How to design sampling strategy and protocol ?

IOTC Species Identification and sampling workshop

Kochi, India – September 29th to October 4th 2025

Food and Agriculture Organization of the United Nations

INDIAN OCEAN Tuna Commission

OUTLINE

- Objectives of the session
- Introduction

1. Building data collection and sampling strategy
2. Sampling design and sampling plan
3. Sample size
4. Sample allocation
5. Sampling protocol

Objectives of the session

- Sample design
- Calculate optimal sample size (regarding tolerated error)
- Allocate your sample regarding inter-strata variability

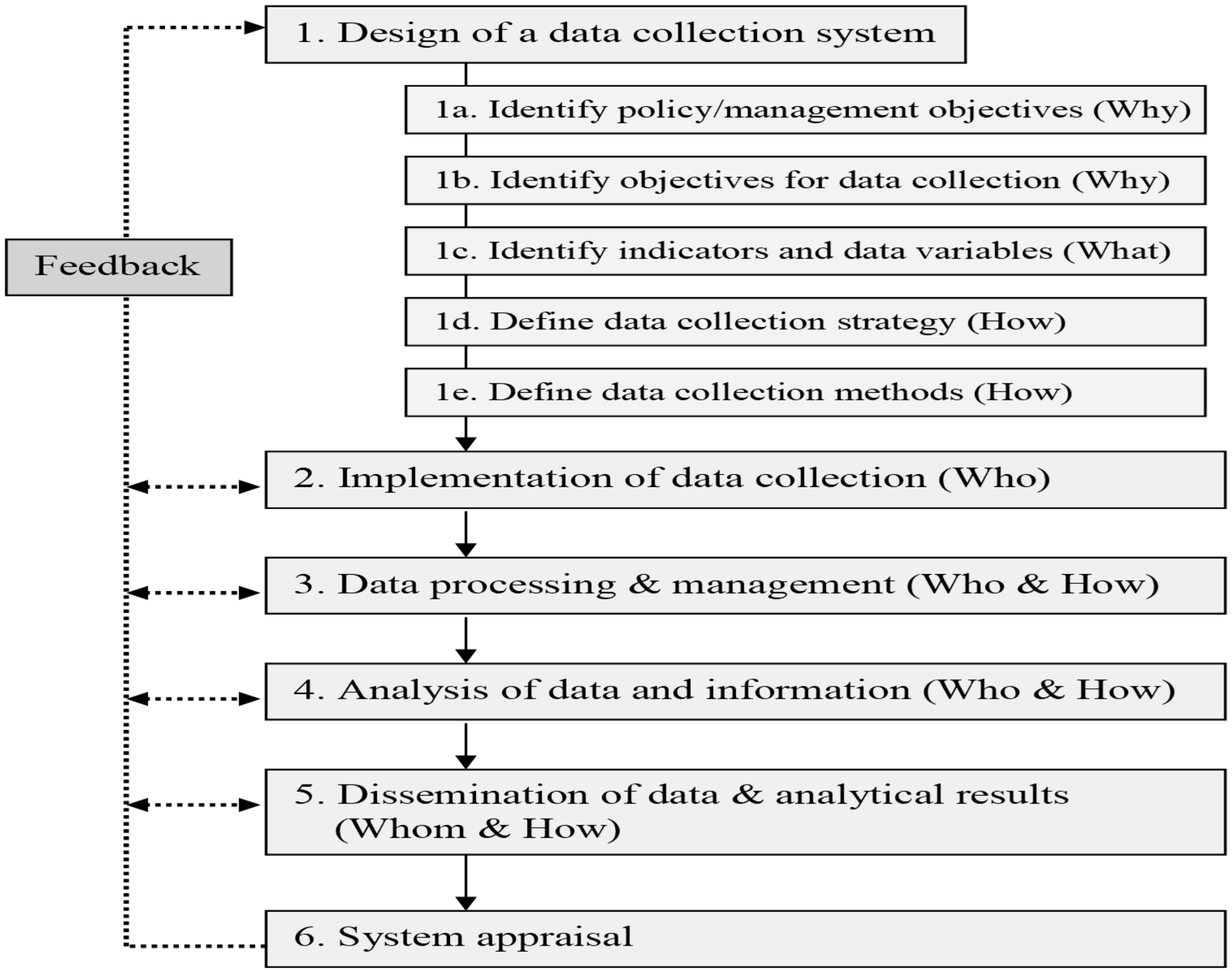
1. Introduction
Generic Statistical Business Process Model

