Food and Agriculture Organization
of the United Nations

Training course to enhance collection of fisheries and aquaculture statistics

Module 3 – Data collection and sampling methods

2018

Module 3: Data Collection and sampling methods

Outline

1. Data sources and collection methods
2. Integrated approach to data collections: the Master Sampling Frame
3. Sampling strategies and considerations

Data sources and collection methods

Main data sources for SSF and aquaculture statistics

Small-scale fishery and aquaculture information are usually collected from various sources:

1) Complete enumerations
2) Sample-based surveys
3) Administrative data or reporting systems

1. Complete enumeration: census or frame survey

Frame surveys and fishery/aquaculture censuses are a common category of data collection for which the complete enumeration approach is applied. Structural data can also be obtained by including fisheries-related questions in surveys implemented by non-fisheries line agencies, such as **population censuses and agricultural household censuses**.

Complete enumeration may be preferred in cases where data sources may be legally obliged to report, thus reducing the costs related to this approach.

Complete enumeration may be required as a statutory obligation, often for regulatory purposes. Examples include fishing vessel registers, exports (for custom tariff purposes), variables related to catch quota management (for example, using fishing logbooks), and variables related to fishing effort limitations (such as days at sea).

Complete enumeration may also be preferred where little effort is saved by sampling, such as if the data population is small or the variable to be measured cannot be realistically time-sampled. This might occur with small fishing fleets, regarding which the CPUE is erratic.

1. Complete enumeration: census or frame survey

An important consideration concerning the complete enumeration approach is the risk of negative bias resulting from incomplete coverage. In practice, there is always a proportion of the population that is not captured by a data collection scheme intended to have complete coverage.

The reasons for these information gaps are most commonly associated with operational difficulties. When the proportion of missing data is known to be relatively small, the results can be adjusted to reflect the actual situation.

Another common source of bias occurs when the data collected are used to control of fisheries regulations (such as catch quotas). In this case, deliberate misreporting may occur, to cover illegal fishing.

Developments in data collection technologies, such as Vessel Monitoring System (VMS), electronic logbooks and automatic logging of market information provide opportunities for complete enumeration in situations that were previously ignored or could only be covered by sampling.

2. Sample-based surveys

The basics of sample-based approaches can be summarized as follows:

- Sampling is used if it is impossible, difficult or expensive to conduct a complete enumeration of a target population.
- Sample surveys can operate on selected subsets of the target population and using a number of assumptions regarding the distribution of the population.
- Sample-based surveys provide estimates of the parameters under study, as well as of the sample error.
- A well-designed sampling survey can often produce accurate and reliable estimates at a much lower cost than that of complete enumeration.

However, with sampling, it is necessary to carefully consider how individuals are selected for measurement, whether selecting fish from a catch, vessels landing their catch from all those landing at a particular port, or fishers for interview.

The sampling methods must be based on sound statistical methods and must be fully documented. **Random sampling** and **stratification** are the two basic pillars of sample-based surveys.

Example of surveys for SSF and aquaculture data: **Landing surveys and household surveys (see Modules 4 and 5)**

3. Administrative data or data reporting systems

- **Administrative data** refers to information collected primarily for administrative (not statistical) purposes by government departments and other organizations, usually during the delivery of a service or for the purposes of registration, record-keeping or documentation of a transaction.

- In the SSF and aquaculture rural context, there are two classes of administrative data:
  - **Traditional administrative data :** These data are measurements of well-defined farm entities, and arise naturally through participation in any government or private programme. Examples include information collected through taxation, regulatory processes (e.g. vessels registers), fishery/aquaculture assistance programs (subsidies and insurance) and monitoring programs.

  - **Administrative reporting systems** : These data are normally based on observation or expert judgement, often administered through the Ministry in charge of fisheries on a regular basis. Thus, an extension officer, village chief or another field officer makes a determination based on his or her observations and expertise, and routinely produces reports.

3. Administrative data or data reporting systems

Example: Vessel **register**

- Kept at home ports or, occasionally, landing sites in cases where vessels migrate frequently and do not stay at home ports
- Records information for **operational unit métiers**, that is, vessel type, number of vessels, vessel characteristics (length, motorized or nonmotorized, etc.), fishing gear operated, species targeted, etc.
- It may be designed to take records daily, weekly or monthly, depending on the frequency with which the frame must be updated
- The register must be based on international classifications (discussed in module 1) to obtain standardized records (for example, the CWP international classification of fishing vessels and gears, or the CWP ASFIS list of fish species and other aquatic resources)

Data collection methods

The choice of collection method is influenced by the following: data collection strategy, type of variable, accuracy required, collection point and enumerator skills.

