in terms of the nev variate y taking up values 1 and 0.

As a consequence, the following results are readily obtained. If a simple random sample (wrp) of n fish is selected from the fish pond and n_A and n_B are the number of fish in the sample belonging to species A and B respectively ($n = n_A+n_B$), an unbiased estimate of P is given by:

(21) $p = \hat{P} = \frac{n_A}{n}$, $(q = \frac{n_B}{n} = 1-p)$

The variance of p is:

$$V(p) = \frac{1}{n}(1-\frac{n}{N})S_y^2$$

where:

$$S_y^2 = \frac{1}{N-1}\left(\sum_{j=1}^{N} y_j^2 - \frac{\left(\sum_{j=1}^{N} y_j\right)^2}{N}\right) = \frac{1}{N-1}(NP-NP^2) = \frac{NPQ}{N-1}$$

and:

(22) $V(p) = \frac{(N-n)}{Nn}\frac{NPQ}{N-1} = \frac{(N-n)}{N-1}\frac{PQ}{n}$

The standard error of p is:

(23) $S_p = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{PQ}{n}}$

The coefficient of variation of p is:

(24) $CV(p) = \frac{S_p}{P} = \sqrt{\frac{N-n}{(N-1)n}}\sqrt{\frac{Q}{P}}$

By using the sample data unbiased estimates of the above magnitudes (22, 23, 24) are obtained:

The estimated variance of p is:

(25) $v(p) = \frac{(N-n)}{Nn}\frac{npq}{n-1} = \frac{N-n}{N}\frac{pq}{n-1}$

The estimated standard error of p is:

(26) $s_p = \sqrt{\frac{N-n}{N}}\sqrt{\frac{pq}{n-1}}$

The estimated coefficient of variation of p is:

(27) $cv(p) = \frac{s_p}{p} = \sqrt{\frac{(N-n)}{N(n-1)}}\sqrt{\frac{q}{p}}$

If N is large relative to n, the above formulae (22-27) are simplified as follows:

(22a) $V(p) = \frac{PQ}{n}$ (25a) $v(p) = \frac{pq}{n-1}$

(23a) $S_p = \frac{1}{\sqrt{n}}\sqrt{PQ}$ (26a) $s_p = \frac{1}{\sqrt{n-1}}\sqrt{pq}$

(24a) $CV(p) = \frac{1}{\sqrt{n}}\sqrt{\frac{Q}{P}}$ (27a) $cv(p) = \frac{1}{\sqrt{n-1}}\sqrt{\frac{q}{p}}$

From the above formula 22a it is obvious that for a given sample size V(p) is maximum when $P = \frac{1}{2}$. Also, formula 24a indicates that the coefficient of variation of p is very high when P is very small. This means that a very large sample size is needed in order to reduce the coefficient of variation of the estimate to reasonable limits, if the item under question is rare in the population (i.e. has a small P). In such a situation, this method becomes very expensive.

Example 5: A fish pond contains two kinds of fish (species A, species B). A simple random sample of n = 220 fish gave 55 fish belonging to class A ($n_A = 55$). Estimate the confidence limits for P in the population (P = 95 percent).

Solution:

$$p = \frac{n_A}{n} = \frac{55}{220} = 0.25$$

$$s_p = \frac{1}{\sqrt{224}}\sqrt{0.25x0.75} = 0.0289$$

$$0.25-1.96x0.0289<P<0.25+1.96x0.0289$$

$$0.1928<P<0.3072$$

15.2.6 Estimation for a subgroup

Quite often we make estimates for subgroups of a population in addition to the entire population. For example, in deck sampling1/ we wish to know the quantity (number, weight) of fish caught by species in addition to the total fish catch. In such a case we are interested for subgroups of the population. The sizes of subgroups are generally unknown. In our example, if a sample of n fish, selected from the population under investigation (fish on deck), contains n_g fish of a given species, these n_g fish form a random sample from the N_g in the subgroup. However, the number n_g is not fixed like our n, but is a random variable in the sense that it is likely to vary from sample to sample. For estimation purposes the following procedure can be used. Those units which do not belong to the subgroup are supposed to have a value of zero. It has been proved that an unbiased estimate of the subgroup mean is given by:

(28) $$\bar{y}_g = \frac{1}{n_g}\sum_{j=1}^{n_g} y_j$$

The variance of the estimate being:

(29) $$V(\bar{y}_g) = \{E(\frac{1}{n_g})-\frac{1}{N_g}\}S_g^2$$

As an estimate of the variance of $\bar{y}$ we may take:

1/ Deck Sampling: An Assessment of a Pilot Trawling Survey on Lake Malawi (Malawi), by G.P. Bazigos, UNDP/SF/MLW.16, February 1973

$$(30) \quad v(\bar{y}_g) = \frac{1}{n_g}(1-\frac{n}{N})s_g^2$$

where:

$$s_g^2 = \frac{1}{n_g-1}\left(\sum_{j=1}^{n_g} y_j^2 - \frac{\left(\sum_{j=1}^{n_g} y_j\right)^2}{n_g}\right)$$

For the subgroup total we have, estimated total:

$$(31) \quad \hat{Y}_g = \frac{N}{n}y_g, \quad (y_g = \sum_{j=1}^{n_g} y_j)$$

Estimated variance of the estimated total:

$$(32) \quad v(\hat{Y}_g) = N^2(\frac{1}{n}-\frac{1}{N})s^{*2}$$

where:

$$s^{*2} = \frac{1}{n-1}\left(\sum_{j=1}^{n} y_j^2 - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}\right)$$

Example 6: To get an estimate of the total number of gill nets owned by the trained fisherman (inland waters) in a given country the following procedure was used. From the general registry of fishermen ($N = 5000$ fishermen) a simple random sample of $n = 100$ fishermen was selected. The sample gave $n_g = 20$ trained fishermen. From each selected fisherman of group g information was selected of the total number of gill nets owned, variable y. The obtained sample magnitudes are:

$$\sum_{j=1}^{20} y_j = 52 \qquad \sum_{j=1}^{20} y_j^2 = 140$$

1. The estimated average number of gill nets owned per trained fisherman is,

   $\bar{y}_g = \frac{52}{20} = 2.6$ gill nets

2. The estimated variance per unit is,

   $s_g^2 = \frac{1}{19}\{140-\frac{(52)^2}{20}\} = 0.2526$ gill nets^2

3. The estimated variance of $\bar{y}_g$ is,

   $v(\bar{y}_g) = \frac{1}{20}(1-\frac{100}{5000})\text{x}0.2526 = 0.012377$ gill nets^2

4. The estimated coefficient of variation of $\bar{y}$ is,

   $cv(\bar{y}_g) = \frac{\sqrt{0.012377}}{2.6}\text{x}100 = 4.2$ percent

5. The estimated total number of gill nets owned by the trained fishermen is,

$$\hat{Y}_g = \frac{5000}{100}x52 = 2600 \text{ gill nets}$$

6. Calculated value of s^{*2} is,

$$s^{*2} = \frac{1}{99}\{140-\frac{(52)^2}{100}\} = 1.14 \text{ gill nets}^2$$

7. The estimated variance of $\hat{Y}_g$ is,

$$v(\hat{Y}_g) = 5000^2(\frac{1}{100}-\frac{1}{5000})x1.14 = 279300 \text{ gill nets}^2$$

8. The estimated coefficient of variation of $\hat{Y}_g$ is,

$$cv(\hat{Y}_g) = \frac{\sqrt{279300}}{2600}x100 = 20.33 \text{ percent}$$

15.3 Stratified sampling

Stratification is a method of making use of auxiliary information for improving the precision of the estimate. It has been seen that in simple random sampling the variance of the estimate (e.g. $V(\bar{y})$) depends, apart from the sample size, on the variability of the characteristic in the population. If the population is heterogeneous1/ it may be possible, by using auxiliary information, to divide it into sub-populations (or strata) each of which is internally homogeneous. If a simple random sample (wrp) is taken from each stratum it should be possible to make a precise estimate of the strata averages. These estimates can then be combined into a precise estimate for the population. Since the main purpose of stratification is to achieve a better precision, a number of problems arise for which solutions must be found i.e. a) Estimators for stratified sampling and their properties; b) Allocation of the overall sample to the strata; c) How to construct strata and how many.

15.3.1 Estimation of population mean and total

The fish catch of an experimental fishing boat is divided into a number of strata on the basis of, say, size of fish, thereby separating the very large ones, the medium sized ones, and the smaller ones. For the i^{th} stratum ($i = 1, 2, \ldots k$) the size will be N_i fish, the population total Y_i (weight of fish in kg), the variance per unit S_i^2. A random sample of size n_i is selected from the i^{th} stratum, the sample mean being $\bar{y}_i$. It is easy to see that an estimate of the population total:

$$Y_{st} = \sum_{i=1}^{k} N_i\bar{Y}_i$$

is given by:

$$(33) \quad \hat{Y}_{st} = \sum_{i=1}^{k} N_i\bar{y}_i$$

where,

$$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$$

1/ Measurements vary considerably from one unit to another.

Further, an estimate of the population mean is given by:

$$(34)\quad \bar{y}_{st} = \frac{\hat{Y}}{N} = \frac{1}{N}\sum_{i=1}^{k} N_i\bar{y}_i$$

The variance of the estimated total is[1]:

$$(35)\quad V(\hat{Y}_{st}) = \sum_{i=1}^{k} N_i^2 V(\bar{y}_i) = \sum_{i=1}^{k} N_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)S_i^2$$

$$= \sum_{i=1}^{k} \frac{N_i(N_i-n_i)}{n_i}S_i^2$$

where,.

$$S_i^2 = \frac{1}{N_i-1}\left(\sum_{j=1}^{N_i} y_{ij}^2 - \frac{\left[\sum_{j=1}^{N_i} y_{ij}\right]^2}{N_i}\right)$$

The estimated variance of $\hat{Y}$ is:

$$(36)\quad v(\hat{Y}_{st}) = \sum_{i=1}^{k} N_i^2 v(\bar{y}_i) = \sum_{i=1}^{k} N_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)s_i^2$$

$$= \sum_{i=1}^{k} \frac{N_i(N_i-n_i)}{n_i}s_i^2$$

where,

$$s_i^2 = \frac{1}{n_i-1}\left(\sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left[\sum_{j=1}^{n_i} y_{ij}\right]^2}{n_i}\right)$$

If the sampling fraction $\left(\frac{n_i}{N_i}\right)$ is negligible in all strata the above formulae (35, 36) are simplified as:

$$(35a)\quad V(\hat{Y}_{st}) = \sum_{i=1}^{k} \frac{N_i^2}{n_i}S_i^2$$

$$(35b)\quad v(\hat{Y}_{st}) = \sum_{i=1}^{k} \frac{N_i^2}{n_i}s_i^2$$

The variance of the estimated mean is:

$$(37)\quad V(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i^2 V(\bar{y}_i)$$

$$= \frac{1}{N^2}\sum_{i=1}^{k} \frac{N_i(N_i-n_i)}{n_i}S_i^2$$

1/ It should be noted that with stratified sampling, there is in general no single finite population correction factor, the factors entering individually into each stratum.

The estimated variance of $\bar{y}_{st}$ is:

$$(38)\quad v(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i^2 v(\bar{y}_i)$$

$$= \frac{1}{N^2}\sum_{i=1}^{k} \frac{N_i(N_i-n_i)}{n_i} s_i^2$$

If the sampling fraction $(\frac{n_i}{N_i})$ is negligible in all strata the above formulae (37, 38) are simplified as follows:

$$(37a)\quad V(\bar{y}_{st}) = \sum_{i=1}^{k} W_i^2 \frac{S_i^2}{n_i}, \quad (W_i = \frac{N_i}{N})$$

$$(38a)\quad v(\bar{y}_{st}) = \sum_{i=1}^{k} W_i^2 \frac{s_i^2}{n_i}$$

Example 7: The table below gives the number of landings (N = 25 landings) on day d_o at a given fishing site along with their catches in kg. Estimate the average catches per landing and total catches by: a) Selecting a simple random sample of size n = 10 landings; b) A stratified random sample of size n = 10 ($n = \sum_{i=1}^{k=5} n_i$, n_i = 2 landings).

Fish catch (kg)

						Parameters of total population:
(Str:1)	1,	3,	2,	5,	4,	
(Str:2)	15,	11,	13,	14,	12,	Y = 575 kg
(Str:3)	23,	21,	22,	25,	24,	$\bar{Y}$ = 23 kg/landing
(Str:4)	31,	32,	35,	33,	34,	S^2 = 210.4kg^2
(Str:5)	43,	42,	41,	45,	44	

a) Simple Random Sample

Sample data, kgs (3, 31, 22, 33, 43, 21, 24, 35, 5, 12)

1. Sample total $y = \sum_{j=1}^{10} y_j = 229$
2. Sample mean $\bar{y} = \frac{y}{n} = 22.9$ kg/landing
3. Estimated total landings $\hat{Y} = N\bar{y} = 25\text{x}22.9 = 572.5$ kg
4. Variance of the estimated total,

   $v(\hat{Y}) = \frac{N(N-n)}{n}s^2 = \frac{25\text{x}15}{10}\text{x}210.4 = 7890\ \text{kg}^2$
5. Coefficient of variation of $\hat{Y}$,

   $CV(\hat{Y}) = \frac{\sqrt{7890}}{572.5} = 15.51$ percent

b) Stratified Sample

Sample data (kg) by stratum:

Stratum-1: (1, 5)
Stratum-2: (11, 13)
Stratum-3: (25, 21)
Stratum-4: (32, 35)
Stratum-5: (43, 45)

1. Calculated magnitudes by stratum:

Stratum	N_i	n_i	$\bar{Y}_i$ (kg)	$\bar{y}_i$ (kg)	S_i^2 (kg²)	s_i^2 (kg²)
Str:1	5	2	3.0	3.0	2.5	8
Str:2	5	2	13.0	12.0	2.5	2
Str:3	5	2	23.0	23.0	2.5	8
Str:4	5	2	33.0	33.5	2.5	4.5
Str:5	5	2	43.0	44.0	2.5	2

2. Estimated population total:

$$\hat{Y}_{st} = \sum_{i=1}^{5} N_i\bar{y}_i = (5x3)+(5x12)+(5x23)+(5x33.5)+(5x44) = 577.5 \text{ kg}$$

3. Estimated population mean:

$$\bar{y}_{st} = \frac{\hat{Y}}{25} = \frac{577.5}{25} = 23.1 \text{ kg/landing}$$

4. Variance of the estimated total:

$$V(\hat{Y}_{st}) = \sum_{i=1}^{5} \frac{N_i(N_i-n_i)}{n_i}S_i^2 = 93.75 \text{ kg}^2$$

5. Coefficient of variation of $\hat{Y}$:

$$CV(\hat{Y}_{st}) = \frac{\sqrt{93.75}}{577.5} = 1.7 \text{ percent}$$

15.3.2 Allocation of the total sample to the strata

We now consider the problem of the allocation of the total sample size n to the different strata. It should be noted that the precision of the estimate depends heavily on the allocation made. Provided that the strata have already been constructed, there are two methods by which the total sample can be distributed among the strata

1. Proportional allocation.
2. Optimum allocation.

1. Proportional allocation

Proportional allocation is used in practice when auxiliary information on strata variances is not available. This method gives the most efficient estimates (i.e. those with the smallest sampling variance) for a given sample size (n), provided all the within-stratum variances are equal. In proportional allocation the sampling fraction is constant from stratum to stratum:

$$(39)\quad \frac{n}{N} = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \ldots = \frac{n_i}{N_i} = \ldots = \frac{n_k}{N_k}$$

$$\text{or,}\quad n_i = n\frac{N_i}{N} = nW_i, \quad (W_i = \frac{N_i}{N})$$

This allocation of the sample size gives a self-weighting sample. If many items are involved, self-weighting estimates are of great interest. It is easy to see that the estimate of the population total takes a simple form:

$$\hat{Y}_{pr} = \sum_{i=1}^{k} N_i \bar{y}_i = \sum_{i=1}^{k} \frac{N_i}{n_i} y_i, \quad (y_i = \sum_{j=1}^{n_i} y_{ij})$$