Overarching Processes

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

1. Introduction

- Data collection process



IOTC Species identification and sampling workshop, 2005

1. Building data collection and sampling strategy Specify your objectives and needs

1. What are your objectives?
   - Production data for policy maker
   - Stock assessment
   - Resource management

2. Define your needs
   - Estimate total catch by species (target and bycatch), by month, by gear, by site…
   - Estimate fishing effort
   - Estimate number of vessels
   - Estimate total of discard
   - Collect size frequency, or other biological data or samples
   - Estimate accidental catch of ETP species

1. Building data collection and sampling strategy

3. Identify variables and indicators = 2nd step of the design of a data collection system
4. Designing a sampling scheme
   - A wide variety of approaches/methods can be used to collect the variables, but in data collection a clear distinction should be made between "routine data collection" and "adhoc/scientific surveys" .
   - Routine data collection systems should be robust and simple and can provide the basic long-term data series on catch, effort, major species and price/value
5. Additional information, such as: Length frequencies, biomass, biodiversity, state of the ecosystem, stock status, socio economics, livelihoods, households could be covered through separate monitoring programmes or "adhoc/scientific surveys".

IOTC Species identification and sampling workshop, 2005

1. Building data collection and sampling strategy
Designing a sampling scheme

- With limited human and financial capacity, the long-term sustainability of small-scale fisheries routine data collection systems can only be guaranteed by focusing on the basics and applying appropriate statistical approaches.
- Reliable statistics or collection of data requires the establishment of a permanent system which may be modified over time to adapt to changing circumstances and objectives

IOTC Species identification and sampling workshop, 2005

2. Sampling design and sampling plan Sampled sites selection

- Random sampling is rarely easy to implement:
  - Various constraints
  - Field reality
- Important points
  - Focus on bigger landing sites
  - Sample fleet with biggest catch
  - Include species with high catch and/or high value
- Verify your sample is representative of the population you want to study
  - Compare your sample composition with your population composition
  - Make sure all strata are correctly sampled
- Define your data analyse in parallel

2. Sampling design and sampling plan Sampled sites selection

- The main **criteria in selecting sampling sites** are:
  - Sampling sites should provide a satisfactory geographical coverage of the statistical area. Limited human resources or transport will usually be the major operational constraint to this coverage.
  - Original frame surveys of the numbers of boats (fishing units) by site and boat/gear type will indicate the relative importance of sites (i.e. very important, important, less important, etc.). Sampling sites should represent boat/gear types involved in the survey.

2. Sampling design and sampling plan Sampled sites selection

- No preliminary data available
  - We use a cautionary approach. During the first year of sampling this type of information will become available, and therefore it is recommended to cover all different types of landing sites and within each stratum the samples could be allocated proportionally to the size of the landing sites. Or you use the more cautious approach of Stamatopoulos (2002) (1).
- Preliminary data available
  - Selection of landing sites is straightforward, if sample sizes for the different strata are determined with preliminary available data and application of relative error.
  - If, however, funds and staff are available to cover more landing sites then you could do, as this minimise risk of sampling errors.

(1) https://www.fao.org/3/y2790e/y2790e.pdf

2. Sampling design and sampling plan
Sampled sites selection

- Good accuracy levels can be achieved at relatively small sample sizes, provided that the samples are representative of the population.

- If there are no differences in CPUE for small and large landing sites then only large landing sites are sampled.

- Beyond a certain sample size the gains in accuracy are negligible, while sampling costs increase significantly.

2. Sampling design and sampling plan Stratification definition

- Minimal stratification level = **Fishing unit**
- **A Fishing Unit** is a group of fishing vessels practicing the same type of fishing operation, targeting the same species or group of species and having a similar economic structure". Examples of operational units are:
- Defining the Fishing Units for small scale fisheries is often straightforward. However, multi gear use in small scale fisheries should be taken care of.
- If gears are changed within the season then a fleet segment is split up in a number of operational units (Table next slide). This means that the number of operational units can be more as the total number of canoes in the frame survey!