The main data collection methods are:

- Registration
- Questionnaires
- Interviews – open-ended, structured
- Direct observation – observers, inspectors, scientific research, data logging
- Reporting – harvest, post-harvest, trade, sale

Integrated approach to data collections: the Master Sampling Frame

Master Sampling Frame for SSF and aquaculture statistics

A Master Sampling Frame (MSF) for SSF and aquaculture statistics is an upcoming development in fishery-sector data collection. It leads to reliable, cost-effective and integrated statistics using a master frame consisting of several frames of landing sites, households, vessels/boats, etc.

- The Global Strategy to improve Agriculture and Rural Statistics (GSARS), is preparing guidelines on using MSF for fisheries and aquaculture surveys
- Research and field tests on the approach are being conducted. The results will be published shortly on the GSARS website (http://gsars.org/en/tag/fisheries/)

Master Sampling Frame for SSF and aquaculture statistics

Statistical models can be used to combine survey data from multiple sources and produce consistent integrated statistics on SSF and aquaculture for a wide range of policy objectives, such as conservation, food security and economic sustenance.

Investing in the construction of an MSF significantly reduces the cost of surveys in the long term, as a stable and reliable master frame made of frames of landing sites, households, vessels/boats, etc. is initially built at a high cost and level of technical know-how, but is then followed by lower usage costs and minimal operating technical know-how requirements as the costs and technical skills required to build many individual frames need not be expended.

By facilitating sample-based surveys, MSFs allow for cheaper long-term data collection, thus enhancing sustainability. In addition, the estimates produced using MSFs also present an acceptable degree of statistical reliability.

Master Sampling Frame for SSF and aquaculture statistics - Main steps

1. **Definition of goals and purpose of activities:** this includes considering important aspects of the study area or population – criteria for stratification, spatial and temporal domains for estimation, species to be included, etc.
2. **Identification of populations** that contribute to the estimation's objectives. This step defines operational aspects, such as defining the population of carp aquaculture to be estimated, as the species can found in ponds, rivers, small cages in lakes, etc.
3. **Definition of variables and measures:** the variables to measure or observe depend on the goals of the estimation and inference. A careful evaluation of options and the feasibility of obtaining different measurements is required for each specific context.

Master Sampling Frame for SSF and aquaculture statistics - Main steps

3. Definition of variables and measures

The variables of interest, as determined by the survey goals, have important implications for the choice of the frame. Common types of frames for surveys of socio-economic variables are lists of households, communities, or telephone numbers. On the other hand, spatio-temporal frames are more commonly used for surveys of production-related variables, such as fishing catch and effort.

Considerations related to this step include the following:

- Are variables the same or different over multiple populations?
- Are the measurement scales the same or different?
- Are the units of measurement consistent across populations, or can they be reasonably combined in one estimation method?

Master Sampling Frame for SSF and aquaculture statistics - Main steps

4. Identification of contact methods

Surveys of fisheries and aquaculture employ different data collection schemes. The contact method is an important component of the overall data collection strategy, which is often decided on the basis of the sampling frame chosen.

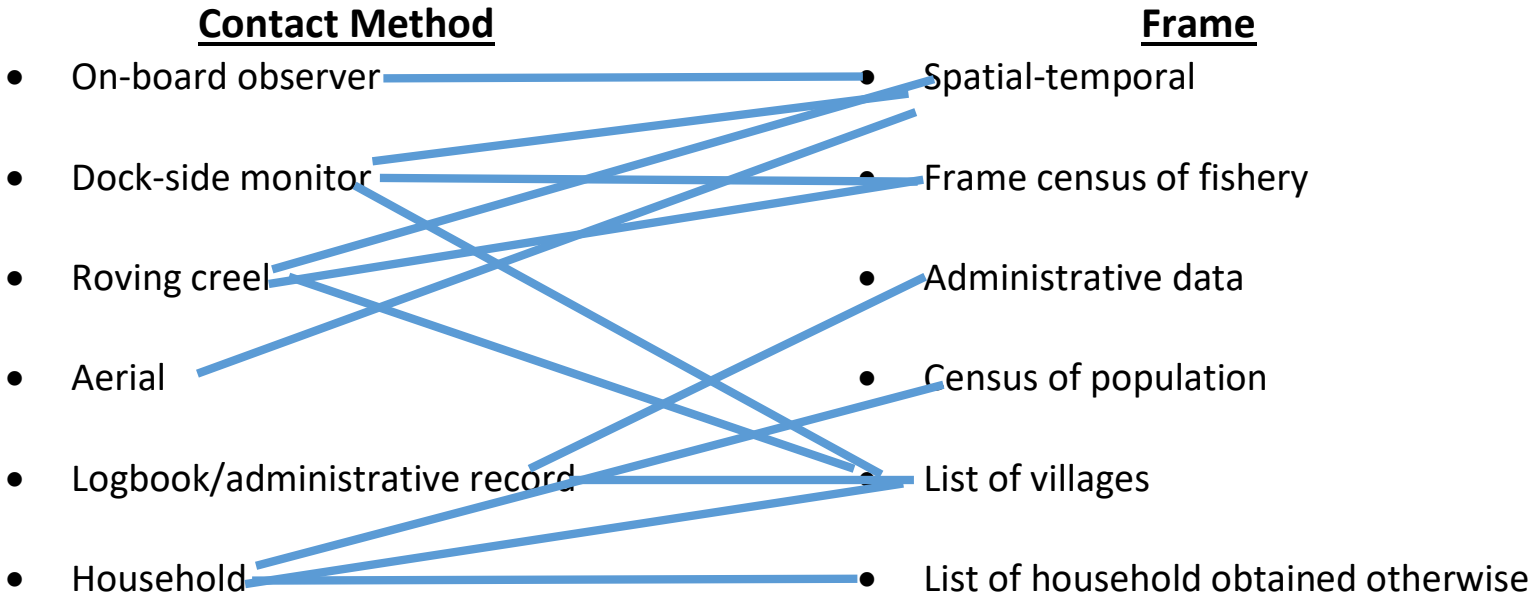
The contact methods for surveys of fisheries may be divided into two broad categories:

- **on-site (direct)** methods: the data are collected at the location where fishing occurs
- **off-site (indirect) methods** include telephone surveys, household surveys, and data collected from logbooks

Master Sampling Frame for SSF and aquaculture statistics - Main steps

4. Identification of contact methods

The data collection method is closely linked to the choice of the frame. Certain contact methods are typically used with certain types of frames.



Examples of questions related to this step include: Have appropriate media been identified for data recording? What amount of training and costs will be involved? Have focal points been identified for implementation?

Master Sampling Frame for SSF and aquaculture statistics - Main steps

5. Outline the sampling design

This step includes details of the type or types of sampling design used – not only stratified sampling, for example, but also the number and identity of strata, how those will be sampled, how primary and secondary units in a multistage design will be sampled, etc. Consideration of all steps in stage should point to one or more potential sampling designs, which may differ depending on the population or variable.

- How many stages are required? Is the design multistage or multiphase?
- What types of frames are needed so that inclusion probabilities can be computed for all *sampled units*?
- Is the design specific enough that its description could allow individuals who have not been involved in this point of the process to implement it?

Master Sampling Frame for SSF and aquaculture statistics - Main steps

6. Construct the sampling frame

The availability of sampling frames affects the considerations that must be made prior to this step. Similarly to sampling designs, there should be a fair amount of general information on possible frames that has been generated or used in previous steps.

Questions related to this step:

- Has the use of each frame been explicitly identified and attached to it?
- Has the maintenance schedule of each frame been determined?
- Have unique identifiers for individual units (population and sampling) been used, so that overlap in frames can be detected?
- Are the frames constructed compatible with the sampling design, variables, and definitions of population units resulting from previous steps of the process?

Master Sampling Frame for SSF and aquaculture statistics - Main steps

7. Develop estimation strategy

This step may involve consultation with external experts, and may be revised and improved without necessarily repeating the entire process. Estimation may be considered separately for each objective identified in the first step of the process, or may be considered as a whole from the outset. Either way, it is important to consider both estimation for each objective individually and the set of estimation problems as a coherent unit.

Questions related to this step:

- Are all quantities identified in the objectives addressed in the estimation?
- Are the quantities associated with the uncertainty (or inference) produced by the strategy developed?
- Does the set of estimation procedures hold together in a unified (or coherent) approach?

Master Sampling Frame for SSF and aquaculture statistics - Main steps

8. Develop framework for storage and retrieval of data and results

This step focuses on the format that can be retrieved, updated and accessed for use by decision-makers. For example, a frame used to sample for the direct observation of catch and a different frame used for an aerial survey of fishing activity may be joined by an edge that represents dependence in estimation (CPUE). In contrast, one frame of registered fishing vessels and another of landing sites that might be sampled for activity could be joined by an edge that represents multiple frames for the same variable (active fishing vessels in a year).

Questions related to this step:

- Can the storage/retrieval system be used to rapidly locate information for the implementation of sampling in the current spatial and temporal window of interest?
- Can the system be used to rapidly retrieve data for use with potential new estimation strategies, or for the production of estimates for population divisions that could not be foreseen when the program was initiated?