If, $\frac{N_i}{n_i} = c = \frac{N}{n}$, then:

$$(40)\quad \hat{Y}_{pr} = c\sum_{i=1}^{k} y_i = c\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$$

With proportional allocation the variance of $\hat{Y}$ is:

$$(41)\quad V(\hat{Y}_{pr}) = \sum_{i=1}^{k} \frac{N_i^2}{n_i}(1-\frac{n_i}{N_i})S_i^2$$

Or:

$$(42)\quad V(\hat{Y}_{pr}) = \frac{N-n}{n}\sum_{i=1}^{k} N_i S_i^2$$

The estimated variance of population total is:

$$(43)\quad v(\hat{Y}_{pr}) = \frac{N-n}{n}\sum_{i=1}^{k} N_i s_i^2$$

If the sampling fraction $(\frac{n_i}{N_i})$ is negligible in all strata the above formulae (42, 43) are simplified as follows:

$$(42a)\quad V(\hat{Y}_{pr}) = \frac{N}{n}\sum_{i=1}^{k} N_i S_i^2 = c\sum_{i=1}^{k} N_i S_i^2$$

$$(43a)\quad v(\hat{Y}_{pr}) = \frac{N}{n}\sum_{i=1}^{k} N_i s_i^2 = c\sum_{i=1}^{k} N_i s_i^2$$

The estimated population mean is:

$$(44)\quad \bar{y}_{pr} = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{y}_i = \frac{1}{n}\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$$

The variance of $\bar{y}_{pr}$ is:

$$(45)\quad V(\bar{y}_{pr}) = \frac{N-n}{nN}\sum_{i=1}^{k} W_i S_i^2$$

The estimated variance of $\bar{y}_{pr}$ is:

(46) $v(\bar{y}_{pr}) = \frac{N-n}{nN}\sum_{i=1}^{k} W_i s_i^2$

If the sampling fraction $(\frac{n_i}{N_i})$ is negligible in all strata the above formulae (45, 46) are simplified as follows:

(45a) $V(\bar{y}_{pr}) = \frac{1}{n_i}\sum_{i=1}^{k} W_i S_i^2$

(46a) $v(\bar{y}_{pr}) = \frac{1}{n_i}\sum_{i=1}^{k} W_i s_i^2$

Example 8: The registered fishing economic units (FEU's) at Lake A have been grouped into three strata by taking as a criterion of stratification the kind of fishing boat used. In order to get estimates on the number of gill nets owned (y) by the fishermen, a random sample of FEU's was selected (stratified sample - proportional allocation). The table below shows the units in the population and in the sample on a stratum basis and the sample totals of the characteristic under investigation. Estimate the confidence intervals of Y and $\bar{Y}$ respectively (P = 95 percent).

Strata of FEU's	Number of FEU's			y_i	s_i^2	$N_i s_i^2$	$W_i s_i^2$
	Population		Sample				
	N_i	W_i(%)	n_i				
Str: 1	400	50.0	50	300	6	2400	3
Str: 2	240	30.0	30	400	10	2400	3
Str: 3	160	20.0	20	300	20	3200	4
TOTAL	N=800	100.0	n=100	1000 $(\sum_i \sum_j y_{ij})$		8000 $(\sum_i N_i s_i^2)$	10 $(\sum_i W_i s_i^2)$

1. Estimated population total:

$\hat{Y}_{pr} = \frac{800}{100}x1000 = 8000$ gill nets

2. Estimated variance of the estimated total:

$v(\hat{Y}_{pr}) = \frac{800-100}{100}x8000 = 56000$ gill nets^2

3. Estimated standard error of $\hat{Y}_{pr}$:

$s\hat{Y}_{pr} = \sqrt{56000} = 236.64$ gill nets

4. Estimated confidence interval of Y (P = 95 percent):

$8000-1.96x236.64<Y<8000+1.96x236.64$

7358.19 gill nets<Y<8461.81 gill nets

5. Estimated population mean:

$\bar{y}_{pr} = \frac{1000}{100} = 10$ gill nets

6. Estimated variance of $\bar{y}_{pr}$:

$$v(\bar{y}_{pr}) = \frac{800-100}{100x800}x10 = 0.875 \text{ gill nets}^2$$

7. Estimated standard error of $\bar{y}_{pr}$:

$$s_{\bar{y}_{pr}} = \sqrt{0.875} = 0.9354 \text{ gill nets}$$

8. Estimated confidence interval of $\bar{Y}$, (P = 95 percent):

$$10-1.96x0.9354<\bar{Y}<10+1.96x0.9354$$

$$8.17 \text{ gill nets}<\bar{Y}<11.83 \text{ gill nets}$$

2. Optimum allocation

If the within-stratum variances differ greatly from one stratum to another, the method of proportional allocation no longer gives the best possible estimates. In this situation the sampling fraction for any stratum should be taken proportional to within-stratum standard deviation. Thus:

$$(47) \quad n_i = n\left(\frac{N_i S_i}{\sum N_i S_i}\right)$$

This allocation (optimum allocation) gives the smallest variance for the estimated mean (total) for a fixed total sample size n.

The variance of the estimated total is:

$$(48) \quad V(\hat{Y}_{opt}) = \frac{1}{n}\left(\sum_{i=1}^{k} N_i S_i\right)^2 - \sum_{i=1}^{k} N_i S_i^2$$

The estimated variance of the estimated total is:

$$(49) \quad v(\hat{Y}_{opt}) = \frac{1}{n}\left(\sum_{i=1}^{k} N_i s_i\right)^2 - \sum_{i=1}^{k} N_i s_i^2$$

If the sampling fraction $(\frac{n_i}{N_i})$ is negligible in all strata the above formulae (48, 49) are simplified as follows:

$$(48a) \quad V(\hat{Y}_{opt}) = \frac{1}{n}\left(\sum_{i=1}^{k} N_i S_i\right)^2$$

$$(49a) \quad v(\hat{Y}_{opt}) = \frac{1}{n}\left(\sum_{i=1}^{k} N_i s_i\right)^2$$

The variance of the estimated mean is:

$$(50) \quad V(\bar{y}_{opt}) = \frac{1}{n}\left(\sum_{i=1}^{k} W_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} W_i S_i^2$$

The estimated variance of the estimated mean is:

$$(51) \quad v(\bar{y}_{opt}) = \frac{1}{n}\left(\sum_{i=1}^{k} W_i s_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} W_i s_i^2$$

If the sampling fraction $(\frac{n_i}{N_i})$ is negligible in all strata the above formulae (50, 51) are simplified as follows:

$$(50a) \quad \bar{V}(\bar{y}_{opt}) = \frac{1}{n}\left[\sum_{i=1}^{k} W_i S_i\right]^2$$

$$(51a) \quad v(\bar{y}_{opt}) = \frac{1}{n}\left[\sum_{i=1}^{k} W_i s_i\right]^2$$

Example 9: By using the data of example 8, estimate the optimum n_i in the strata (n = 100).

Strata of FEU's	Number of FEU's N_i	s_i^2	$N_i s_i^2$		$n_i' = nW_i$	Remarks
				W_i' (%)		
Str:1	400	6	2400	30.0	30	
Str:2	240	10	2400	30.0	30	
Str:3	160	20	3200	40.0	40	
TOTAL	N=800		8000 ($\sum_i N_i s_i^2$)	100.0	n=100	

Example 10: The table below gives the sample totals by stratum for the optimum allocation of n. Estimate the confidence intervals of Y and $\bar{Y}$ respectively (P = 95 percent).

Strata of FEU's	Population		n_i'	y_i'	$N_i \bar{y}_i'$	$N_i s_i$	$N_i s_i^2$
	N_i	W_i (%)					
Str: 1	400	50.0	30	180	2400	979.76	2400
Str: 2	240	30.0	30	400	3200	756.95	2400
Str: 3	160	20.0	40	800	3200	715.53	3200
TOTAL	N=800	100.0	n=100		8800	2452.24 ($\sum_i N_i s_i$)	8000 ($\sum_i N_i s_i^2$)

1. Estimated population total:

$\hat{Y}_{opt}$ = 8800 gill nets

2. Estimated variance of the estimated total:

$v(\hat{Y}_{opt}) = \frac{1}{100}x6013481-8000 = 52134.81$ gill nets^2

3. Estimated confidence interval for Y:

$8800-1.96\times 228.33<Y<800+1.96\times 228.33$

8352 gill nets $<Y<$ 9247 gill nets

4. Estimated population mean:

$\bar{y}_{opt} = \frac{8800}{800} = 11$ gill nets

5. Estimated variance of the population mean:

$v(\bar{y}_{opt}) = \frac{1}{800^2}5213 4.81 = 0.8146$ gill nets^2

6. Estimated confidence interval of $\bar{Y}$:

$11-1.96\times 0.902<\bar{Y}<11+1.96\times 0.902$

9.23 gill nets $<\bar{Y}<$ 12.77 gill nets

15.3.3 Some properties of the estimators

It has been stated that one of the basic considerations involved in the use of stratification is the achievement of better precision for the estimated magnitudes. In this section we will describe in what way the gain due to stratification is achieved.

In the cases of simple random sample and stratified random sample with proportional and optimum allocation the variances of the estimated means are denoted by:

V_{ran} : Simple random sample

V_{pr} : Stratified sample/proportional allocation

V_{opt} : Stratified sample/optimum allocation.

If the total sample size is fixed and the sampling fraction $(\frac{n_i}{N_i})$ is negligible in all strata it could be proved that the following equations are valid (approximation):

(52) $$V_{ran} = V_{pr}+\frac{1}{n}\sum_{i=1}^{k}\frac{N_i}{N}(\bar{Y}_i-\bar{Y})^2$$

(53) $$V_{pr} = V_{opt}+\frac{1}{n}\sum_{i=1}^{k}\frac{N_i}{N}(S_i-\bar{S})^2$$

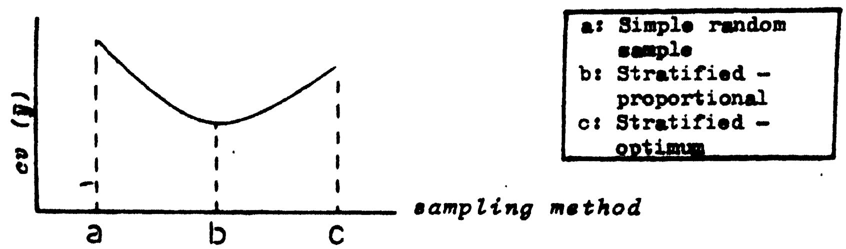
Where:

$$\bar{S} = \sum_{i=1}^{k}\frac{N_i}{N}S_i$$

The above equation (52) shows that proportional allocation would be very beneficial if the strata averages $\bar{Y}_i$ differ greatly from one stratum to another.

Equation (53) shows that optimum allocation would be very beneficial indeed if the strata variances S_i^2 differ greatly from one stratum to another.

In the figure below there is a graphical presentation of the level of precision for the population mean achieved through the sampling methods of simple random sample, stratified-proportional, stratified-optimum and for a fixed sample size n (CAS - Volta Lake):



15.3.4 **Estimation of the sample size**

In a stratified sample the principal steps involved in the choice of the sample size are:

1. We must know what is the expected precision of the estimate e.g. Vo: expected variance of $\bar{y}_{st}$.
2. Estimation of the within-strata variances.
3. Determining the method of allocating the total sample size to the strata.

In the case of proportional allocation the required total sample size for given Vo is:

$$(54) \quad n = \frac{\sum_{i=1}^{k} W_i S_i^2}{Vo + \frac{1}{N} \sum_{i=1}^{k} W_i S_i^2}$$

In the case of optimum allocation the required total sample size for given Vo is:

$$(55) \quad n = \frac{\left[\sum_{i=1}^{k} W_i S_i\right]^2}{Vo + \frac{1}{N} \sum_{i=1}^{k} W_i S_i^2}$$

If the sampling fraction $(\frac{n_i}{N_i})$ is negligible in all strata the above formulae (54, 55) are simplified as follows:

$$(54a) \quad n = \frac{1}{Vo} \sum_{i=1}^{k} W_i S_i^2$$

$$(55a) \quad n = \frac{1}{Vo} \left(\sum_{i=1}^{k} W_i S_i \right)^2$$

Example 11: At Lake A the active fishing economic units have been allocated into four strata by taking into account the kind of fishing boat used (dug-out canoe, plank-canoe, small plank-boat, large plank-boat). Within each stratum information is available on the variance per unit of the variate (y) {(y) : total number of gill nets used by FEU}. How many FEU's should be selected to achieve an estimate of the average number of gill nets used per FEU with a coefficient of variation of 5 percent? Calculations for the size of n should be made for the following cases:

1. Proportional allocation
2. Optimum allocation.

Strata of FEU's	Number of FEU's in the population		S_i^2	S_i	$W_iS_i^2$	W_iS_i
	N_i	W_i (%)				
Str: 1	200	50.00	4	2	2	1
Str: 2	100	25.00	16	4	4	1
Str: 3	80	20.00	25	5	5	1
Str: 4	20	5.00	36	6	1.8	0.3
TOTAL	N=400				12.8	3.3

$$CV(\bar{y}) = \frac{\sqrt{V(\bar{y})}}{\bar{Y}} = 0.05$$

Or:

$$Vo = V(\bar{y}) = (0.05x\bar{Y})^2$$

If the estimated $\bar{y}$=12, then:

$$\hat{V}o = 0.0025x144 = 0.36$$

In the case of proportional allocation:

$$n = \frac{12.8}{0.36+0.32} = \frac{12.8}{0.68} = 19 \text{ FEU's}$$

In the case of optimum allocation:

$$n = \frac{(3.3)^2}{0.36+0.32} = \frac{10.89}{0.68} = 16 \text{ FEU's}$$

15.3.5 Estimation of proportions

In vrp simple random samples of size n_i, $\sum_{i=1}^{k} n_i = n$, within strata, an unbiased estimate of the population proportion P is given by:

$$(56) \quad p_{st} = \frac{1}{N}\sum_{i=1}^{k} N_i p_i = \sum_{i=1}^{k} W_i p_i$$

Where:

p_i is the sample proportion of units in class A in i^{th} stratum, and W_i is the stratum weight:

$$(W_i = \frac{N_i}{N})$$

The variance of p_{st} is:

$$(57) \quad V(p_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} \frac{N_i(N_i-n_i)}{n_i} S_i^2$$

Where:

$$S_i^2 = \frac{N_i P_i Q_i}{N_i - 1}$$

The estimated variance of p_{st} is:

$$(58) \quad v(p_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} \frac{N_i(N_i - n_i)}{n_i} s_i^2$$

Where:

$$s_i^2 = \frac{n_i p_i q_i}{n_i - 1}$$

If N_i/N_i-1 can be taken as unity, the above formula (57) can be simplified as follows:

$$(57a) \quad V(p_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} \frac{N_i(N_i - n_i)}{n_i} P_i Q_i$$

From the above formula (57a) it is obvious that the variance of p_{st} depends on the product P_i and $Q_i = (1-P_i)$. The product is small if P_i is near to zero or to unity. Thus higher precision can be achieved if the strata can be formed in such a way that units belonging to the class under investigation (for which the proportion is sought) can be allocated to the same stratum.

In a stratified sample with proportional allocation ($n_i = n\frac{N_i}{N}$) the estimator of the variance of the proportion is:

$$(59) \quad V(p_{st}) = \frac{N-n}{Nn}\sum\frac{N_i}{N}\,\frac{N_i P_i Q_i}{N_i - 1} = \frac{N-n}{Nn}\sum W_i P_i Q_i$$

It has been indicated that if $1 = N_i/N_i-1$, then, $S_i^2 = P_i Q_i$. In such a case the optimum sample size n_i is given by:

$$(60) \quad n_i = n\frac{N_i\sqrt{P_i Q_i}}{\sum N_i\sqrt{P_i Q_i}}$$

Example 12: At a given inland water place the registered fishermen have been grouped into three strata by taking as a criterion of stratification their place of residence. In order to estimate the proportion of fishermen owners of boats, a sample of n = 200 fishermen was selected from the registry. The table below provides the sample values of p_i of the fishermen owners of fishing boats. Estimate the confidence interval for P (P = 95 percent).