2. Sampling design and sampling plan Data collection requirements

- 4 main components:
  - CPUE -> sampling ⇨ what, where, when ?
  - Active vessels -> complete enumeration (registry or frame survey)
  - Active days -> information collected at landing sites (fishers, neighbour)
  - Activity coefficient -> sampling ⇨ what, where, when ?

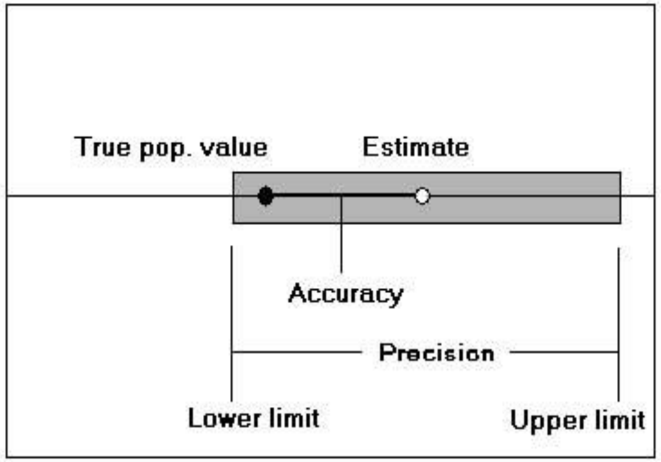
=> Sampling in space and time is the most efficient

- Additional component: biological data

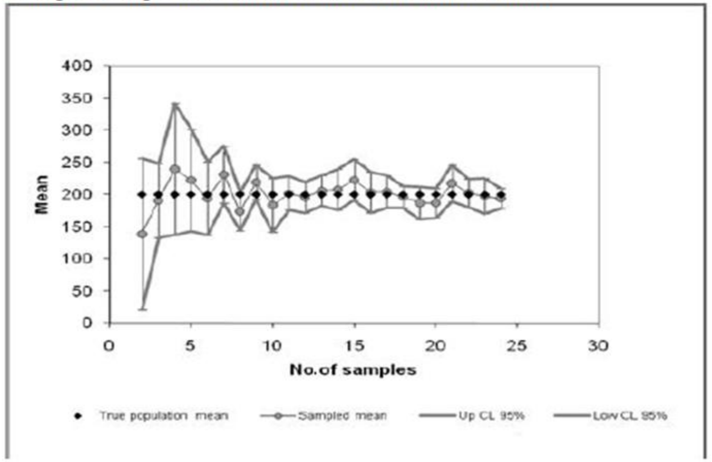
2. Sampling design and sampling plan
Accuracy and precision in sampling

- **<u>Precision</u>**

Precision describes the closeness of repeated measurements of the same quantity. Precision examines how well-clustered around the mean the sample observations are. Note its relation to the confidence interval (upper and lower limits).



Example of sample sizes, confidence limits and estimated means



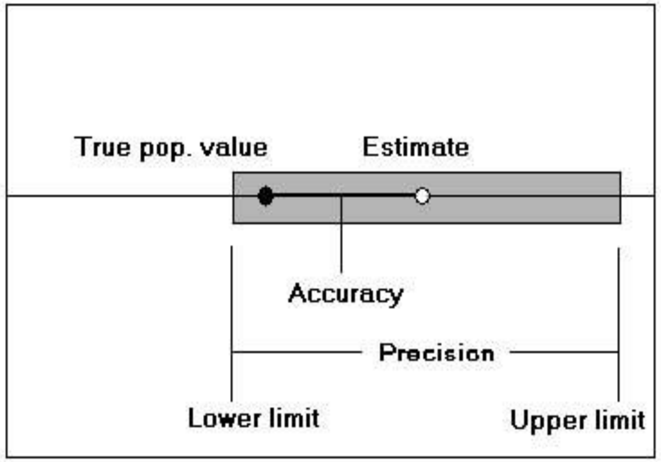
IOTC Species identification and sampling workshop, 2005