Sampling strategies and considerations

Data collection strategy – steps

- Evaluate existing data sets in relation to the objectives of the programme, including data accessibility
- Describe the operating characteristics of the sector or subsector (fishery, market, fleet, community, institutional environment, etc.), also known as the census or frame survey
- Decide on the approach to be taken: complete enumeration or sampling, including a cost/benefits and cost effectiveness analysis, and an evaluation of operational considerations
- Design methods according to the approach adopted, including the form of stratification to be used in sampling
- Implement a test phase to validate the method, including the participation of other stakeholders
- Establish a continuing feedback mechanism between data sources and data users to ensure that the data types, quantity, quality and origin are consistent with the requirements to determine the performance indicator in question

Complete enumeration and sampling

- Data collection is the recording of one or more data variables (catch, value, fishing duration, etc.) from the members of a population of "data units" (the population of fishing vessels, fishers, etc.).
- Two basic data collection approaches are possible:
  - complete enumeration, where all members of the whole population are measured: CPUE, prices, vessel registers, etc.
  - sampling, where only a proportion of members of the whole population are measured: species comp., etc.

Example

- Lake Bosumtwi
- Area: 100 ha
- 20 boats
- One landing site
- Well-trained staff
- Unlimited funds.

- How would you monitor this small lake to obtain the most accurate results?

Plenary discussion

Example

❖ Lake Victoria

- Area: 35 000 km^2
- 30 208 boats
- 634 landing sites
- Limited staff
- Limited funds

- How would you monitor this big lake to obtain the most accurate results?

Example

Conclusions

- Lake Bosumtwi: small area, limited number of boats. Can be covered completely

  Full enumeration or census

- Lake Victoria: large area, large numbers of landing sites, large number of boats. Sampling is essential.

Sample-based approaches

- Sampling is done if it is impossible, difficult or expensive to observe all elements of a target population.
- Sample surveys can operate on selected subsets of the **target population** and using a number of assumptions on the distribution of the population.
- Sample-based surveys provide estimates of the parameters under study, as well as the sample error.
- A well-designed sampling survey can often produce accurate and reliable estimates at a much lower cost than complete enumeration.

Objectives of sampling

- Obtain data on stocks and their exploitation (for example, what fish exist in the body of water being studied, how many there are, what species)
- Analyse the characteristics of the resources (such as the more striking characteristics (size, feeding habits, etc.) of each species)
- Analyse the effects of exploitation on the abundance of resources
- Determine appropriate fishing levels to obtain best possible catches for the present and future years

Sampling considerations

- What is sampling?
- Objective?
- Design or select sample sites or stratification
- Sample size
- Sampling frequency
- Sample methods
- Habitat
- Others/resources

What is sampling?

Sampling is a ‘short-cut’ method to investigate an entire group.

What is sampling?

Example

It costs 1 USD ($) to investigate one square of the puzzle

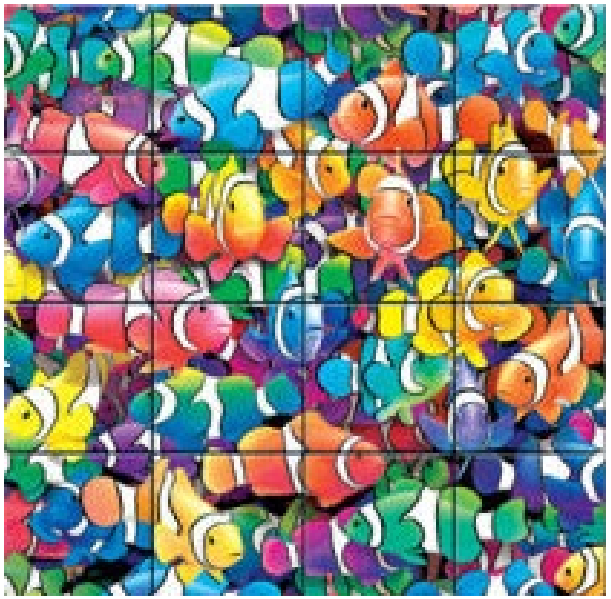
Total cost = 16$

Cost of sample = 4$

Good estimate

What is sampling?

Can the cost be reduced and still provide an acceptable estimate?