Geographical strata	Number of fishermen in the population N_i	W_i (%)	n_i	p_i	$W_i p_i$	S_i^2 1/	$\frac{N_i(N_i-n_i)}{n_i} S_i^2$
Str: 1	1200	50.0	60	0.70	0.350	0.2136	4870
Str: 2	800	30.0	36	0.75	0.225	0.1929	3275
Str: 3	400	20.0	24	0.86	0.172	0.1256	787
TOTAL	N=2400	100.00	n=120		0.747		8932

1/ Population values

Estimated population proportion:

$$p_{st} = \sum_{i=1}^{k} W_i p_i = 0.747$$

Calculated variance of p_{st}:

$$V(p_{st}) = \frac{1}{(2400)^2} x8932 = 0.00155$$

Calculated confidence interval of P:

$$0.747-1.96x0.0394<P<0.747+1.96x0.0394$$

$$0.6698<P<0.8542$$

15.3.6 X-proportional allocation

In section 15.3.2 the allocation of the total sample size to the strata was made by using the N-proportional allocation. If measures of the control variable (x) are available for all units in the population, the sample sizes of n_i may be found as a proportion of X_i:

$$(61) \quad n_i = n\frac{X_i}{X}$$

Where:

$$X_i = \sum_{j=1}^{N_i} x_{ij}, \quad X = \sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij}$$

This allocation is called the X-proportional allocation. The variance of $\hat{Y}$ is given by:

$$(62) \quad V_{x-pr} = \frac{N}{n}\sum_{i=1}^{k} \frac{N_i S_i^2}{\bar{X}_i/\bar{X}}$$

15.3.7 Construction and number of strata

With the construction of strata the ideal situation is that in which the distribution of the control variate is known. In such a case strata are formed by cutting up the distribution at suitable points. In the absence of this information the next best thing is the frequency of some other quantity that is highly correlated with the control variable. It should be noted that with skew populations in which a small proportion of the units accounts for a large proportion of the total (of y) a practical solution to the problem is to take the largest units into the sample with certainty and select a sample from the rest.

As far as the number of strata is concerned, at first sight the answer is that multiplicity of strata would improve the precision of the estimates. This is one reason that the survey statistician favours the use of a large number of strata. Multiplication of strata, beyond a reasonable number, is attended by some disadvantages, i.e. the cost and time of processing the data should be increased considerably.

15.4 Systematic sampling

Suppose a fish pond contains N fish, numbering from 1 to N. To select a sample of n units the following procedure can be used:

1. We calculate the spacing interval $z = \frac{N}{n}$
2. We select a unit of random from the z first units (j^{th} selected unit)
3. The sample contains the n units with serial numbers in the population, j, j+z, j+2z, ... , j+(n-1)z.

From the above analysis it is obvious that the first selected unit determines the whole sample1/. This type of sample has been called an every z^{th} systematic sample (with the z being the spacing interval).

It has been proved that in systematic sampling with spacing interval z, an unbiased estimate of the population mean is given by:

$$(63) \quad \bar{y}_s = \frac{y_s}{n} = \frac{1}{n}\sum_{j=1}^{n} y_{sj}$$

Where suffix s stands for the selected systematic sample (s = 1, 2, ... , z). The variance of the estimated mean is:

$$(64) \quad V(\bar{y}_s) = \frac{1}{z}\sum_{s=1}^{z}(\bar{y}_s-\bar{Y})^2$$

The above formula (64) shows that best results will be achieved if the variability within clusters is very high, i.e. the clusters are as heterogeneous as possible. It is easy to show that the variance of the systematic sample mean can also be written as:

$$(65) \quad V(\bar{y}_s) = \frac{S^2}{n}\{(1-\frac{1}{N})+(n-1)\rho_s\}$$

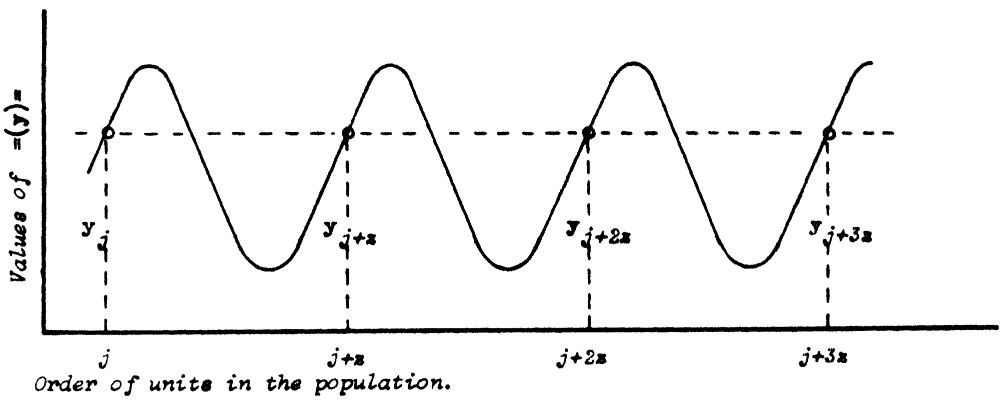
where ρ_s is the correlation coefficient between units in the same systematic sample (intracluster correlation coefficient). From formula (65) it is obvious that positive correlation between units in the same systematic sample increases the variance of estimate.

In practice, however, when we have reason to believe that the listing of units in the population can be considered to be random, the formula for estimating the variance of $\bar{y}$ for simple random sample (wrp) is applied (see section 15.2.1) for calculating $V(\bar{y}_s)$.

It should also be noted that the systematic sample may give very poor results if the population consists of a periodic trend. In fact, if the spacing interval z is an even multiple of the half-period, the sample is no more precise than a single observation selected at random from the population (see figure below). However, if the sampling interval is an odd multiple of the half-period, the estimate would be very precise.

Systematic sampling is a great advantage in a large scale survey where the sample is to be selected at the spot in the field and where we want to make sure that the recorder has made no mistakes in the process of selection. Again, since the sample is evenly spread over the whole population, we expect to obtain fairly precise results.

1/ Another way of looking at systematic sampling is that the population has been divided into z large sampling units (clusters) each of which contains n of the original units. The operation of choosing a randomly located systematic sample is just the operation of choosing one cluster at random from the z clusters.



Further notes on the variance of $\bar{y}_s$: With regard to the estimation of error of the sample estimate it is obvious from formula (64) that it is not possible to get an unbiased estimate from a single sample. However, if more than one systematic sample is selected, one could estimate the variance of $\bar{y}_s$ easily. Another estimator for the variance of $\bar{y}_s$ is given by:

$$(66) \quad V(\bar{y}) = \frac{N-1}{N}S^2 - \frac{n-1}{n}S_w^2$$

Where:

$$S^2 = \frac{1}{n-1}\sum_{s=1}^{z}\sum_{s=1}^{n}(y_{sj}-\bar{Y})^2$$

And:

$$S_w^2 = \frac{1}{z(n-1)}\sum_{s=1}^{z}\sum_{j=1}^{n}(y_{sj}-\bar{Y}_s)^2$$

The S_w^2 is also called "variance within the clusters".

Example 13: At a beach B, a systematic sample of landings was selected on a given day. From each selected landing information was collected on the total fish catch (y), by using the "real measurement" approach. The sample data are given below. Estimate the coefficient of variation of $\bar{y}_s$.

Sample data:

$N = 50$
$n = 10$

$$\sum_{j=1}^{10} y_i = 250\text{kg}$$

$$\sum_{j=1}^{10} y_j^2 = 6430\text{kg}^2$$

Estimated population mean:

$\bar{y}_s = 25$kg per landing

Estimated variance per unit:

$$s_y^2 = \frac{1}{9}\{6430-\frac{(250)^2}{10}\} = 20\text{kg}^2$$

Estimated variance of $\bar{y}_s$:

$$v(\bar{y}_s) = (\frac{1}{10}-\frac{1}{50})20 = 1.6kg^2$$

$$cv(\bar{y}_s) = \frac{\sqrt{1.6}}{25}x100 = 5.06 \text{ percent}$$

15.5 Ratio estimation

In this section we will present another method in which emphasis is laid on the use of auxiliary information for improving the precision of estimates. If the values of (x) (auxiliary variate) are known for all the units in the population and the ratio of y (survey variate) to x does not differ considerably from unit to unit, it may be advantageous to estimate the population ratio:

$$R = \frac{Y}{X}$$

from the sample and thereby estimate the population mean or total. This estimate is called ratio estimate. In the case of simple random sampling an estimate (biased) of the population total is given by:

$$(67) \quad \hat{Y}_{rat} = \frac{y}{x}X = \hat{R}X$$

where:

$$y = \sum_{j=1}^{n} y_j, \quad x = \sum_{j=1}^{n} x_j, \quad X = \sum_{j=1}^{N} x_j$$

An estimate of the population mean is given by:

$$(68) \quad \bar{y}_{rat} = \frac{y}{x}\bar{X} = \hat{R}\bar{X}$$

It should be noted that in large samples the bias of the estimate is negligible[1/].

The exact expression for the variance of the estimate is very complicated. As an approximation of the variance of Y we may take:

$$(69) \quad V(\hat{Y}_{rat}) = N^2(\frac{1}{n}-\frac{1}{N})S^2_{yx}$$

Where:

$$S^2_{yx} = \frac{1}{N-1}\sum_{j=1}^{N}(y_j-Rx_j)^2$$

Or:

$$V(\hat{Y}_{rat}) = \frac{N(N-n)}{n}(S^2_y+R^2S^2_x-2R\rho S_yS_x)$$

Where ρ: coefficient of correlation between x_j and y_j.

As an estimate of the variance of $\hat{Y}$ we may take:

$$(70) \quad v(\hat{Y}_{rat}) = N^2(\frac{1}{n}-\frac{1}{N})s^2_{yx}$$

1/ It has been proved that the estimators $\hat{Y}_{rat}$ and $\bar{y}_{rat}$ are biased i.e. $E(\hat{Y}_{rat})\neq Y$ and $E(\bar{y}_{rat})\neq \bar{Y}$. This bias is negligible in large samples (n>50). In such a case, the ratio is normally distributed and the large-scale formula for its variance is valid

Where:

$$s_{yx}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j-\hat{R}x_j)^2, \qquad \hat{R} = \frac{\bar{y}}{\bar{x}}$$

Or:

$$v(\hat{Y}_{rat}) = \frac{N(N-n)}{n}(s_y^2+\hat{R}^2s_x^2-2\hat{R}\hat{\rho}s_ys_x)$$

As an approximation of the variance of $\bar{y}_{rat}$ we may take:

$$(71) \quad V(\bar{y}_{rat}) = \frac{1}{N^2}V(\hat{Y}) = \frac{(N-n)}{Nn(n-1)}\sum_{j=1}^{N}(y_j-Rx_j)^2$$

Or:

$$V(\bar{y}_{rat}) = (\frac{N-n}{Nn})(S_y^2+R^2S_x^2-2R\rho S_xS_y)$$

As an estimate of the variance of $\bar{y}_{rat}$ we may take:

$$(72) \quad v(\bar{y}_{rat}) = \frac{(N-n)}{Nn(n-1)}\sum_{j=1}^{n}(y_j-\hat{R}x_j)^2$$

Or:

$$v(\bar{y}_{rat}) = \frac{(N-n)}{Nn}(s_y^2+\hat{R}^2s_x^2-2\hat{R}\hat{\rho}s_ys_x)$$

We shall now compare the variance of the estimate based on the sample mean of y's (simple random sample) and the ratio estimate. We have:

1. $V(\hat{Y}) = N^2(\frac{1}{n}-\frac{1}{N})S_y^2$

2. $V(\hat{Y}_{rat}) = N^2(\frac{1}{n}-\frac{1}{N})(S_y^2+R^2S_x^2-2\hat{R}\rho S_xS_y)$

Hence the ratio estimate has the smaller variance if:

$$\rho > \frac{1}{2}\,\frac{CV(x)}{CV(y)}$$

Example 14: At a given man-made lake there are 110 fishing economic units. In order to estimate the total number of fishing operations of the units on the lake in the course of a given year, the following procedure was used. A simple random sample of FEU's was selected (n = 15). From each sample unit information was obtained on the survey characteristic (y) and the number of active fishermen (x). The total number of active fishermen at the lake is known (results of a Frame Survey). The table below gives the obtained sample data. Estimate the confidence interval of Y (total number of fishing operations at the given year).

Fishing site	x_j	y_j	Fishing site	x_j	y_j	Fishing site	x_j	y_j	Remarks
01	2	220	06	3	220	11	2	200	$\sum_{j=1}^{110} x_i = 340$ active fishermen (Frame Survey)
02	3	200	07	2	150	12	2	200	
03	4	230	08	3	190	13	3	220	
04	2	180	09	4	250	14	4	220	
05	3	200	10	4	250	15	4	250	
							$\sum x_j = 45$	$\sum y_j = 3180$	

Estimated total number of fishing operations on the lake at the given year:

$$\hat{Y}_{rat} = \frac{3180}{45}x340 = 22249 \text{ fishing operations}$$

Estimated variance of $\hat{Y}$:

$$v(\hat{Y}_{rat}) = \frac{N(N-n)}{n(n-1)}\sum_{j=1}^{n}(y_j-\hat{R}x_j)^2$$

$$= \frac{N(N-n)}{n(n-1)}\left[\sum_{j=1}^{n}y_j^2+\hat{R}^2\sum_{j=1}^{n}x_j^2-2\hat{R}\sum_{j=1}^{n}x_jy_j\right]$$

Or:

$$v(\hat{Y}_{rat}) = \frac{110x95}{15x14}(685000+4992.84x145-2x70.66x9790) = 25439 \text{ fishing operations}^2$$

Estimated confidence interval of Y (P = 95 percent):

$$22249-1.96x1123.76<Y<22249+1.96x1123.76$$

$$20047 \text{ fishing operations}<Y<24451 \text{ fishing operations}$$

15.5.1 <u>The use of ratio estimation in estimating proportions</u>

Suppose that on a given day the total number of baskets of smoked fish which reached a lakeside market is N(j = 1, 2, ... , N). Each basket contains fish of two different species, A and B. A simple random sample of size n of baskets is selected and from each sample basket information is collected on the total weight of the basket (x_j) and the weight of species A (y_j).