2. Sampling design and sampling plan
Accuracy and precision in sampling

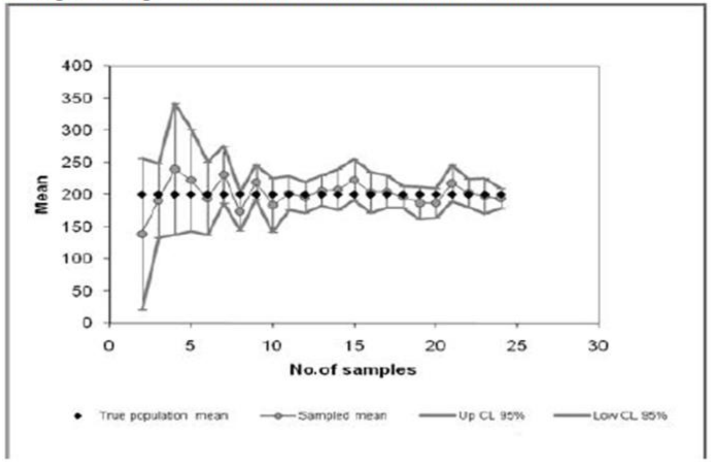
- **<u>Accuracy</u>**

Closeness of measured or computed value to its <u>true target population</u> value. Accuracy examines whether the estimate mean is close to the true population parameter. A sample that gives an inaccurate estimate is called a BIASED SAMPLE.

ACCURACY = PRECISION + BIAS



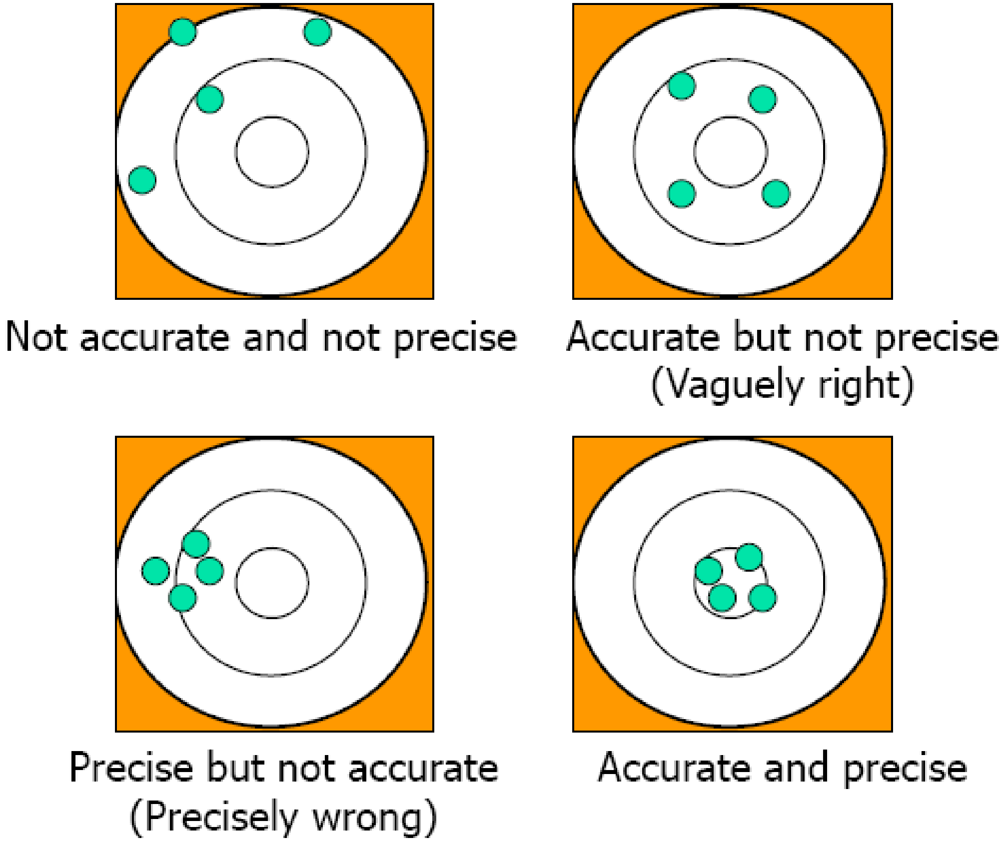
Example of sample sizes, confidence limits and estimated means



2. Sampling design and sampling plan
Accuracy and precision in sampling

- Sample estimates can be precise but not accurate

- Do you know an example of this?