Total cost = 16$

1$ 

Biased estimates

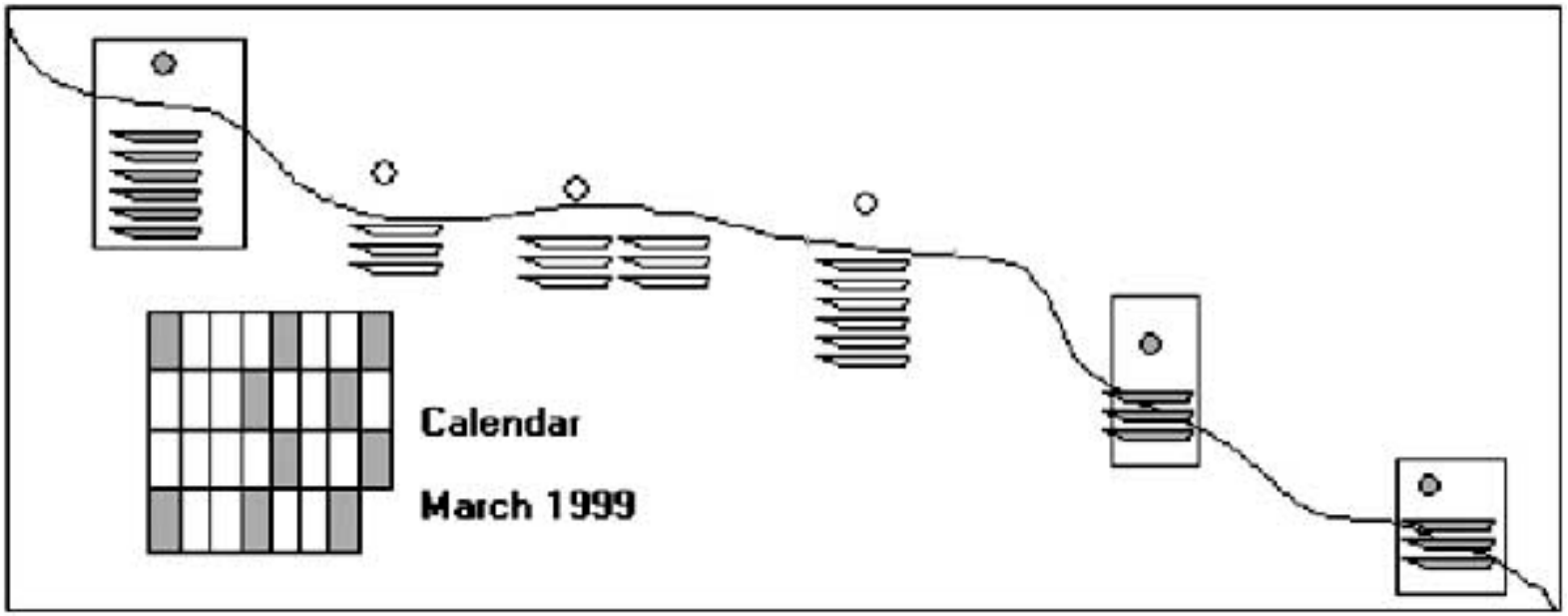
1$ 

2$

Combined sample

Sufficiently good estimate?

Sampling in time and space



- Most commonly used
- At some sites, boats are sampled (space)
- Sampling is done on certain days only (time)