The proportion of species A reaching the given market is:

$$(73) \quad P = \frac{\sum_{j=1}^{N} y_j}{\sum_{j=1}^{N} x_j}$$

As an estimate of P we take:

$$(74) \quad p = \frac{\sum_{j=1}^{n} y_j}{\sum_{j=1}^{n} x_j} = \frac{y}{x} = \hat{R}$$

The variance of p can be taken as follows:

a. $p = \hat{R} = \frac{\hat{Y}}{X}$

b. $V(\hat{R}) = \frac{1}{X^2}v(\hat{Y}) = \frac{1}{X^2}\,\frac{N(N-n)}{n}s_{yx}^2$

Or:

$$(75) \quad V(p) = V(\hat{R}) = \frac{N(N-n)}{nX^2}\,\frac{\sum_{j=1}^{N}(y_j-Rx_j)^2}{N-1}$$

As an estimate (approximation) of the variance of p we may take:

$$X^2 = N^2\bar{x}^2$$

$$(76) \quad v(p) = v(\hat{R}) = \frac{(N-n)}{nN}\frac{1}{\bar{x}^2}\frac{\sum_{j=1}^{n}(y_j-\hat{R}x_j)^2}{n-1}$$

Example 15: At a given inland water place there are 1500 registered FEU's. A simple random sample of 30 FEU's was selected and information was obtained from each sample FEU on the employment status of the active fishermen. By using the sample data estimate the confidence interval of the proportion of salaried (cash) assistant fishermen in the population. In the table, the symbol x_j expresses the total number of active fishermen per sample FEU and the symbol y_j the respective salaried assistant fishermen.

x_j	y_j	x_j	y_j	x_j	y_j	x_j	y_j	x_j	y_j	Remarks
6	3	3	1	7	3	3	2	2	1	
5	1	3	1	4	3	3	1	4	3	
2	1	4	2	3	2	4	1	3	1	
3	1	4	3	5	3	3	2	4	2	
3	1	3	2	4	3	3	2	2	1	
3	1	2	1	4	3	1	0	4	2	
								$\sum_{j=1}^{30} x_j=104$	$\sum_{j=1}^{30} y_j=53$	

Estimated proportion of salaried (cash) assistant fishermen:

$$p = \frac{53}{104} = 0.5096$$

Estimated variance of p:

$$v(p) = \frac{(N-n)}{Nn}\frac{1}{\bar{x}^2(n-1)}\left(\sum_{j=1}^{30} y_j^2+\hat{R}^2\sum_{j=1}^{30} x_j^2-2\hat{R}\sum_{j=1}^{30} x_jy_j\right)$$

Or:

$$v(p) = \frac{(1500-30)}{1500x30}x\frac{1}{(3.466)^2x29}\{117+(\frac{53}{104})^2x404-2x(\frac{53}{104})x206\} = 0.000937$$

Estimated confidence interval of P:

$$0.4496<P<0.5696$$

15.5.2 The use of ratio estimation in stratified random sampling

There are mainly two ways in which a ratio estimate of the population total Y can be made:

1. Separate Ratio Estimate: A separate estimate is made of the total of each stratum and these totals are added:

$$(77) \quad \hat{Y}_{rat.s} = \sum_{i=1}^{k}\frac{y_i}{x_i}X_i = \sum_{i=1}^{k}\frac{\bar{y}_i}{\bar{x}_i}X_i = \sum_{i=1}^{k}\hat{R}_iX_i$$

If the sample sizes n_i are large in all strata the variance of $\hat{Y}$ is:

$$(78) \quad V(\hat{Y}_{rat.s}) = \sum_{i=1}^{k} V_i$$

Where:

$$V_i = \frac{N_i(N_i-n_i)}{n_i} \frac{1}{N_i-1} \sum_{j=1}^{N_i} (y_{ij}-R_i x_{ij})^2$$

The estimated variance of $\hat{Y}$ is:

$$(79) \quad v(\hat{Y}_{rat.s}) = \sum_{i=1}^{k} v_i$$

Where:

$$v_i = \frac{N_i(N_i-n_i)}{n_i} \frac{1}{n_i-1} \sum_{j=1}^{n_i} (y_{ij}-\hat{R}_i x_{ij})^2$$

2. Combined Ratio Estimate: From the sample data we first compute:

$$\hat{Y}_{st.} = \sum_{i=1}^{k} N_i \bar{y}_i, \qquad \hat{X}_{st.} = \sum_{i=1}^{k} N_i \bar{x}_i, \qquad \hat{R} = \frac{\hat{Y}}{\hat{X}}$$

If the total sample size n is large, the variance of $\hat{Y}_{rat.s}$ is given by:

$$(80) \quad V(\hat{Y}_{rat.s}) = \sum_{i=1}^{k} V_i$$

Where:

$$V_i = \frac{N_i(N_i-n_i)}{n_i} \frac{1}{N_i-1} \sum_{j=1}^{N_i} (y_{ij}-R x_{ij})^2$$

The estimated variance of $\hat{Y}_{rat.s}$ is:

$$(81) \quad v(\hat{Y}_{rat.s}) = \sum_{i=1}^{k} v_i$$

Where:

$$v_i = \frac{N_i(N_i-n_i)}{n_i} \frac{1}{n_i-1} \sum_{j=1}^{n_i} (y_{ij}-\hat{R} x_{ij})^2$$

15.6 Difference estimation

Another method which is used in the field of large-scale sample surveys in order to improve the precision of the estimates, by making use of supplementary information, is the difference estimation. We have seen that the ratio method will give a good estimate if the line representing the relationship between y and x passes through the origin. If this line does not go through the origin, it would be better to use an estimate based on the regression y on x, rather than on the ratio of y to x. When several items are to be estimated from the same survey the difference estimation replaces the regression estimation because of the simplicity to calculate the estimates of the survey characteristics.

As an estimate of the population mean we may take:

$$(82) \quad \bar{y}_D = \bar{y}+k(\bar{X}-\bar{x})$$

Where:

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j, \quad \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j, \quad \bar{X} = \frac{1}{N}\sum_{j=1}^{N} x_j$$

k : the constant k is usually taken equal to unity1/.

It should be noted that $\bar{y}_D$ is an unbiased estimate.

$$\begin{aligned} E(\bar{y}_D) &= E(\bar{y})+kE(\bar{X}-\bar{x}) \\ &= \bar{Y}+k\bar{X}-kE(\bar{x}) \\ &= \bar{Y}+k\bar{X}-k\bar{X} = \bar{Y} \end{aligned}$$

The variance of $\bar{y}_D$ is given by:

$$(83) \quad V(\bar{y}_D) = \frac{(N-n)}{Nn}\,\frac{1}{N-1}\sum_{j=1}^{N}\{(y_j-kx_j)-(\bar{Y}-k\bar{X})\}^2$$

An unbiased estimate of the variance of $\bar{y}_D$ is:

$$v(\bar{y}_D) = \frac{(N-n)}{Nn}\,\frac{1}{n-1}\sum_{j=1}^{n}\{(y_j-kx_j)-(\bar{y}-k\bar{X})\}^2$$

Or:

$$(84) \quad v(\bar{y}_D) = \frac{(N-n)}{Nn(n-1)}\left(\sum_{j=1}^{n} y_j^2+k^2\sum_{j=1}^{n} x_j^2-2k\sum_{j=1}^{n} x_j y_j-\frac{y^2}{n}-k^2\frac{x^2}{n}+2k\frac{xy}{n}\right)$$

An estimate of the population total is given by:

$$(85) \quad \hat{Y}_D = N\bar{y}_D = N\{\bar{y}+k(\bar{X}-\bar{x})\}$$

The variance of $\hat{Y}_D$ is:

$$(86) \quad V(\hat{Y}_D) = N^2V(\bar{y}_D)$$

An unbiased estimate of the variance of $\hat{Y}_D$ is:

$$\begin{aligned} (87) \quad v(\hat{Y}_D) &= N^2v(\bar{y}_D) \\ &= \frac{N(N-n)}{n(n-1)}\left(\sum_{j=1}^{n} y_j^2+k^2\sum_{j=1}^{n} x_j^2-2k\sum_{j=1}^{n} x_j y_j-\frac{y^2}{n}-k^2\frac{x^2}{n}+2k\frac{xy}{n}\right) \end{aligned}$$

Example 16: By using the sample data of example 14, calculate the $\bar{y}_D$ and $cv(\bar{y}_D)$.

Calculated magnitudes:

$\bar{y}$ = 212 fishing operations
$\bar{x}$ = 3 active fishermen
$\bar{X}$ = 3.09 active fishermen
k = 1

Estimated average number of fishing operations per FEU during the given year:

$\bar{y}_D$ = 212+1(3.09-3.00) = 212.09 fishing operations

1/ The precision of k increases as $k \to \beta$, where β is the regression coefficient in the regression equation $y = a+\beta x$.

Estimated variance of $\bar{y}_D$:

$$v(\bar{y}_D) = \frac{95}{110 \times 15 \times 14}(685000+145-2 \times 9790-\frac{3180^2}{15}-\frac{45^2}{15}+\frac{2 \times 3180 \times 45}{15})$$

$$= 42.44 \text{ fishing operations}^2$$

Estimated coefficient of variation of $\bar{y}_D$:

$$cv(\bar{y}_D) = \frac{\sqrt{42.44}}{212.09} \times 100 = 3.07 \text{ percent}$$

15.7 Estimation in unequal probability sampling

Stratification, ratio, regression and difference estimation are some of the techniques by which the variability in the size of the units is controlled. Another such technique is pps sampling where the units are selected with probabilities proportionate to their size. This method has a good deal of application in the field of large-scale surveys where the sampling of clusters of units, with or without sub-sampling within clusters, is preferred to direct sampling of individual units for one of two reasons: One is the difficulty of organizing a sample of individual units in the absence of a reliable frame. The second reason for cluster sampling is that sampling of individual units does not make the most economical use of the available resources.

pps method has the advantage of giving unbiased and easily calculated estimates of means, totals and variances. In order that this shall be so selection must be made with replacement. The measure of cluster size is the supplementary information available for the units. For example, the measure of size of a fishing site may be its number of fishing economic units, FEU's (data available from a Frame Survey). In this method a unit with a larger size has a higher chance of selection as compared to one with a smaller size. A simple method of selection of the sample would consist in allotting numbers to the units in proportion to their size and then use the table of random numbers as usual. As a demonstration of this process of selection, consider the problem of selecting three fishing sites from those listed in the table below;

Ser.No. of fishing site	No. of FEU's (x)	Cumulative total of x	Allotted numbers	Selected fishing sites
01	12	12	001–012	Random No. 011, Fishing site sel. No.01
02	5	17	013–017	
03	20	37	018–037	Random No. 027, Fishing site sel. No.03
04	2	39	038–039	
05	30	69	040–069	Random No. 064, Fishing site sel. No.05
06	15	84	070–084	
07	8	92	085–092	
08	6	98	093–098	
09	8	106	099–106	
10	14	120	107–120	
Total	120			

If the number of units in the population is very large so that the cumulation of the sizes of the units becomes very laborious, the following equivalent procedure may be used:

1. N : Number of clusters in the population, e.g. fishing sites.
2. x_i : Size of the i^{th} cluster, e.g. number of FEU's ($i = 1, 2, \ldots, N$).
3. x_{max} : Maximum size.

Choose a pair of random numbers, one between 1 and N and the other between 1 and x_{max}. If the second random number is smaller than the size of the unit selected provisionally by the first random number, this unit is finally selected in the sample. If not, this unit is rejected and the selection is made afresh.

As a consequence of choosing units with probability proportional to (x) the estimation procedure is particularly simple. Let the sample selected by this procedure be given by:

$$\begin{cases} y_1, y_2, \cdots y_i \cdots y_n \\ p_1, p_2, \cdots p_i \cdots p_n \end{cases}$$

Where:

$p_i = \frac{x_i}{X}$, $\quad (X = \sum_{i=1}^{N} x_i)$, is the probability in which the i^{th} unit is selected in the sample of size x_i.

As an estimate of the population total Y, we take:

$$(88) \quad \hat{Y} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i} = \frac{X}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}$$

Note that in formula (88) the term $(1-\frac{n}{N})$ does not appear, as sampling is with replacement (wr).

The above estimator (88) is unbiased, i.e.:

$$E(\hat{Y}) = Y$$

The variance of the estimated total is:

$$(89) \quad V(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{N} p_i\left(\frac{y_i}{p_i}-Y\right)^2 = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{y_i^2}{p_i}-Y^2\right)$$

The estimated variance of $\hat{Y}$ is:

$$(90) \quad v(\hat{Y}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{p_i}-\hat{Y}\right)^2$$

Note that if $y_i=x_i$ then:

$$V(\hat{Y}) = 0$$

Example 17: At a new man-made lake there are 20 fishing sites. In order to get an estimate of the number of gill nets owned by the fishermen (y) the following survey method was applied. By using the "water approach" a Frame Survey was initiated on the basis of which eye observations were obtained on the number of fishing boats (x). A sample of four fishing sites was selected from the sampling frame (pps) and information was obtained on the survey variate. The table below gives the sample data. Estimate the $\hat{Y}$ and $cv(\hat{Y})$.

- sample data -

Sample fishing sites	Number of fishing boats seen x_i	Number of gill nets owned y_i	$p_i=\frac{x_i}{X}$	$t_i=\frac{y_i}{p_i}$	$t_i^2=(\frac{y_i}{p_i})^2$	Remarks
1	22	81	0.0443	1828	3341584	$(\sum_{i=1}^{20} x_i=496$ fishing boats seen)
2	30	118	0.0605	1950	3802500	
3	30	118	0.0605	1950	3802500	
4	42	170	0.0847	2007	4028049	
				7735 $(t=\sum_i t_i)$	14974633	

Estimated total number of gill nets owned by the fishermen:

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i} = \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{7735}{4} = 1934 \text{ gill nets}$$

Estimated variance of $\hat{Y}$:

$$v(\hat{Y}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\frac{y_i}{p_i}-\hat{Y})^2 = \frac{1}{n(n-1)}\sum_{i=1}^{n}(t_i-\bar{t})^2 = \frac{1}{n(n-1)}\left[\sum_{i=1}^{n} t_i^2-\frac{(\sum_{i=1}^{n} t_i)^2}{n}\right]$$

$$= \frac{1}{n(n-1)}\left[\sum_{i=1}^{n} t_i^2-(\frac{t^2}{n})\right] = \frac{1}{4\times3}\left[14974633-\frac{(7735)^2}{4}\right] = 1423 \text{ gill nets}^2$$

Estimated coefficient of variation of $\hat{Y}$:

$$cv(\hat{Y}) = \frac{\sqrt{v(\hat{Y})}}{\hat{Y}}x100 = \frac{\sqrt{1423}}{1934}x100 = 1.9 \text{ percent}$$

15.8 <u>Two-stage sampling</u>

When the clusters are large it is difficult to enumerate them completely. At the same time it is unnecessary to collect information on every individual in the sample clusters. Instead, it may be better to take a further sample of survey units from each selected cluster and collect information of the survey characteristics from the sample survey units. The sample is thus selected in two stages. At the first stage a sample of clusters are selected and at the second stage a sample of survey units are selected within the sample clusters. The sample design is called two-stage sampling.

We shall now study how to form estimates with their standard errors from a two-stage design.

Notation:

N : Number of first stage units (Primary Sampling Units, PSU's) $i = 1, 2, 3, \ldots, N$.

n : First stage sample size.

M_i: Size of the i^{th} PSU, i.e. number of survey units, here called Secondary Sampling Units, SSU's ($j = 1, 2, 3, \ldots, M_i$).

m_i: Second stage sample units within the i^{th} PSU.

$f_1 = \frac{n}{N}$: First-stage sampling fraction.

$f_{2i} = \frac{m_i}{M_i}$: Second-stage sampling fraction for i^{th} PSU.

15.8.1 **Estimation in equal probability sampling**

Let the fishing industry at a given inland water place consist of N fishing sites out of which a simple random sample of n fishing sites is selected. From a given sample fishing site containing M_i fishing economic units m_i FEU's are selected at random and investigated for the number of fishermen (y). Let y_{ij} be the number of fishermen of the j^{th} fishing economic unit from the i^{th} fishing site. As an estimate of the population total we take:

$$(91) \quad \hat{Y} = \frac{N}{n}\sum_{i=1}^{n} M_i \bar{y}_i$$

Where:

$$\bar{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$$

The above estimator (91) is unbiased, i.e.:

$$E(\hat{Y}) = Y$$

The variance of the estimated total is given by:

$$(92) \quad V(\hat{Y}) = \frac{N(N-n)}{n}\,\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i-\frac{Y}{N}\right)^2+\frac{N}{n}\sum_{i=1}^{N}\frac{M_i(M_i-m_i)}{m_i}S_i^2$$

Where:

$$\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i-\frac{Y}{N}\right)^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N}Y_i^2-\frac{Y^2}{N}\right)$$

And:

$$S_i^2 = \frac{1}{M_i-1}\sum_{j=1}^{M_i}(y_{ij}-\bar{Y}_i)^2 = \frac{1}{M_i-1}\left(\sum_{j=1}^{M_i}y_{ij}^2-\frac{Y_i^2}{M_i}\right)$$

Note that:

$$Y = \sum_{i=1}^{N}Y_i, \qquad Y_i = \sum_{j=1}^{M_i}y_{ij}$$

The estimated variance of $\hat{Y}$ is:

$$(93) \quad v(\hat{Y}) = \frac{N(N-n)}{n}\,\frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{Y}_i-\frac{\sum_{i=1}^{n}\hat{Y}_i}{n}\right)^2+\frac{N}{n}\sum_{i=1}^{n}\frac{M_i(M_i-m_i)}{m_i}s_i^2$$

Where:

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{Y}_i-\frac{\sum_{i=1}^{n}\hat{Y}_i}{n}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}\hat{Y}_i^2-\frac{\left(\sum_{i=1}^{n}\hat{Y}_i\right)^2}{n}\right)$$

And:

$$s_i^2 = \frac{1}{m_i-1}\left(\sum_{j=1}^{m_i}y_{ij}^2-\frac{\left(\sum_{j=1}^{m_i}y_{ij}\right)^2}{m_i}\right)$$

Example 18: At a given inland water place there are eight fishing sites. In order to get an estimate of the number of fish traps owned by the fishermen (y) a simple random sample of fishing sites was selected (n = 3). Within each selected fishing site a number of fishing economic units was taken ($m_i = m = 3$) and information was obtained from the selected units on the survey characteristic. The table below provides the obtained sample data. Calculate the magnitudes $\hat{Y}$ and $cv(\hat{Y})$.