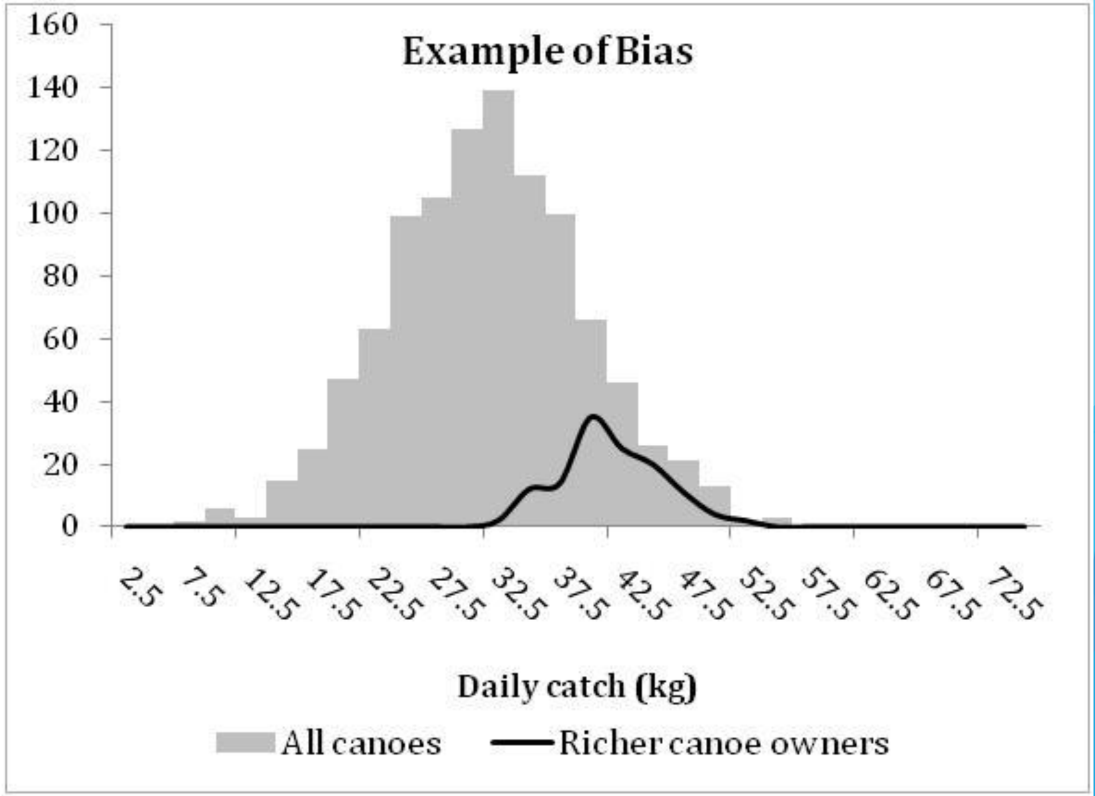


2. Sampling design and sampling plan
Avoid bias of sample estimates

- Bias is the tendency for sample estimates to centre upon a value that is different from the true value, as data accumulate

- In many cases bias is the result of systematic errors in the sampling (e.g., repeated sampling of the same vessels).

- IMPORTANT: biased data may produce inaccurate but precise estimates.



IOTC Species identification and sampling workshop, 2005

Potential bias in landing surveys (catch and effort)

In practice, it is not easy to implement full random sampling at landing sites:

- Enumerators may prefer to work with vessel captains who are cooperative, etc., which means the same vessels are repeatedly selected (non-random selection of vessels).
- Vessels often land their catch very early in the morning or late in the evening - when enumerators are not present. Similarly, enumerators may visit the landing site on the same day, at the same time.
- When boats land within a short period, enumerators may tend to sample those with a small catch in order to cover as many landings as possible.
- If landings occur over longer periods and recorders must visit other sites during the day, only the first landings at the first site will be sampled.

- *** All can lead to systematic bias in type of vessels (and the associated catches/species composition of landings) that are sampled ***

Potential bias in biological surveys

Other case studies: biological sampling

1. Data collectors tending to choose larger fish size (self-selection, non-random samples).

2. Catches that are unloaded may already be sorted in size and quality.
   - Sampling protocols based on the sampling of the first 20 fish, for example, would therefore be unrepresentative of the length frequencies of total catches unloaded.
   - Adjusting the sampling protocols (e.g., sampling nth fish based on a random number generator) will ensure the data are more representative.