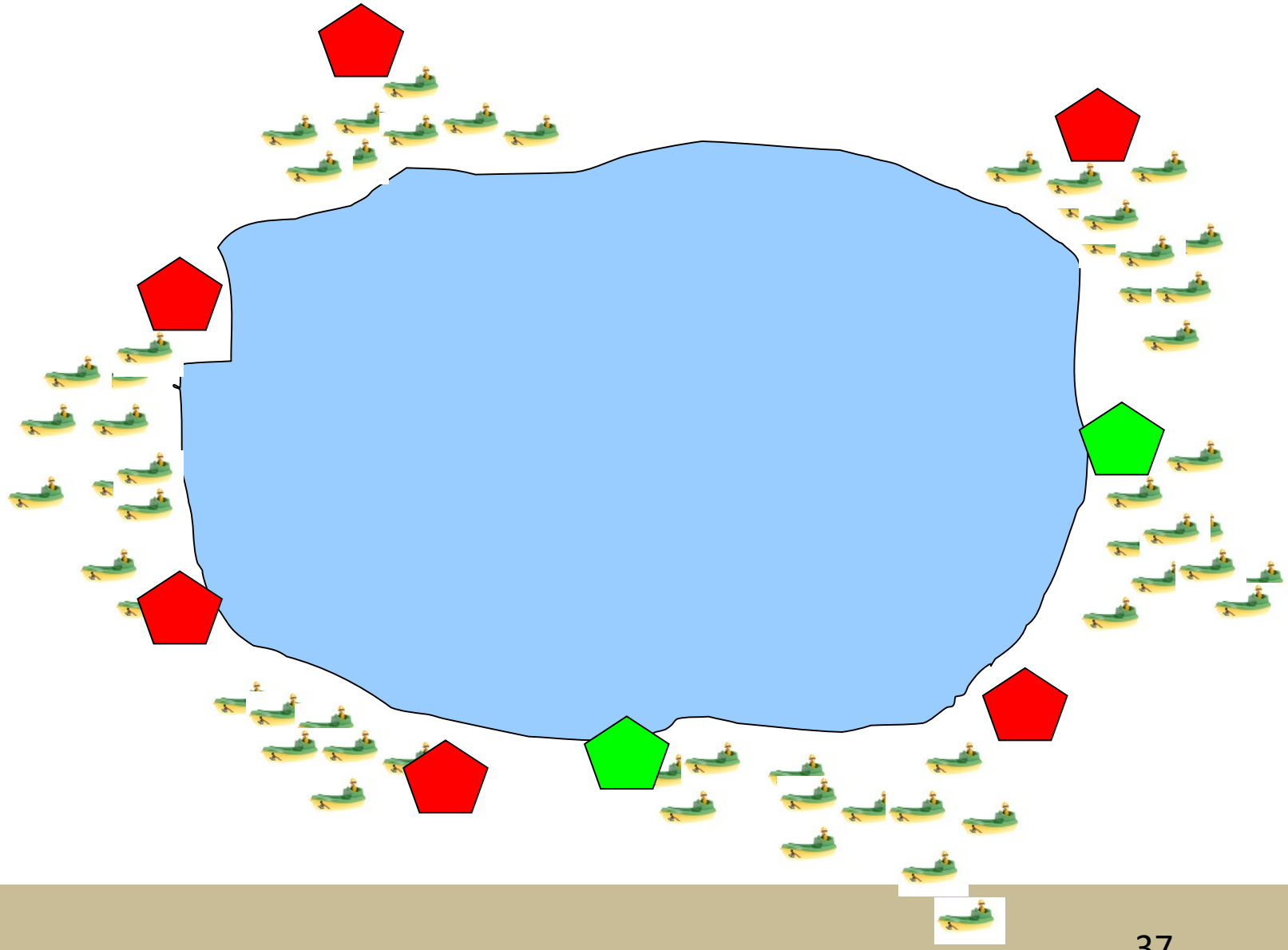
The sampling method is crucial

Random sampling (without choosing) is the best scientific way to sample. It prevents investigators from applying their judgement on choosing sites, time, etc.

Random sampling methods

- Simple Random Sampling
- Stratified Random Sampling

Example - 8 Landing sites – where and what do we sample?



Example - 8 Landing sites – where and what do we sample?

Selecting at random

- Random selection of two landing sites -> high probability that only red sites are chosen
- May work for boats, but ***not*** for landing sites
- Thus, systematic site selection through stratification is used

Stratification

- **Major strata** – done for administrative classification, such as region or district
  - Example: Ecological strata

- **Minor strata** – improves sampling for higher accuracy and reducing costs!
  - Example: Fishing units

Stratification

- Aims to reduce variability in the sampled data, which will:
  - Improve the reliability of the collected data
  - Reduce the number of samples needed
  - Consequently, reduce the sampling cost

Example

Major stratum: Gomoa District

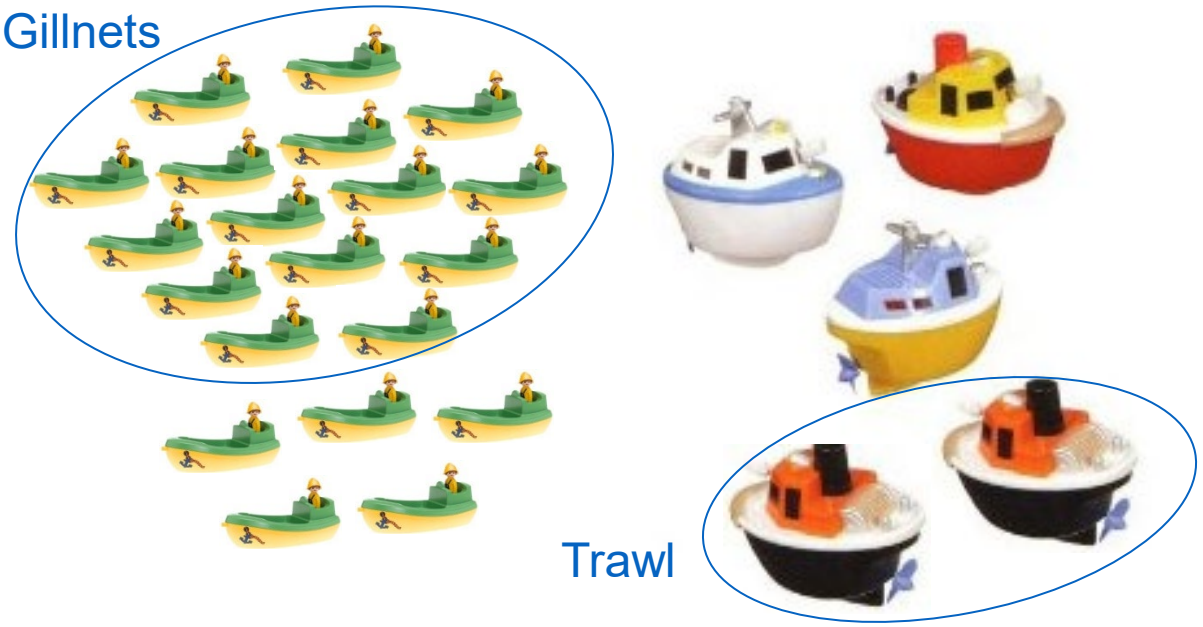
Minor stratum: River Tana

Fishing units: Boats with traps
Boats with gillnets

Structure of sampling

Similar boats and similar gear are already grouped and counted separately in the frame survey