- Number of traps -

Sample FEU's	Sample fishing sites 1st	2nd	3rd	Remarks
1	13	5	12	$s_1^2 = 12.3$ traps^2
2	9	7	8	$s_2^2 = 6.3$ traps^2
3	6	10	13	$s_3^2 = 7.0$ traps^2
TOTAL	28	22	33	
	$M_1=6$, $m_1=3$	$M_2=9$, $m_2=3$	$M_3=7$, $m_3=3$	

Estimated total number of traps:

$$\hat{Y} = \frac{N}{n}\sum_{i=1}^{n} \hat{Y}_i$$

Where:

$$\hat{Y}_1 = 28x\frac{6}{3} = 56 \text{ traps}$$

$$\hat{Y}_2 = 22x\frac{9}{3} = 66 \text{ traps}$$

$$\hat{Y}_3 = 33x\frac{7}{3} = 77 \text{ traps}$$

$$\sum_{i=1}^{3} \hat{Y}_i = 199 \text{ traps}$$

And:

$$\hat{Y} = \frac{8}{3}x199 = 531 \text{ traps}$$

Estimated variance of $\hat{Y}$:

$$v(\hat{Y}) = \frac{8x5}{3x2}x221+\frac{8}{3}\left[\{\frac{6x3}{3}x12.3\}+\{\frac{9x6}{3}x6.3\}+\{\frac{7x4}{3}x7.0\}\right] =$$

$$= 1473.3 \quad + \quad 673.3 \quad = 2146.6 \text{ traps}^2$$

$$(68.63\%)+(31.37\%) = (100\%)$$

$$cv(\hat{Y}) = \frac{\sqrt{2146.6}}{531}x100 = 8.73 \text{ percent}$$

15.8.2 Estimation in unequal probability sampling

In two-stage design another scheme of selection consists of selecting n fishing sites (PSU's) with replacement with probabilities p_i. An independent simple random sample of m_i fishing economic units is taken from every sample primary sample unit. We shall call this scheme two-stage sampling with unequal probabilities.

Within a selected PSU the estimated total of the survey characteristic is given by:

$$(94) \quad \hat{Y}_i = \frac{M_i}{m_i}\sum_{j=1}^{m_i} y_{ij}$$

In the present case an unbiased estimate of the population total is:

$$(95) \quad \hat{Y} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_i}{p_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\frac{M_i}{m_i}\sum_{j=1}^{m_i} y_{ij}$$

If:

$$t_i = \frac{1}{p_i}\frac{M_i}{m_i}\sum_{j=1}^{m_i} y_{ij}$$

Then the above estimator (95) can be written:

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

The estimated variance of the estimated total is:

$$v(\hat{Y}) = \frac{1}{n}\left(\sum_{i=1}^{n} t_i^2 - \frac{\left(\sum_{i=1}^{n} t_i\right)^2}{n}\right)$$

Example 19: To estimate the number of fishing operations (y) during the period t_o at a given inland water place a two-stage sampling scheme was adopted. A sample of fishing sites was selected with unequal probabilities (probabilities proportionate to the number of existing fishing boats) and within each sample fishing site a simple random sample of fishing economic units was taken. The table below gives the obtained sample data. Calculate $\hat{Y}$ and $cv(\hat{Y})$.

Sample fishing site	p_i	$\frac{M_i}{m_i}$	$\sum_{j=1}^{m_i} y_{ij}$	$\hat{Y}_i$	$t_i = \frac{\hat{Y}_i}{p_i}$	t_i^2 (000)	Remarks
1	$\frac{30}{1000}$	6	120	720	24000	576000	
2	$\frac{50}{1000}$	5	200	1000	20000	400000	
3	$\frac{10}{1000}$	2	110	220	22000	484000	
					66000 ($\sum_i t_i = t$)	1460000 ($\sum_i t_i^2$)	

Estimated total number of fishing operations during the period t_o:

$$\hat{Y} = \frac{1}{3}\sum_{i=1} t_i = \frac{66000}{3} = 22000 \text{ fishing operations}$$

Estimated variance of $\hat{Y}$:

$$v(\hat{Y}) = \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} t_i^2 - \frac{\left(\sum_{i=1}^{n} t_i\right)^2}{n}\right)$$

$$= \frac{1}{6}(1460000000-1452000000)$$

$$= 1333333 \text{ fishing operations}^2$$

Estimated coefficient of variation of $\hat{Y}$:

$$cv(\hat{Y}) = \frac{\sqrt{1333333}}{22000} x100 = 5.25 \text{ percent}$$

15.8.2.1 <u>Self-weighting system</u>

In practice the selection of SSU's within the sample PSU's is made by using the method of systematic sampling. Further, considerable simplification in the analysis of the data can be achieved when the sample selected is made self-weighting. This is done by adjusting the size of the overall sample of SSU's in such a way that the estimate can be obtained by multiplying the value of (y) in the sample by a known constant factor c. To give an illustration, suppose that the population consists of $M = \sum_{i=1}^{N} M_i$ secondary sampling units and we want to select a total sample of $m = \sum_{i=1}^{n} m_i$. In a self-weighting system the sampling interval $\frac{M_i}{m_i}$ within a sample primary sampling unit is given by:

$$(96) \quad \frac{M_i}{m_i} = cnp_i, \text{ if, for example } c = 50$$

Then:

$$\frac{M_i}{m_i} = 50\ np_i$$

In such a case the estimator for $\hat{Y}$ is modified as follows:

$$(97) \quad \hat{Y} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\frac{M_i}{m_i}\sum_{j=1}^{m_i} y_{ij} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{p_i}50np_i\right)\sum_{j=1}^{m_i} y_{ij}$$

$$= 50\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} = 50y$$

15.8.3 <u>Stratified two-stage sampling</u>[1/]

The theory discussed in the previous sections is applicable when PSU's are selected from a stratum. To get an estimate of population total, as well as variance, we simply add the independent estimates obtained within the established strata. Thus, in the case of sampling with unequal probabilities, the estimated population total is taken by:

1/ See also, Case Study, chapter 17.

$$\hat{Y} = \sum_{\ell=1}^{L}\left(\frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\frac{1}{p_{\ell i}}\frac{M_{\ell i}}{m_{\ell i}}\sum_{j=1}^{m_{\ell i}} y_{\ell i j}\right) = \sum_{\ell=1}^{L}\hat{Y}_\ell$$

Where the suffix ℓ stands for the strata ($\ell = 1, 2, \ldots, L$).

Also, the estimated variance of $\hat{Y}$ is:

$$v(\hat{Y}) = \sum_{\ell=1}^{L} v(\hat{Y}_\ell)$$

Where:

$$v(\hat{Y}_\ell) = \frac{1}{n_\ell(n_\ell-1)}\left(\sum_{i=1}^{n_\ell} t_{\ell i}^2 - \frac{\left(\sum_{i=1}^{n_\ell} t_{\ell i}\right)^2}{n_\ell}\right)$$

15.9 Area sampling

The first requirement for a probability sample of any nature is the establishment of the Sampling Frame. A Sampling Frame is a collection of sampling units which may be unambiguously defined and identified. For certain types of samples a complete and accurate list of the survey units covered by the survey is used as a Sampling Frame. For other samples, such a list may not exist or cannot be obtained inexpensively. An Area Sampling Frame is a geographic frame of well defined area units whereby any element has an association with the established area units. In an area sample the ultimate sampling units are the area units and the survey units can only be identified by geographic rules associating them with the sample area units. The method of area sampling is a necessity for the statistical surveys at inland water places which are characterized by a high level of peripheral mobility of the fishing economic units. For the selection of the sample units the following steps can be taken:

1. Up-dating of topographical sheets, preferably of scale 1:50000.
2. Proper stratification of the area covered by the survey population.
3. Proper delineation of the area units into the established strata.
4. Grouping the area units into minor strata by using auxiliary information (e.g. data of an aerial survey).
5. Selection of the sample of area units.
6. Selection of survey units within the sample area units.

16. SUMMATORS, EXPECTATION TECHNIQUES

16.1 One summator

The sum of n consecutive integral numbers can be written:

(1) $S = 1 + 2 + 3 + \ldots + n$

If suffix $i = 1, 2, 3 \ldots n$ then:

$$1 + 2 + 3 + \ldots + n = \sum_{i=1}^{i=n} i = \sum_{1}^{n} i$$

The expression $\sum_{i=1}^{i=n} i$ is called the "summator"

Exercises:

$$\sum_{0}^{1} i = 0 + 1 = 1$$

$$\sum_{1}^{2} i = 1 + 2 = 3$$

$$\sum_{1}^{3} i = 1 + 2 + 3 = 6$$

$$\sum_{1}^{4} i = 1 + 2 + 3 + 4 = 10$$

$$\sum_{1}^{5} i = 1 + 2 + 3 + 4 + 5 = 15$$

$$\sum_{0}^{6} i = 0 + 1 + 2 + 3 + 4 + 5 + 6 = 21$$

$$\sum_{-1}^{0} i = (-1) + 0 = -1$$

$$\sum_{-1}^{1} i = (-1) + 0 + 1 = 0$$

$$\sum_{-2}^{1} i = (-2) + (-1) + 0 + 1 = -2$$

$$\sum_{-3}^{2} i = (-3) + (-2) + (-1) + 0 + 1 + 2 = -3$$

$$\sum_{4}^{8} i = 4 + 5 + 6 + 7 + 8 = 30$$

Prove:

$$\sum_{1}^{n} i = \sum_{n}^{1} i$$

If we multiply each term of (1) by c then:

$$S' = c\times 1 + c\times 2 + c\times 3 + \ldots + c\times n$$

Or:

$$c\times 1 + c\times 2 + c\times 3 + \ldots + c\times n = \sum_{i=1}^{i=n} ci = \sum_{1}^{n} ci$$

Exercises:

$$\sum_{1}^{3} 3i = 3\times 1 + 3\times 2 + 3\times 3 = 18$$

$$\sum_{0}^{3} 3i = 3\times 0 + 3\times 1 + 3\times 2 + 3\times 3 = 18$$

$$\sum_{-1}^{1} 3i = 3\times(-1) + 3\times 0 + 3\times 1 = 0$$

$$\sum_{1}^{4} \frac{i}{2} = \frac{1}{2} + \frac{2}{2} + \frac{3}{2} + \frac{4}{2} = \frac{1 + 2 + 3 + 4}{2} = 5$$

$$\sum_{1}^{3} \frac{2i}{3} = \frac{2\times 1}{3} + \frac{2\times 2}{3} + \frac{2\times 3}{3} = \frac{2 + 4 + 6}{3} = 4$$

$$\sum_{0}^{3} (-5)i = (-5)\times 0 + (-5)\times 1 + (-5)\times 2 + (-5)\times 3 = -30$$

$$\sum_{-2}^{3} (-4)i = (-4)\times(-2) + (-4)\times(-1) + (-4)\times 0 + (-4)\times 1 + (-4)\times 2 + (-4)\times 3 = -12$$

$$\sum_{1}^{4} (-2)^i = (-2)^1 + (-2)^2 + (-2)^3 + (-2)^4 = -2 + 4 - 8 + 16 = 10$$

Prove:

1. $\sum_{1}^{n} ci = c \sum_{1}^{n} i$

2. $\sum_{1}^{n} (i+c) = \sum_{1}^{n} i + nc$

3. $\sum_{1}^{n} (ci+b) = c\sum_{1}^{n} i + nb$

16.1.1 Two summators

Let us assume that the suffix i takes the values:

$i = 1, 2, 3, \ldots, n$

and the suffix j takes the values:

$j = 1, 2, 3, \ldots, m$

The double summator $\sum_{i=1}^{2} \sum_{j=3}^{6} (i+j)$ is:

$$\sum_{i=1}^{2} \sum_{j=3}^{6} (i+j) = (1+3) + (1+4) + (1+5) + (1+6) + (2+3) + (2+4) + (2+5) + (2+6) = 54$$

Generally the double summator:

$$\begin{aligned} \sum_{i=1}^{n} \sum_{j=1}^{m} (i+j) = {} & (1+1) + (1+2) + (1+3) + \ldots + (1+m) \\ & (2+1) + (2+2) + (2+3) + \ldots + (2+m) \\ & (3+1) + (3+2) + (3+3) + \ldots + (3+m) \\ & \quad " \qquad\quad " \qquad\quad " \qquad\qquad\quad " \\ & \quad " \qquad\quad " \qquad\quad " \qquad\qquad\quad " \\ & (n+1) + (n+2) + (n+3) + \ldots + (n+m) \end{aligned}$$

Another presentation of the double summator $\sum_{i=1}^{i=n} \sum_{j=1}^{j=m} (i+j)$ is given in the table below:

j→ / i↓	j=1	j=2	j=3	.	.	.	.	.	.	j=m	Marginal total
i=1	1+1 =2	1+2 =3	1+3 =4	.	.	.	.	.	.	1+m	$T_{\cdot 1}$
i=2	2+1 =3	2+2 =4	2+3 =5	.	.	.	.	.	.	2+m	$T_{\cdot 2}$
i=3	3+1 =4	3+2 =5	3+3 =6	.	.	.	.	.	.	3+m	$T_{\cdot 3}$
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
i=n	n+1	n+2	n+3	.	.	.	.	.	.	n+m	$T_{\cdot n}$
Marginal total	$T_{1\cdot}$	$T_{2\cdot}$	$T_{3\cdot}$	.	.	.	.	.	.	$T_{m\cdot}$	General Total T

<u>Prove</u>:

1. $$\sum_{i=1}^{i=n} \sum_{j=1}^{j=m} (i+j) = m \sum_{i=1}^{i=n} i + n \sum_{j=1}^{j=m} j$$

2. $$\sum_{i=1}^{i=n} \sum_{j=1}^{j=m} (ci+bj) = cm \sum_{i=1}^{i=n} i + bn \sum_{j=1}^{j=m} j$$

The double summator $\sum_{i=1}^{i=2} \sum_{j=3}^{j=6} ij$ is:

$$\sum_{i=1}^{i=2} \sum_{j=3}^{j=6} ij = (1\times3) + (1\times4) + (1\times5) + (1\times6) + (2\times3) + (2\times4) + (2\times5) + (2\times6) = 54$$

Generally the double summator:

$$\begin{aligned}\sum_{i=1}^{i=n} \sum_{j=1}^{j=m} ij = {} & (1\times1) + (1\times2) + (1\times3) + \ldots + (1\times m) \\ & (2\times1) + (2\times2) + (2\times3) + \ldots + (2\times m) \\ & (3\times1) + (3\times2) + (3\times3) + \ldots + (3\times m) \\ & \quad " \qquad " \qquad " \qquad\qquad " \\ & \quad " \qquad " \qquad " \qquad\qquad " \\ & (n\times1) + (n\times2) + (n\times3) + \ldots + (n\times m)\end{aligned}$$

Another presentation of the double summator $\sum_{i=1}^{i=n} \sum_{j=1}^{j=m} ij$ is given in the table below:

j→ / i↓	j=1	j=2	j=3	.	.	.	.	.	.	j=m	Marginal total
i=1	1×1 =1	1×2 =2	1×3 =3	.	.	.	.	.	.	1×m	$T_{\cdot 1}$
i=2	2×1 =2	2×2 =4	2×3 =6	.	.	.	.	.	.	2×m	$T_{\cdot 2}$
i=3	3×1 =3	3×2 =6	3×3 =9	.	.	.	.	.	.	3×m	$T_{\cdot 3}$
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
i=n	n×1	n×2	n×3	.	.	.	.	.	.	n×m	$T_{\cdot n}$
Marginal total	$T_{1\cdot}$	$T_{2\cdot}$	$T_{3\cdot}$	.	.	.	.	.	.	$T_{m\cdot}$	General Total T

Prove:

1. $$\sum_{i=1}^{i=n} \sum_{j=1}^{j=m} ij = \sum_{i=1}^{i=n} i \sum_{j=1}^{j=m} j$$

2. $$\sum_{i=1}^{i=n} \sum_{j=1}^{j=m} (ci)(bj) = \left(c \sum_{i=1}^{i=n} i\right)\left(b \sum_{j=1}^{j=m} j\right)$$

General exercises

Calculate the following sums:

1. $\sum_{1}^{3} i + \sum_{2}^{5} i$

2. $\sum_{1}^{3} i^2$

3. $\sum_{2}^{5} \frac{2i}{5}$

4. $\sum_{1}^{4} \frac{4i^2}{10}$

5. $\sum_{1}^{2} (3i^2+2i)$

6. $\sum_{1}^{3} i-1$

7. $\sum_{2}^{4} (3+\frac{4}{i})$

8. $\sum_{i=1}^{i=2} \sum_{j=3}^{j=4} (i+j)$

9. $\sum_{i=1}^{i=2} \sum_{j=3}^{j=4} ij$

10. $\sum_{i=4}^{i=5} \sum_{j=1}^{j=2} (i-j)$

11. $\sum_{i=-2}^{i=2} \sum_{j=-1}^{j=1} (3i+4j)$

12. $\sum_{i=1}^{i=2} \sum_{j=1}^{j=2} \frac{i+j}{ij}$

13. $\sum_{i=3}^{i=4} \sum_{j=1}^{j=2} \frac{i-j}{i+j}$

14. $\sum_{i=0}^{i=2} \sum_{j=3}^{j=5} \frac{(2i+3j)}{(4i+5j)}$

16.2 The use of summators in statistics

16.2.1 The sum of the numerical values of variables

Assuming that the numerical values (real values) of a variable (x) are:

$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}$

If (x) expresses total monthly fish catch of a Fishing Economic Unit (FEU) at Lake Tanganyika, then the yearly total catches of the FEU is given by:

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12}$$

By using summators the above sum can be written:

$$\sum_{i=1}^{i=12} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12}$$

Where:

$i = 1, 2, 3, \ldots, 12$

(the suffix, i, stands for the twelve months of the given year). For simplicity:

$$\sum_{i=1}^{i=12} x_i = \sum_{1}^{12} x_i$$

In general, the sum of n values of x is:

$$\sum_{i=1}^{i=n} x_i = \sum_{1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

Where:

$i = 1, 2, 3, \ldots, n$

Exercises:

1. $\sum_{0}^{2} x_i = x_0 + x_1 + x_2$

2. $\sum_{2}^{5} x_i = x_2 + x_3 + x_4 + x_5$

3. $\sum_{-3}^{3} x_i = x_{-3} + x_{-2} + x_{-1} + x_0 + x_1 + x_2 + x_3$

Prove:

1. $\sum_{1}^{n} x_i = \sum_{n}^{1} x_i$

2. $\sum_{1}^{n} x_i = \sum_{1}^{h} x_i + \sum_{h+1}^{n} x_i, \ (h<n)$

3. $\sum_{1}^{n} cx_i = c\sum_{1}^{n} x_i, \ (c = \text{constant})$

4. $\sum_{1}^{n} (c+x_i) = nc + \sum_{1}^{n} x_i$

5. $\sum_{1}^{n} (c+bx_i) = nc + b\sum_{1}^{n} x_i$

6. $\sum_{1}^{n} (x_i+y_i) = \sum_{1}^{n} x_i + \sum_{1}^{n} y_i$

7. $\sum_{1}^{n} (cx_i+by_i) = c\sum_{1}^{n} x_i + b\sum_{1}^{n} y_i$

8. $\sum_{1}^{n} (x_i+y_i+z_i) = \sum_{1}^{n} x_i + \sum_{1}^{n} y_i + \sum_{1}^{n} z_i$

9. $\sum_{1}^{n} (cx_i+by_i+az_i) = c\sum_{1}^{n} x_i + b\sum_{1}^{n} y_i + a\sum_{1}^{n} z_i$

10. $\sum_{1}^{n} x_i^2 = x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2$

11. $\sum_{1}^{n} x_i^\mu = x_1^\mu + x_2^\mu + x_3^\mu + \ldots + x_n^\mu$

12. $\sum^{n} (x_i-y_i) = (x_1-y_1) + (x_2-y_2) + (x_3-y_3) + \ldots + (x_n-y_n)$

13. $\sum_{1}^{n} x_i y_i = x_1y_1 + x_2y_2 + x_3y_3 + \ldots + x_ny_n$

In certain statistical problems summators are used to express the sum of terms with two suffixes, i, j:

a: The suffix, i, is a constant and suffix, j, is a variable:

i) If i=1 we have,

$$\sum_{j=1}^{m} x_{1j} = x_{11} + x_{12} + x_{13} + \dots + x_{1m}$$

ii) If i=2 we have,

$$\sum_{j=1}^{m} x_{2j} = x_{21} + x_{22} + x_{23} + \dots + x_{2m}$$

etc.

b: The suffix, j, is a constant and suffix, i, is a variable:

i) If j=1 we have,

$$\sum_{i=1}^{n} x_{i1} = x_{11} + x_{21} + x_{31} + \dots + x_{n1}$$

ii) If j=2 we have,

$$\sum_{i=1}^{n} x_{i2} = x_{12} + x_{22} + x_{32} + \dots + x_{n2}$$

etc.

c: The suffixes, i, and, j, are variables:

$$\begin{aligned}\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} = {} & x_{11} + x_{12} + x_{13} + \dots + x_{1m} \\ & + x_{21} + x_{22} + x_{23} + \dots + x_{2m} \\ & + x_{31} + x_{32} + x_{33} + \dots + x_{3m} \\ & \quad " \qquad " \qquad " \qquad " \qquad " \\ & \quad " \qquad " \qquad " \qquad " \qquad " \\ & \quad " \qquad " \qquad " \qquad " \qquad " \\ & + x_{n1} + x_{n2} + x_{n3} + \dots + x_{nm}\end{aligned}$$

<u>Prove</u>:

1. $$\sum_{i=1}^{n} \sum_{j=1}^{m} x_i x_j = \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{m} x_j\right)$$

2. $$\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j = \left(\sum_{i=1}^{n} x_i\right)^2$$

3. $$\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{l} x_i x_j x_k = \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{m} x_j\right)\left(\sum_{k=1}^{l} x_k\right.$$

4. $$\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} x_i x_j x_k = \left(\sum_{i=1}^{n} x_i\right)^3$$

16.3 Expectation techniques

We assume a discontinuous random variable, x, with:

(1) $P(x = x_i) = p_i$

The expectation of the random variable is defined as:

$$E(x) = \sum_i p_i x_i$$

where the sum is taken over all possible values the random variable may take. The expectation of any function of, x, say f(x), is defined as:

(2) $$Ef(x) = E\{f(x)\} = \sum_i p_i f(x_i)$$

If x is a characteristic random variable taking values 0 and 1 with probabilities p and q respectively, the E(x) equals:

(3) $$E(x) = 1 \times p + 0 \times q = p$$

(4) $$E(x^k) = 1^k \times p + 1^k \times 0 = p$$

or the expected value of any power of the characteristic random variable is the same as the expected value of the random variable.

Let x be a random variable which may take values, $x_1, x_2, \ldots x_i, \ldots x_n$ with probabilities $p_1, p_2, \ldots p_n$, and let y be another variable which may take values $y_1, y_2, \ldots y_m$. The values of x and y are mutually exclusive and the only possible ones. Further, let P_{ij} be the joint probability that x takes the value x_i and y the value y_j. The expected value of the sum of the two variables is given by:

(5) $$E(x+y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i+y_j)P_{ij}$$

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} P_{ij} + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} P_{ij}$$

$$= \sum_{i=1}^{n} x_i P_i + \sum_{j=1}^{m} y_j P_j = E(x) + E(y)$$

Generally if we have a set of discontinuous random variables then:

(6) $$E\left[\sum_{i=1}^{k} x_i\right] = \sum_{i=1}^{k} E(x_i)$$

The above equality (6) implies that the expectation of the sum of a number of random variables is equal to the sum of their separate expectations. Equation (6) stands either for independent or non-independent random variables.

For the same two random variables the expected value of the product of the variables is given by:

(7) $$E(xy) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i y_j)P_{ij}$$

$$= \sum_{i=1}^{n} x_i P_i \sum_{j=1}^{m} y_j \; P\left[(y=y_j)\middle|(x=x_i)\right]$$

If x and y are independent then:

$$(8)\quad E(xy) = \sum_{i=1}^{n} x_i P_i \sum_{j=1}^{m} x_j P_j = E(x)E(y)$$

Generally if we have a set of k mutually independent variables then:

$$(9)\quad E \prod_{i=1}^{k} x_i = \prod_{i=1}^{k} E(x_i)$$

The above equality (9) implies that the expectation of the product of a number of mutually independent variables is equal to the product of their expectations.

16.3.1 Expectation of some statistical functions

Where s_x^2 sample variance, and σ_x^2 population variance, prove that $E(s_x^2) = \sigma_x^2$:

$$E(s_x^2) = \frac{1}{n-1} E\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (x_i - m + m - \bar{x})^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} \{(x_i - m) - (\bar{x} - m)\}^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (x_i - m)^2 + \sum_{i=1}^{n} (\bar{x} - m)^2 - 2(\bar{x} - m) \sum_{i=1}^{n} (x_i - m)\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (x_i - m)^2 - n(\bar{x} - m)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(x_i - m)^2 - nE(\bar{x} - m)^2\right]$$

$$= \frac{1}{n-1}(n\sigma_x^2 - n\sigma_{\bar{x}}^2) \quad , \quad (\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n})$$

$$= \frac{1}{n-1}(n\sigma_x^2 - \sigma_x^2)$$

$$= \sigma_x^2 \frac{n}{n-1}(1 - \frac{1}{n})$$

$$= \sigma_x^2 \frac{n}{n-1} \times \frac{n-1}{n} = \sigma_x^2$$

Prove that $V(ax) = a^2V(x)$:

$$V(ax) = V(x'), \quad x' = ax \text{ and } a = \text{constant}$$
$$= E(x'^2)-\{E(x')\}^2$$
$$= E(ax)^2-\{E(ax)\}^2$$
$$= a^2\left[E(x)^2-\{E(x)\}^2\right]$$
$$= a^2V(x)$$

Note: the operator, V, stands for Variance

Prove that $V(ax+b) = a^2V(x)$:

$$V(ax+b) = V(x"), \quad x" = ax+b \text{ where } a, b = \text{constants}$$
$$= E(x")^2-\{E(x")\}^2$$
$$= E(ax+b)^2-\{E(ax+b)\}^2$$
$$= \left[a^2E(x^2)+2abE(x)+b^2\right]-\left[a^2\{E(x)\}^2+2abE(x)+b^2\right]$$
$$= a^2\left[E(x^2)-\{E(x)\}^2\right]$$
$$= a^2V(x)$$

Prove that, if $t = \frac{x-m}{\sigma_x}$, then $V(t) = 1$:

$$V(t) = E(t^2)-\{E(t)\}^2$$
$$= E\left[\frac{x-m}{\sigma_x}\right]^2-\left[E\{\frac{x-m}{\sigma_x}\}\right]^2$$
$$= \frac{1}{\sigma_x^2}\left[E(x^2)-2mE(x)+m^2\right]-\frac{1}{\sigma_x^2}\left[\{E(x)\}^2-2mE(x)+m^2\right]$$
$$= \frac{1}{\sigma_x^2}\left[E(x2)-\{E(x)\}^2\right] = \frac{\sigma_x^2}{\sigma_x^2} = 1$$

Prove that $E(t) = 0$:

$$E(t) = E\left(\frac{x-m}{\sigma_x}\right)$$
$$= \frac{1}{\sigma_x}\{E(x)-m\} = \frac{0}{\sigma_x} = 0$$

Prove that if y, x, are two independent random variables, $V(x+y) = V(x)+V(y)$:

$$V(x+y) = E(x+y)^2-\{E(x+y)\}^2$$
$$= \left[E(x^2)+2E(xy)+E(y^2)\right]-\left[\{E(x)\}^2+2E(x)E(y)+\{E(y)\}^2\right]$$
$$= \left[E(x^2)-\{E(x)\}^2\right]+\left[E(y^2)-\{E(y)\}^2\right]$$
$$= V(x)+V(y)$$

For the same two variables (x, y) prove that $V(x-y) = V(x)+V(y)$:

$$V(x-y) = \left[E(x-y)^2-\{E(x-y)\}^2\right]$$

$$= \left[E(x^2)-2E(xy)+E(y^2)\right]-\left[\{E(x)\}^2-2E(x)E(y)+\{E(y)\}^2\right]$$

$$= \left[E(x^2)-\{E(x)\}^2\right]+\left[E(y^2)-\{E(y)\}^2\right]$$

$$= V(x)+V(y)$$

If x and y are non-independent variables, prove that $V(x+y) = V(x)+V(y)+2Cov.(x, y)$ (Cov. = Covariance):

$$V(x+y) = \left[E(x^2)+2E(xy)+E(y^2)\right]-\left[\{E(x)\}^2+2E(x)E(y)+\{E(y)\}^2\right]$$

$$= \left[E(x^2)-\{E(x)\}^2\right]+\left[E(y^2)-\{E(y)\}^2\right]+2\left[E(xy)-E(x)E(y)\right]$$

$$= V(x)+V(y)+2\ Cov.(x, y)$$

where $Cov.(x, y) = E(xy)-E(x)E(y) = E(x-m_x)(y-m_y)$

Prove that for the same two variables x, y (as above) $V(x-y) = V(x)+V(y)-2Cov.(x, y)$.

Prove that ρ (coefficient of correlation) is given by:

$$\rho_{xy} = cov.(t_1, t_2)$$

Where:

$$t_1 = \frac{x-mx}{\sigma_x}, \quad t_2 = \frac{y-my}{\sigma_y}$$

Prove that $cov.(x, y) = \rho_{xy}\cdot\sigma_x\cdot\sigma_y$

Prove that $cov.(\bar{x}, \bar{y}) = (\frac{1}{n}-\frac{1}{N})\ Cov.(x, y)$

Where $\bar{x}$, $\bar{y}$, are sample means.