3. Sample size
Determining the total samples required

Two approaches are possible, depending on the data we have available prior to sampling:

1. **Preliminary data** are available for the different fishing units and minor strata.
   - *Usually obtained from a pilot study, prior to sampling.*

2. Only **frame survey** and structural data are available.

Bias of sample estimates

- Biased estimates may be found to be systematically above or below the true (but unknown) population value - generally because they are derived from samples that are not representative of the true population.
- Bias is not easily detectable, and at times not detectable at all.
- Consequently, users may be unaware of the problem since they also do not know the true population value.
- To minimise the risk of bias requires effective design and monitoring of sampling in the field.
- Local knowledge of the characteristics of the fisheries + well trained (and well motivated!) enumerators is equally as important as applying the principle of sampling theory.

Role of staff – OFFICE

- Primary data collected by field staff are requires an appropriate statistical office infrastructure
- Responsibilities and functions of statistical office staff:
  - Designing and planning of fishery surveys activities (implementation scheduling, training, equipment and logistics support)
  - Coordinating and monitoring field and office activities
  - Data processing (compiling and reviewing primary data obtained from the field, data checking, editing, correcting and analysing data)
  - Reporting results of data analysed to stakeholders

Role of Staff – FIELD

- The field staff are the data collectors and their supervisors. They are the primary interface between fishers (and also aquaculture-practicing households) and fisheries management
- They collect and submit data to the fishery statistical office for further processing
- The important role of the field staff involved in data collection can be summarized as follows:
  - **Winning and maintaining cooperation of respondents**
  - **Ensuring quality of data by correctly carrying out instructions on data collection received in regular training**
  - **Providing feedback to office for evaluation**

3. Sample size
Method 1: Preliminary data are available

- If preliminary data are available, the total number of samples required for sampling can be estimated using the relationship between the relative error (in the preliminary samples) and sample size.
- **The relative error** is set by the survey designer and is expressed in percentages (usually set at around 10%). It represents the *tolerable* level of relative error.

3. Sample size
Method 1: Preliminary data are available

- The relative error formula requires some prior knowledge of the population variance/standard deviation and mean - from preliminary data

Relative error is: $\varepsilon = \frac{CL}{\bar{x}} = \frac{tn{-}1s}{\sqrt{n}\,\bar{x}}$

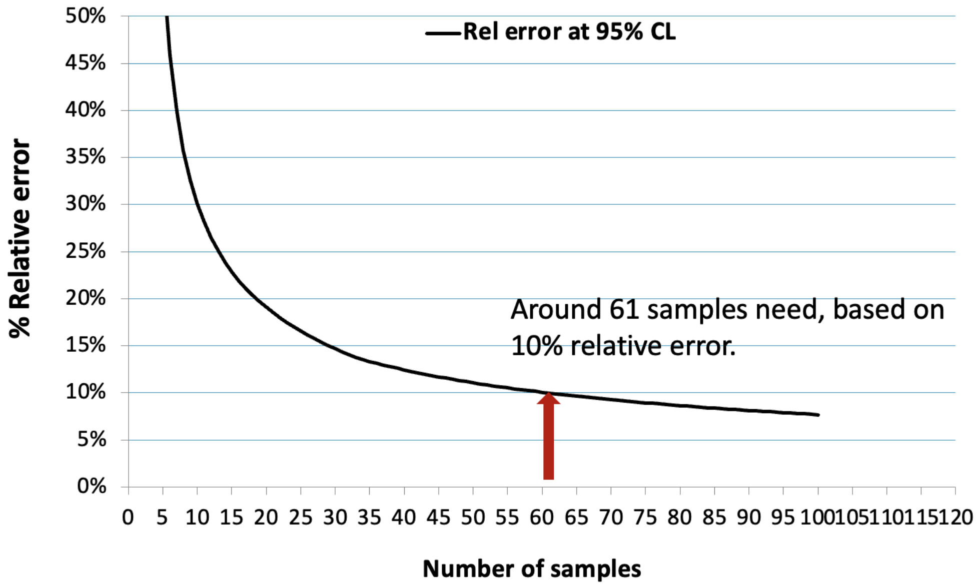
- In the equation we see ε on the left and, the square root of the sample size (n) on the right n.
- With some mathematical manipulation we can swap them (‘change of subject’) and the number of samples needed can be estimated with:

$$n = \left[\frac{t_{n-1}s}{\varepsilon\bar{x}}\right]^2$$

n = sample size needed
$\bar{x}$ = average of preliminary samples
s = standard deviation of preliminary samples
tn-1 = value of t fractiles
ε = Tolerated relative error