Structure of sampling

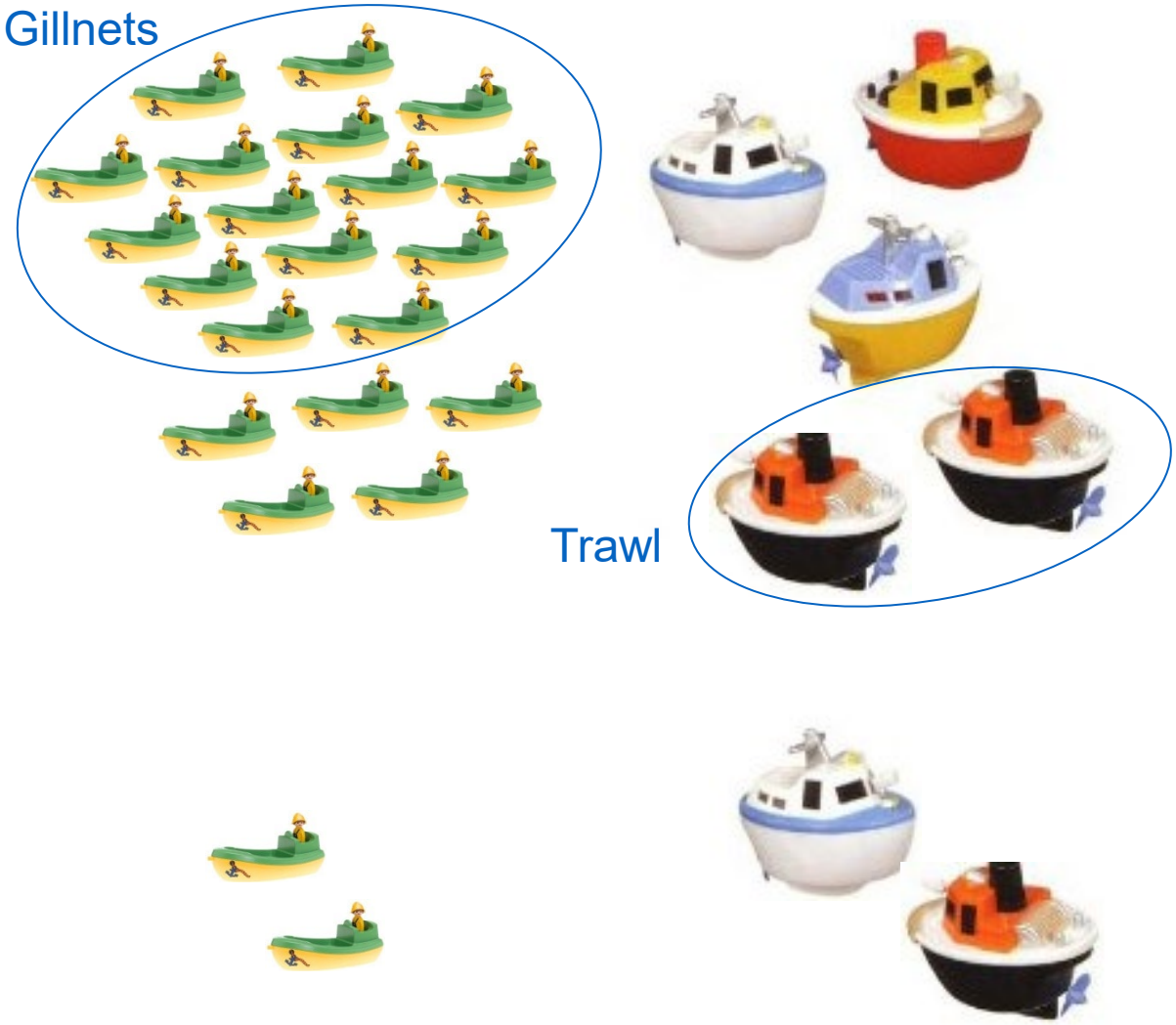


Similar boats and similar gear are already grouped and counted separately in the frame survey

Structure of sampling

- 1. Similar boats and similar gear are already grouped and counted separately in the frame survey