APPENDIX IIIa - TABLE OF RANDOM NUMBERS

1	2	3	4	5	6	7	8	9	10	11	12
13	70	43	69	38	81	87	42	12	20	41	15
26	99	82	78	99	05	22	99	52	32	80	91
72	53	95	81	07	98	14	74	52	58	73	10
22	08	08	68	37	16	36	62	20	02	35	98
21	61	90	53	85	72	86	94	87	18	50	11
47	38	55	66	50	96	96	78	34	45	52	78
96	68	13	07	31	29	70	09	16	66	81	09
45	92	93	44	87	72	26	75	82	31	72	69
78	85	71	45	32	16	57	91	52	05	93	20
51	99	50	88	62	54	90	51	01	39	18	70
67	62	30	02	88	17	37	25	42	86	00	32
03	08	89	77	12	41	15	25	52	30	93	11
45	10	04	66	94	70	33	74	97	23	40	97
62	48	46	97	04	36	31	27	29	84	85	35
59	59	33	63	53	43	60	30	15	81	67	59
72	63	67	17	24	55	68	32	24	80	13	92
46	28	15	70	28	98	53	36	03	89	83	74
21	03	09	16	31	48	05	10	98	62	14	15
84	82	53	39	92	14	07	84	04	01	66	17
75	68	40	90	39	95	46	10	94	68	39	10
42	77	29	80	73	38	92	11	81	72	50	88
63	55	09	84	66	56	92	13	97	14	87	27
54	29	70	14	85	95	79	72	77	48	57	92
42	97	50	61	19	55	38	55	85	57	85	08
52	30	47	73	26	54	18	05	75	92	95	08
88	44	33	02	47	97	47	04	12	38	93	25
49	91	93	73	14	15	01	47	02	70	30	96
45	42	46	06	93	60	41	09	31	29	52	49
50	69	74	10	51	89	66	51	57	21	54	95
18	56	73	16	02	87	41	05	13	87	13	61

13	14	15	16	17	18	19	20	21	22	23	24
76	96	85	27	81	21	75	39	43	77	80	81
38	51	09	17	41	85	13	20	66	59	22	20
40	91	90	51	74	23	54	88	84	12	16	77
44	53	23	87	91	53	86	97	42	80	83	37
31	25	22	30	16	17	32	34	00	07	25	52
36	35	20	92	81	12	15	28	42	98	67	52
36	12	17	03	83	93	48	64	50	32	57	94
25	51	40	74	85	16	86	09	22	62	06	38
72	38	33	97	36	58	90	91	23	91	19	04
17	20	75	03	85	53	06	41	29	78	51	15
75	57	37	77	67	60	70	44	56	91	03	49
12	47	35	37	15	17	96	24	95	08	39	55
73	67	55	64	16	38	58	74	29	71	49	62
16	02	29	14	16	78	44	49	34	05	46	96
48	98	13	29	19	71	98	71	19	51	86	82
73	65	42	09	39	92	56	68	36	54	55	46
22	96	06	41	55	75	08	62	55	19	15	75
57	26	11	28	98	16	85	39	67	49	02	30
47	76	60	92	22	79	70	66	78	13	97	42
31	80	30	86	08	54	39	88	38	46	74	21
91	55	48	36	26	40	17	70	39	94	05	76
83	70	10	91	20	64	12	33	15	59	43	28
28	35	53	14	30	57	07	34	09	56	26	81
86	91	62	94	83	96	96	17	02	10	89	71
24	86	86	52	67	59	63	22	28	76	43	45
43	73	70	73	19	41	04	60	25	42	09	50
52	69	34	01	65	33	19	62	22	41	29	65
01	15	92	69	53	78	68	58	74	08	05	11
94	46	83	72	49	19	98	09	56	83	25	40
44	42	06	32	95	17	32	67	80	84	09	69
81	58	85	33	16	11	87	12	17	39	12	11
60	25	84	42	22	94	38	96	52	03	38	97
53	12	75	59	76	42	73	48	95	57	51	31
02	68	01	17	09	00	38	12	31	52	22	24
09	68	53	92	82	11	96	03	47	31	35	59

17. A CASE STUDY : THE SAMPLING DESIGN OF THE CATCH ASSESSMENT SURVEY (CAS) AT LAKE TANGANYIKA (TANZANIA)1/

17.1 Purpose of the survey

The Catch Assessment Survey at Lake Tanganyika (Tanzania) is a probability sample survey, conducted on a lunar month basis by the Lake Tanganyika Fisheries Project (Tanzania). The primary objective of the survey is to obtain reliable current estimates on a regional basis, and for the Tanzanian part of Lake Tanganyika, of the total quantity of fish harvested by the fishermen at the lake (in terms of live weight in metric tons). Secondary objectives include the species composition of fish catch and the fishing effort involved in obtaining the catch, from which the estimate "catch per unit of effort" can be obtained. This type of information can be used, among other things, in determining the management practices that might be necessary in the future and provides a base line on which these practices can be rationally evaluated.

17.2 The sampling method of the survey

The sampling method used for the CAS can be described as "sampling in space and time". For space, the area sampling method on a stratified multistage basis with unequal probabilities (pps) was used. As far as time is concerned the method of the stratified random sample was applied.

17.3 Sampling in space

For the stratification of Lake Tanganyika (Tanzania) limnological information and data concerning the area distribution of the fishing industry were used as control characteristics of stratification. Specifically, the lake was divided first into seven areas here called "strata". In order to take full advantage of possible gains from area stratification the sample design adopted called for a further stratification of the surveyed population. Specifically, each stratum was divided into a number of area zones, here called minor-strata, by taking into account the level of localization of the fishing industry.

17.3.1 Sampling units

A sampling procedure presupposes the division of the surveyed population into a finite number of distinct and identifiable units called the sampling units. Specifically, the smallest units into which a population can be divided are called the elements of the population (survey units), and groups of elements are called clusters. In our case the Fishing Economic Unit (FEU) was taken as the survey unit or reporting unit and the Fishing Site as a cluster unit or Primary Sampling Unit (PSU). An FEU is an integral unit composed of fishing boat, fishing gear and fisherman(men) to carry out fishing operations.

17.3.2 The sample of Primary Sampling Units (PSU's)

For sampling purposes, a number of fishing sites have been selected within each established minor-stratum. Specifically, the sampling design called for the selection of two PSU's within each minor-stratum with probabilities proportionate to the number of fishing boats (inter-penetrating sub-sampling method with unequal probabilities). For the survey, the selected sample of PSU's was kept fixed over time. In the table below (Table 17.3.2.1) the selected sample of PSU's is given.

1/ A report prepared for the Lake Tanganyika Fishery Research Project, by G.P. Bazigos FI:DP/URT/71/012/1 December 1973, Rome, Italy.

Table 17.3.2.1 Selected sample of PSU's

Stratum : 1 - Selected fishing sites/CAS

C.No.	Name	Location (see map)
111 112	Makombe Ngonya Ntambala (R)	 (Reserved)
121 122	Makasa Kitwe Bugamba (R)	 (Reserved)
131 132	Ngeru Kalalangobo Kogongo (R)	 (Reserved)
141 142	Katonga Kasaba Kampande (R)	 (Reserved)
151 152	Mwakizega Kasigo Mchangani (R)	 (Reserved)

Sizes:		
	M-Str:11	690c
	M-Str:12	599c
	M-Str:13	584c
	M-Str:14	439c
	M-Str:15	493c
	Total	2805c

Stratum : 2 - Selected fishing sites/CAS

C.No.	Name	Location (see map)
211 212	Kite Kirando Lugufa (R)	 (Reserved)
221 222	Ngangasima Kangweno-II Mkuyu (R)	 (Reserved)
231 232	Helembe Ngondozi Camp-II Kashe (R)	 (Reserved)

Sizes:		
	M-Str:21	629c
	M-Str:22	587c
	M-Str:23	721c
	Total	1937c

Stratum : 3 - Selected fishing sites/CAS

C.No.	Name	Location (see map)
311 312	Makola Lufubu Kalya (R)	 (Reserved)

Sizes: M-Str:31 258c

Stratum : 4 - Selected fishing sites/CAS

C.No.	Name	Location (see map)
411 412	Ikola Karema Kasanga (R)	 (Reserved)

Sizes: M-Str:41 140c

Stratum : 5 - Selected fishing sites/CAS

C.No.	Name	Location (see map)
511 512	Mkombe-I Kambwe-I Katoba (R)	 (Reserved)
521 522	Shashete-I Chongo Katete (R)	 (Reserved)
531 532	Kipili Uwile (ISL) Mtakuja (R)	 (Reserved)

Sizes: M-Str:51 252c
M-Str:52 360c
M-Str:53 393c

Total 1005c

Stratum : 6 - Selected fishing sites/CAS

C.No.	Name	Location (see map)
611 612	Katale Chombo Kisambala-I (R)	 (Reserved)
621 626	Msamba Wampembe Kizumbi (R)	 (Reserved)

Sizes: M-Str:61 227c
M-Str:62 349c

Total 576c

Stratum : 7 - Selected fishing sites/CAS

C.No.	Name	Location (see map)
711 712	Chowe Kilambo Tundu (R)	 (Reserved)
721 722	Kipanga Kasanga Musi (R)	 (Reserved)

Sizes: M-Str:71 308c
M-Str:72 367c

Total 675c

Estimated size of the fishing industry by Minor-stratum
(Frame Survey - Jun/Jul 1973, map-data)

Stratum	Min-str	Number of fishing boats	Remarks
		Total 7396	
Str:1		2805	
	1.1	690	
	1.2	599	
	1.3	584	
	1.4	439	
	1.5	493	
Str:2		1937	
	2.1	629	
	2.2	587	
	2.3	721	
Str:3		258	
	3.1	258	
Str:4		140	
	4.1	140	
Str:5		1005	
	5.1	252	
	5.2	360	
	5.3	393	
Str:6		576	
	6.1	227	
	6.2	349	
Str:7		675	
	7.1	308	
	7.2	367	

17.3.3 The sample of Fishing Economic Units (FEU's)

From a statistical point of view a fishing site (PSU) can be considered as a compound area unit consisting of two parts, the residential area of the fishing site where the headquarters of the FEU's are located and the beach of the fishing site where the producing FEU's are located. According to the established "survey method" of the survey, within each selected PSU information on the static characteristics of the FEU's i.e. items of information on the components of the FEU's, are collected by using the census approach, whereas items of information on the dynamic characteristics of the FEU's i.e. input and output data of the fishing operations of the units, are obtained by using the sampling approach. For the selection of the samples of FEU's within the sample fishing sites, the method of systematic sampling with a random starting point is used.

17.4 Sampling in time

One of the purposes of the CAS is to provide reliable estimates of the trends describing the yield seasonality pattern at the lake. By taking into account the type of fishery at the lake it was decided that the "reference period" of the survey characteristics of the CAS would be a lunar month (typical period).

Within each typical period the selected PSU's (fixed sample) are randomly allocated in survey periods (one survey period covers four days); the obtained sample data of the survey within a typical period are used to provide estimates on a lunar month basis. Annual estimates are calculated by adding up the monthly estimates.

17.5 The survey period of the Catch Assessment Survey (CAS)

For the CAS the length of the survey period was determined by taking into account the sampling error attributed to day-by-day variations of surveyed characteristics. It was decided that the optimum survey period of the CAS is four consecutive days. Specifically, the first day is used to set up the sampling frame of the existing FEU's within the sample fishing sites. Items of information of the survey characteristics are collected within each of the remaining three days.

17.6 Survey operations

The description of the survey operations provides an indication of the linkage between the sample and survey design and the actual collection and processing of the data. In this section a summary is given of the field operation of the survey and data processing.

17.6.1 Field personnel

An intensive training course for field recorders, lasting for more than two weeks, was held at Kigoma. The course was also attended by Regional Officers. An objective evaluation of the trainees was made through a series of exercises, discussions on methodological problems and a critical analysis of the results of the conducted Mini Pilot Survey.

For the field operations of the CAS seven working groups (WG's) were formed. Each WG consisted of a statistical recorder and an assistant. One WG was assigned to each Stratum.

For the transportation of the field personnel between the selected fishing sites two boats are going to be used. One stationed in Kigoma (Project boat) and the other one in Stratum 5 of the survey.

17.6.2 Field operations - control

The best assurance of accurate field work is that the statistical recorders are well trained and are capable, conscientious and keen. Nevertheless, it is important even with good recorders to keep a close watch on the progress of the work.

In the CAS the main ingredients of the field supervision as far as control is concerned can be described as follows:

1. Field editing: The primary purpose of the field editing, carried out by the supervisors of the survey, is to catch omissions, inconsistencies, illegible entries and errors, before schedules are sent to Kigoma Office (HQ) and correction is still possible.

2. The supervisors must also check the quality of the work of the recorders. Within a lunar month each supervisor is required to observe the work of the recorders working in the Supervision Area for which he is responsible. During the period the supervisor accompanies a recorder, he observes how well the recorder "sells" the survey, how he checks the coverage of the survey units, how he selects the samples of the FEU's, whether he asks questions properly, whether he conducts measurements properly, and how well he conducts himself generally.

3. Every month the supervisor will give a report to each recorder of any errors detected in the course of reviewing his work in the office. As required, the report will specify what special steps the recorder should take to avoid making similar errors in the future.

17.6.3 Source documents

For the collection of the information at the selected fishing sites four forms are used, Form: Bo, B1, A1, A2. Specifically, the purpose of Form:Bo is to set up the sampling frame of the existing FEU's at the sample fishing sites and collect items of information on the components of the FEU's. The form is also used for the classification of the existing FEU's into groups according to the fishing method used.

Form: B1 is used to select the samples of active FEU's at the sample fishing sites. Specifically, a sample of FEU's is selected in each survey day within the established fishing methods.

Form: A1 is used to collect items of information on fishing effort and fish catch from the selected landings using Lusenga, Liftnets or Beach seine net for dagaa as a fishing method.

Form: A2 is used to collect items of information on fishing effort and fish catch from the selected landings using Gillnet, Beach seine net for fish or Handlines as a fishing method. In Appendix IIIb the format of the forms used for the CAS are given.

17.6.4 Processing operations

The processing of the material and the highly skilled task of analyzing begins soon after the completion of the field operations of the survey (on a lunar month basis). Before the questionnaires can be regarded as ready for tabulation they must be checked once more by the supervisors for completeness, accuracy and conformity. At the same time the quality of coding must be checked by the supervisors.

It has been decided that tabulation be done in two stages. In the first stage the process is done manually. For final presentation within a month a number of basic tables are constructed providing estimates for the most important characteristics of the survey.

In the second stage machine tabulation will be used. The computerization scheme will be discussed in another report.

In Appendix IIIc the working sheets which will be used for the manual processing of the results of the CAS are given. Also in Appendix IIIc the instructions for the completion of the working sheets are given.

APPENDIX IIIb - FORMS USED FOR THE CATCH ASSESSMENT SURVEY (CAS)

LAKE TANGANYIKA FISHERIES RESEARCH AND DEVELOPMENT PROJECT (TANZANIA)
CATCH ASSESSMENT SURVEY (CAS)

FORM: Bo(1.0)

Page ____

LIST OF EXISTING FISHING ECONOMIC UNITS

Name of the Recorder ____________ ☐

Survey date ☐☐☐

Sample fishing site: C.No. ☐ Name ____________

FEU's	FISHING METHOD		FISHING BOAT		CREW					Remarks
	Prim	Other	Type	Reg. No.	Total	Boat owner	Crew leader	Assis-tants	Other (specify)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)

1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

LAKE TANGANYIKA FISHERIES RESEARCH AND DEVELOPMENT PROJECT (TANZANIA) CATCH ASSESSMENT SURVEY (CAS)

FORM: B1 (1.1)

SELECTION OF FEU'S

Name of the Recorder ____________________ ☐

Landing date ☐☐☐

Sample fishing site: C.No ☐ Name ____________

Primary Fishing Method: C. No.	LUSENGA (1)		GILLNET (2)		BEACH-DAGAA (3)		BEACH-FISH (4)		LIFTNET (5)		HANDLINES (6)		Remarks
1. No. of existing FEU's													
2. No. of inactive FEU's													
3.(1-2) No.of active FEU's													
	Ser.No.		Ser.No.		Ser.No.		Ser.No.		Ser.No.		Ser.No.		
	Pop. (1)	Sam. (2)	Pop. (1)	Sam (2)	Pop. (1)	Sam. (2)	Pop. (1)	Sam (2)	Pop. (1)	Sam (2)	Pop. (1)	Sam. (2)	

LAKE TANGANYIKA FISHERIES RESEARCH AND DEVELOPMENT PROJECT (TANZANIA)
CATCH ASSESSMENT SURVEY (CAS)

FORM: A1 (1.2)

CATCH RECORDS
(LUSENGA, BEACH DAGAA, LIFTNETS: 1, 3, 5)

Name of the Recorder ____________________ ☐

Selected FEU:C.No. ☐☐☐

Landing date ☐☐☐

Name of fishing site ____________

1. Particulars of fishing unit	(ASK): Complete the following table:											Remarks
	A. Fishing boat		B. Crew					C. Gear				
								Nets		Other		
	Reg. No. (1)	Type (2)	Total (3)	Boat owner (4)	Crew leader (5)	Asst (6)	Other (spec) (7)	Kind (8)	No. (9)	Kind (10)	No. (11)	(12)

2. Fishing operation

(ASK): Time: 1. Started: 2. Completed:

3. Number of hauls:

4. Did your boat pick up other boat(s) catches?

Yes ☐ 1 No ☐ 2

3. Fish Catch							4. Fish Sold			Remarks
Species	Weight kgs.	Boxes		Baskets		No. of fish	Quantity		Shs/Unit	
		Size	No.	Size	No.		Unit of measure.	No. of units		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
01.0 Dagaa										
02.0 Mikebuka (Luciolates)										
03.0 Lates										
99.9 Other fish										

Sizes of boxes/or baskets: 1 = small size
2 = medium size
3 = large size

LAKE TANGANYIKA FISHERIES RESEARCH AND DEVELOPMENT PROJECT (TANZANIA)
CATCH ASSESSMENT SURVEY (CAS)

FORM: A2(1.3)

CATCH RECORDS
OTHER FISHING METHODS

Name of the Recorder ____________ ☐ Selected FEU:C.No. ☐☐☐

Landing date ☐☐☐ Name of fishing site ____________

1. Particulars of fishing unit	(ASK): Complete the following table:							Remarks (8)
	A. Fishing boat		B. Crew					
	Reg. No. (1)	Kind (2)	Total (3)	Boat owner (4)	Crew leader (5)	Assis-tants (6)	Other (spec.) (7)	

2. Fishing operation	(ASK): Time: 1. Started: 2. Completed: 3. Number of hauls (for beach-fish): 4. Did your boat pick up other boat(s) catches? Yes ☐ 1 No ☐ 2

3. Number of fishing gear used	2. Gillnet		4. Beach-fish		6. Handlines		7. Other (specify)	Remarks
	1. Mesh size		Mesh size	No.	No. HL	No. Ho.		
	2. No. of nets							

4. Fish catch								5. Fish sold			Remarks (12)
Species		Weight kgs.	Boxes		Baskets		Number of fish	Quantity		Shs/ Unit	
C.No. (1)	Name (2)	(3)	Size (4)	No. (5)	Size (6)	No. (7)	(8)	Unit of measure. (9)	No. of units (10)	(11)	

Sizes of boxes/or baskets: 1 = small size 2 = medium size 3 = large size

LAKE TANGANYIKA FISHERIES RESEARCH AND DEVELOPMENT PROJECT (TANZANIA)
CATCH ASSESSMENT SURVEY (CAS)

FORM: R.T.T.