3. Sample size
Method 1: Preliminary data are available

- The value of *n* can be obtained, graphically or through iteration.
- ϵ is calculated for a range of values of *n* until the *n* value where ϵ approaches the tolerated value (10%) is found.



3. Sample size
Method 1: Preliminary data are available

- To monitor this fleet correctly, a total of 190 canoes per month must be sampled. The results also clearly illustrate that the total number of canoes of each fishing unit has a limited influence on the number of samples needed.
- There is no large difference in the number of samples needed, whether the target population (sub-population) includes 6,000 vessels or 1000 vessels.
  - ***What is more important is the variance within each strata, as expressed by the coefficient of variation.***

Canoe type	No of canoes	Avg daily catch	STD daily catch	Coefficient of variation (%)	No of samples for 10 % error
Large, purse seines	1000	125	45	36.0	51
Large, hook and line	3500	90	17	18.9	18
Small, gill nets	6000	18	6	33.3	44
Small, traps	1500	9	4	44.4	77

Total number of samples required for 10% error = 190

IOTC Species identification and sampling workshop, 2005

3. Sample size
Method 2: No preliminary data are available

Sampling requirement at various accuracy level and data population size for boat activity survey

Accuracy (%)	90	91	92	93	94	95	96	97	98	99
Data Population size	**Safe sample size for BACs**									
300	73	85	100	119	141	168	200	234	267	291
400	77	91	109	132	160	196	240	291	343	384
500	81	96	115	141	174	217	273	340	414	475
600	83	99	120	148	185	234	300	384	480	565
700	84	101	124	153	193	248	323	423	542	652
800	86	103	126	157	200	260	343	457	600	738
900	87	105	129	161	206	269	360	488	655	823
1000	88	106	130	164	211	278	375	516	706	906
2000	92	112	140	179	235	322	462	696	1091	1655
*see notes **3000**	**93**	**114**	**143**	**184**	**245**	**341**	**500**	**787**	**1334**	**2286**
4000	94	115	145	187	250	350	522	842	1500	2824
5000	94	116	146	189	253	357	536	879	1622	3288
6000	95	116	146	190	255	361	546	906	1715	3693
7000	95	117	147	191	257	364	553	926	1788	4049
8000	95	117	147	191	258	367	558	942	1847	4364
9000	95	117	148	192	259	368	563	954	1895	4646
10000	95	117	148	192	260	370	566	964	1936	4899
15000	95	118	149	193	262	375	577	996	2070	5855
20000	96	118	149	194	263	377	583	1013	2144	6488
25000	96	118	149	194	264	378	586	1023	2191	6939
30000	96	118	149	195	264	379	588	1030	2223	7275
35000	96	118	149	195	265	380	590	1036	2247	7536
40000	96	118	150	195	265	381	591	1039	2265	7745
45000	96	118	150	195	265	381	592	1042	2279	7915
50000	96	118	150	195	265	381	593	1045	2291	8057
> 50000	96	119	150	196	267	384	600	1067	2401	9602

IOTC Species identification and sampling workshop, 2005

3. Sample size
Method 2: No preliminary data are available

- Sampling requirement at various accuracy level and data population size for landings survey