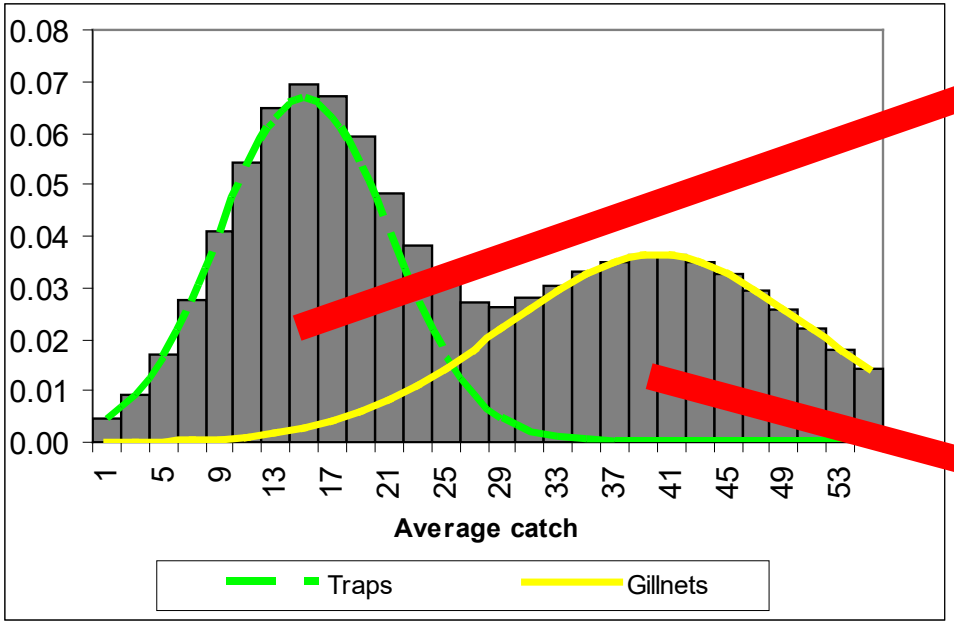
- 2. **Samples** are taken from all gear and boat groups, so that all fish species and sizes have a chance of being included

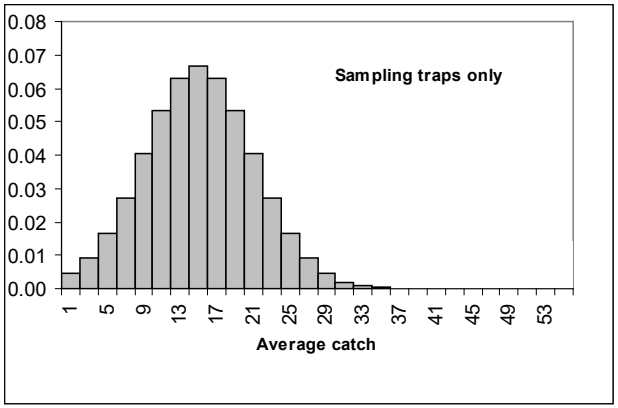


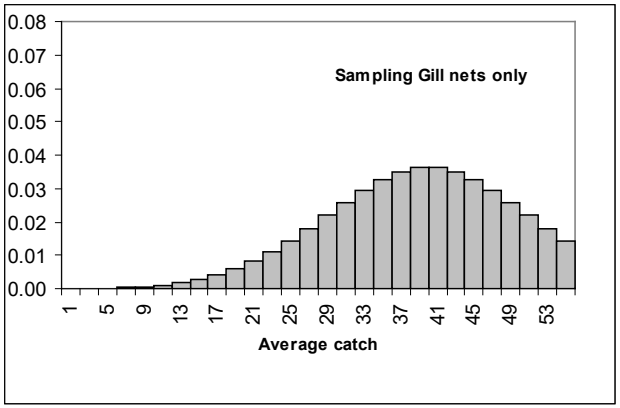
Structure of sampling

- 1. Similar boats and similar gear are already grouped and counted separately in the frame survey
- 2. **Samples** are taken from all gear and boat groups, so that all fish species and all fish sizes have a chance of being included

Stratification







Combined sampling of traps and gillnets

Large variation

Why this stratification?

- Even if several samples are drawn from traps and gillnets combined, it would still not be possible to obtain accurate estimates

- By separating traps and gillnets, a more accurate estimate may be obtained for a lower cost (smaller samples needed)

References

- **Stamatopoulos, C.** 2002. *Sample-based fishery surveys. A technical handbook.* FAO Fisheries Technical Paper No. 425.
- **de Graaf, G.J., Nunoo, F., Ofori Danson, P., Wiafe, G., Lamptey, E. & Bannerman, P.** 2015. *International training course in fisheries statistics and data collection*. FAO Fisheries and Aquaculture Circular No. 1091. FAO Publication: Rome.

Thank you