RECORDER'S TIMETABLE

Name of the Recorder ______________________ ☐

AREA UNIT
C.No. ☐

Survey period ______________________

R E C O R D E R ' S T I M E T A B L E								
Departure		Points of travel				Arrival		Remarks
date	time	From:		To:		date	time	
		Name	C.No.	Name	C.No.			

APPENDIX IIIc - WORKING SHEETS 01 AND 02 (WS:01, WS:02) AND INSTRUCTIONS FOR COMPLETION

LAKE TANGANYIKA FISHERIES RESEARCH AND DEVELOPMENT PROJECT (TANZANIA)
CATCH ASSESSMENT SURVEY (CAS)
a. Round: ______
b. Minor Stratum: ______

WORKING SHEET 01 (WS:01)
(EXISTING FEU's AND THEIR COMPONENTS)

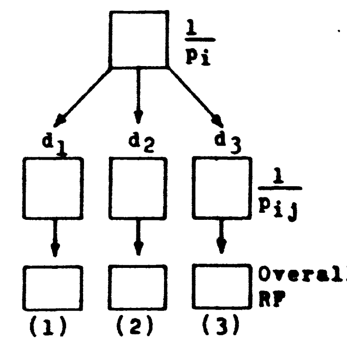
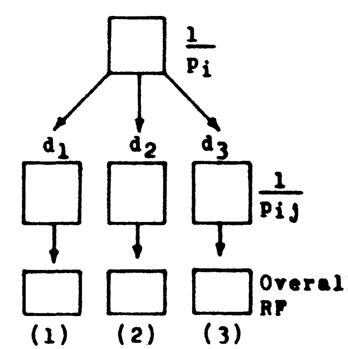
Selected fish.site/ fish.meth.	RF $\left(\frac{1}{P_i}\right)$	T_i D_i	No.of existing FEU's	Components of the existing FEU's										
				(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1)	(2)	(3)	(4)											
1		T_1												
		D_1												
2		T_2												
		D_2												
3		$\bar{D}$												
1		T_1												
		D_1												
2		T_2												
		D_2												
3		$\bar{D}$												
1		T_1												
		D_1												
2		T_2												
		D_2												
3		$\bar{D}$												

LAKE TANGANYIKA FISHERIES RESEARCH AND DEVELOPMENT PROJECT (TANZANIA)
CATCH ASSESSMENT SURVEY (CAS)

a. Round:
b. Minor Stratum:

c. Fishing method:
d. Characteristic(s):
e. Unit(s) of measurement:

WORKING SHEET 02 (WS:02)
(FISHING EFFORT, FISH CATCH)

Selected fishing site/fishing method (1)	Selected fishermen (2)	Sample day: d_1 (3)	(4)	Sample day: d_2 (5)	(6)	Sample day: d_3 (7)	(8)	$d=d_1+d_2+d_3$ (9)	(10)	RAISING FACTORS (11)
	01									
	02									
	03									
	04									
	05									
	06									
	07									
	08									
	09									
	10									
	T_1									
	D_1									
	01									
	02									
	03									
	04									
	05									
	06									
	07									
	08									
	09									
	10									
	T_2									
	D_2									

f:			→ $\frac{D_1+D_2}{2} = \hat{D}$
g:			→ $\hat{T} = c \times \hat{D}$

c = Time Raising Factor

Note:

$T_{1,2}$: Sample total of a given variate (sample day basis)

$D_{1,2}$: $T_{1,2}$ x Overall Raising Factor (sample day basis)

PROCESSING

INSTRUCTIONS FOR THE COMPLETION OF WS:01

1. The purpose of WS:01	WS:01 is used to calculate estimates (totals) of the existing Fishing Economic Units and their components within the established minor strata and within the fishing methods used. Estimates are calculated on a lunar month basis. Annual estimates are obtained by adding up the monthly estimates.
2. The format of WS:01	WS:01 consists of two parts, the heading of the sheet and the body of the sheet.
3. The heading of WS:01	Insert: a) The number indicating the round of the survey: R:01, stands for the first lunar month; R:02, stands for the second lunar month, etc. b) The code number (two digits) of the minor stratum you are dealing with e.g. 11 or 31, etc.
4. The body of WS:01	Within each minor stratum two fishing sites have been selected. Further, for each sample fishing site, Form:Bo has been completed. In the form items of information are collected on the number of existing FEU's, the fishing method(s) used by the units and their components.
4a. Recording the items of information	The body of the sheet is divided into three sections (each section consists of three parts, 1, 2, 3). A section is used to calculate estimates of the survey characteristics on a fishing method (primary fishing method) basis within the given minor stratum. The input data of the sheet are given in the completed Form:Bo. In each section of the body of WS:01 items of information are recorded on a sample fishing site basis. Specifically, line $1.T_1$ is used to record the <u>total value</u> of a given characteristic in the first selected fishing site and within the given fishing method e.g. in Col:4 you should insert the total number of FEU's using Lusenga as primary fishing method. Line $2.T_2$ is used to record the total value of the same given characteristic in the second selected fishing site and within the same given fishing method. In lines $1.D_1$ and $2.D_2$ the sample totals are infered to the minor stratum totals. Specifically, to get the values in line $1.D_1$ one must multiply the recorded values in $1.T_1$ by the respective raising factor (Col:2). Also, to get the values in line $2.D_2$ one must multiply the recorded values in $2.T_2$ by the respective raising factor (Col.2). The estimated value of each survey characteristic is obtained in line 3D: $3\bar{D} = \frac{1}{2}(1.D_1+2.D_2)$
4b. The raising factor, RF	The RF in a given part of a section is calculated by the formula: $RF = \frac{\text{Total number of existing FEU's in the minor stratum}}{\text{Total number of existing FEU's in the selected fishing site}}$ Note that data on the total number of existing FEU's (fishing boats) on a minor stratum basis are given in section 3.2 of the report. Col:4 of the sheet is used to calculate estimates of the total number of existing FEU's by fishing method used. Cols:5-15 are used to calculate estimates of the various components of the FEU's. For each characteristic one column should be used. In the headings of the columns insert the names of the survey characteristics.

PROCESSING

INSTRUCTIONS FOR THE COMPLETION OF WS:02

1. The purpose of WS:02

WS:02 is used to calculate estimates (totals) of the dynamic characteristics of the CAS i.e. fishing effort, fish catch, etc. within the established minor strata and within the fishing methods used. Estimates are calculated on a lunar month basis. Annual estimates are obtained by adding up the monthly estimates.

2. The format of WS:02

WS:02 consists of two parts, the heading of the sheet and the body of the sheet.

3. The heading of WS:02

Insert:

a) The number indicating the round of the survey: R:01, stands for the first lunar month; R:02, stands for the second lunar month, etc.

b) The code number of the minor stratum you start calculations.

c) The name of the fishing method and its code number e.g. LUSENGA, 1.

d) The name of the characteristic(s) under estimation.

e) The unit(s) of measurement of the survey characteristic(s).

4. The body of WS:02

Within each minor stratum two fishing sites have been selected. Further, in each sample fishing site and within the fishing methods used, items of information are collected from the sample FEU's for three consecutive days (completed questionnaires, Form:A1 or A2). The body of WS:02 is used to receive the items of information from the completed forms A1 or A2 and infer them to the minor stratum totals.

4a. Recording the items of information

The body of WS:02 is divided into two identical sections in their structure. The first section is used to receive the items of information from the first selected fishing site and the other section to receive the data from the second selected fishing site. Within each section one line is allocated to record the data per selected FEU, (completed Form:A1 or A2). As you know, an independent sample of FEU's have been selected within each sample day (there are three sample days within each lunar month/fishing site). Therefore, within each section of the body of WS:02 items of information must be recorded on a sample day basis.

The columns of the body of WS:02
Col: 1, 2, stand for the sample code numbers of the selected FEU (see Form: A1 or A2). Specifically; in the large box insert the code number of the selected fishing site, in the small box insert the code number of the fishing method. In Col:3 the serial numbers of the selected FEU's are presented (there is room to insert up to 10 FEU's).
Col: 3, 5, 7, are used to record the values of the characteristic under estimation of the selected FEU's in sample day-1, sample day-2 and sample day-3 respectively.
Col: 4, 6, 8, are spare columns to be used for recording the values of a second characteristic.

4b. Calculation of the overall raising factors, RF

Calculations of the overall raising factors must be made on a sample day basis and within each section of the body of WS:02.
Col: 11: In this column there is a pyramid of boxes in which you have to insert certain arithmetical values i.e.:

a) In the box on the top of the pyramid ($\frac{1}{p_i}$) you have to write a fraction with the enumerator indicating the total number of FEU's in the minor stratum (see section 17.3.2) and the denominator indicating the total number of the existing FEU's in the selected fishing site (see Form:Bo).

b) The three boxes in the middle of the pyramid stand for the three sample days d_1, d_2, d_3. For a given sample day you have to insert in the proper box a fraction, the enumerator indicating the total number of active FEU's within the given fishing method and the denominator indicating the number of selected FEU's (see Form:B1).

c) The three boxes at the bottom of the pyramid are used to estimate the overall raising factors RF within the three sample days {(1), (2), (3)}. The estimation of the values of RF is a mechanical operation i.e. the value of the overall raising factor of the first sample day is given by: "fraction in box d_1" × "fraction in top box"; the overall raising factor of the second sample day is given by: "fraction in box d_2" × "fraction in top box", the overall raising factor of the third sample day is given by: "fraction in box d_3" × "fraction in top box".

4c. Calculation of estimated totals

For the calculation of the estimated totals the following procedure must be used:

1. In each section of the body of the WS:02 and within the respective sample days total the sample values of the survey characteristic. The calculated values must be recorded on the lines with indications T_1 and T_2.
2. Multiply the calculated values (sample totals) with the respective overall raising factors RF, and insert the obtained products on the lines with indications D_1 and D_2.
3. Add horizontally the calculated products of the three days and insert the obtained totals in column 9 or 10 respectively (remember: Col: 3, 5, 7, 9, are reserved for one characteristic and Col: 4, 6, 8, 10, for another characteristic).
4. Calculate vertically the average D_1 and D_2 totals ($\frac{D_1+D_2}{2}$) within Col: 9, 10, and insert the estimated average totals in the boxes with indication f.
5. Multiply the estimated average totals by the respective Time Raising Factor c, and insert the obtained results in the respective boxes with indication g. The obtained magnitudes are the calculated monthly total estimates (lunar month) of the survey characteristics.

Note:

$$c = \frac{\text{Total number of days in a lunar month}}{\text{Number of sample days (=3)}}$$

LIST OF TECHNICAL REPORTS

The technical reports produced by the same author up to the present time, on the application of sampling techniques in fishery statistics (inland waters), are given under sections A-D in the list below. Copies can be obtained from Fishery Statistics Unit, Department of Fisheries, FAO/UN, Rome, Italy.

A: Statistical Studies (St.S.)

1. Sampling Techniques in Inland Fisheries with Special Reference to Volta Lake. St.S./1, FIO:SF/GHA/10, May 1970
2. Frame Surveys at Volta Lake. St.S./2, FIO:SF/GHA/10, March 1970
3. Yield Indices in Inland Fisheries with Special Reference to Volta Lake. St.S./3, FIO:SF/GHA/10, September 1971
4. Frame Survey at Kainji Lake. St.S./1, FI:SF/NIR/24, January 1971
5. The Yield Pattern at Lake Nasser. St.S./1, UNDP/SF/EGY/66/558, September 1972
6. The Yield Pattern at Kainji Lake. St.S./2, UNDP/SF/NIR/24, October 1972
7. Aerial Survey on the Lakes Malombe and Malawi. Analysis of the Results of the Survey. St.S./1, UNDP/SF/MLW/16, November 1972

†8. The Yield Pattern at Volta Lake. St.S./4, UNDP/SF/GHA/10, May 1973

**9. The Improvement of the Fisheries Statistical System at Lake Kossou. St.S./1, UNDP/SF/IVC/71/526, April 1973

**10. Coverage Check Survey of the Aerial Survey at Lake Kossou (CCS-AS). St.S./2, UNDP/SF/IVC/71/526, April 1973

11. The Improvement of the Fisheries Statistical System at Lake Tanganyika (Tanzania). St.S./1, UNDP/SF/URT/71/012, May 1973
12. Statistical Analysis of the Results of the Aerial Survey. St.S./1, Lake Victoria Fisheries Research Project, UNDP/SF/RAF/71/242, August 1973

†13. Recent Trends in the Yield Pattern of Kainji Lake. St.S./3, UNDP/SF/NIR/24, April 1974

†14. Analysis of the Results of Frame Surveys 2 and 3 at Kainji Lake. St.S./4, UNDP/SF/NIR/24, April 1974

B: Statistical Efficiency (St.E.)

1. Efficiency of Different Sampling Methods for Large Scale Biological and Fishery Statistical Sample Surveys at Large African Lakes:
   1. Cove-Rotenone Sample Survey at Kariba Lake, St.E./1, UNDP/SF/ZAM/11, April 1972
2. Deck Sampling: An Assessment of a Pilot Trawling Survey on Lake Malawi (Malawi), St.E./2, UNDP/SF/MLW/16, February 1973

C: Fishery Statistical Surveys/Training (FSS.T.)

**1. Training Courses on Fishery Statistical Surveys (Inland Waters). FSS.T./1, UNDP/SF/ZAM/11, March 1973

** Also available in French.

† Will be available shortly.

D: Technical Reports

**1. Variation in Catchability of Fish with AVB and BOZO Gillnets (Lake Kossou), Report No.3, UNDP/SF/IVC/71/526, August 1973

**2. A Qualitative Evaluation of the Present Statistical System (SS) (Lake Kossou), Report No.8, UNDP/SF/IVC/71/526, August 1973

3. The Sampling Design of the Catch Assessment Survey (CAS), Lake Tanganyika (Tanzania), Report No.1, UNDP/SF/URT/71/012, December 1973

†*4. An Assessment of the Present Status of the Fishery at Lake Tanganyika (Burundi), Report No.1, UNDP/SF/BDI/70/508, January 1974

†*5. The Yield Indices at Lake Tanganyika (Burundi), Report No.2, UNDP/SF/BDI/70/508, January 1974

†*6. A Statistical Analysis of the Results of the Biological Sample Survey (Industrial Fishery) at Lake Tanganyika (Burundi). Report No.3, UNDP/SF/BDI/70/508, January 1974

†*7. Variations in Catchability of Fish with AVB and BOZO Gillnets (Lake Kossou): Analysis of the Results (Dry Season and in Coastal Areas). Report No.18, UNDP/SF/IVC/71/526, April 1974

** Also available in French.

* Will be available in French.

† Will be available shortly.

10:11397