Accuracy (%)	90	91	92	93	94	95	96	97	98	99
Data Population size	**Safe sample size for BACs**									
300	29	35	43	54	69	90	120	163	218	274
400	30	36	44	56	73	97	133	188	267	356
500	30	37	45	58	75	102	143	208	308	432
600	30	37	46	59	77	106	150	223	343	505
700	31	37	47	60	79	108	156	236	373	574
800	31	38	47	60	80	110	160	246	400	640
900	31	38	47	61	81	112	164	255	424	703
1000	31	38	48	61	82	114	167	262	445	762
2000	32	39	49	63	85	120	182	302	572	1231
*** see notes 3000**	32	39	49	64	86	123	188	318	632	1549
4000	32	39	49	64	87	124	191	327	667	1778
5000	32	39	50	64	87	125	192	332	690	1952
6000	32	39	50	65	88	125	194	336	706	2088
7000	32	39	50	65	88	126	195	339	718	2197
8000	32	39	50	65	88	126	195	341	728	2286
9000	32	39	50	65	88	126	196	342	735	2361
10000	32	39	50	65	88	126	196	343	741	2425
15000	32	39	50	65	88	127	197	347	760	2638
20000	32	39	50	65	89	127	198	349	770	2760
25000	32	39	50	65	89	127	198	351	776	2838
30000	32	39	50	65	89	128	199	352	780	2893
35000	32	39	50	65	89	128	199	352	782	2933
40000	32	39	50	65	89	128	199	353	785	2964
45000	32	39	50	65	89	128	199	353	786	2989
50000	32	39	50	65	89	128	199	353	788	3009
> 50000	32	40	50	65	89	128	200	356	800	3201

IOTC Species identification and sampling workshop, 2005

Sample size exercise

Practical and operational recommendations:

- Sample 8 to 12 days a month per site
- Effort survey require more sample than landings survey
- At least 32 landings sampled by stratum
- If no previous data available:
  - cover all different types of landing sites;
  - within each stratum the samples could be allocated proportionally to the size of the landing sites.

4. Sample allocation

- How to distribute your sampling effort? For a certain target samples number.

- 3 methods
  - Proportional allocation (when not previous data are available)

    number of samples for each stratum is simply proportional to the total number of units of each stratum.

  - Neyman allocation

    or “optimum stratified sampling equation”: it uses the assumption that larger strata required larger sample and strata with hight large dispersion (= high standard deviation). To calculate it, variables or indicators of interest must be known from previous years.

  - Coefficient of Variation allocation

    it considers that the higher the variability of the strata is the bigger the sample must be. To calculate it, variables or indicators of interest must be known from previous years too.

Proportional allocation

Proportional allocation: number of samples for each stratum is simply proportional to the total number of units of each stratum.

$$Stratum\ sample\ size = \ Max.No.samples\ \frac{No.vessels\ in\ stratum}{Total\ no.vessels}$$

$$n_i = n\frac{N_i}{\sum_{i=1}^{n} N_i}$$

Neyman allocation

Neyman allocation: or "optimum stratified sampling equation": it uses the assumption that larger samples required larger strata and strata with hight large dispersion (= high standard deviation). To calculate it, variables or indicators of interest must be known from previous years.

$$n_i = n \frac{N_i s_i}{\sum_{i=1}^{n} N_i s_i}$$

Coefficient of variation allocation

Coefficient of variation (CV) allocation: it considers that the higher the variability of the stratum is the bigger the sample must be. To calculate it, variables or indicators of interest must be known from previous years too.

$$n_i = Ni \frac{CV_i}{\sum_{i=1}^{n} CV_i} \quad \text{with} \quad CV_i = \frac{s_i}{\bar{x}_i}$$

IOTC Species identification and sampling workshop, 2005

Exercises

3. Sampling protocols

- Sampling protocols are the instructions on how to collect data
- Sampling protocols must be documented
- Conditions in the field must be reviewed periodically, and protocols changed if necessary

- They are step-by-step instructions on how to collect data in a port, landing site, or vessel
- They are written by an expert who knows the areas well and understands the local conditions
- They may be revised if significant changes occur

Prepare your data collection tools

- Instruction for data collectors
- Data sheet

3. Sampling protocols

- It contains:
- The context why sample
- What
- How
  - The organisation of data collection
  - The forms
  - Equipment
  - All activities of data collector on daily and weekly basis
  - Practical advice
  - Species identification aid memory

Landings survey – What ?

- Focus on your objectives and needs
- Mandatory variables are very few and simple
- Do you have the capacity to analyse other information?

- Choose hard or soft copy

Part A - Document identification

Part B - Sampling Activity

Date
Landing site
Minor stratum
Recorder's name

Part C - Fishing Operation

Boat/gear type
Number of Units landing
Duration of trip
Total landing

Part D – Species information

Landing by species
Price
Number of fish in sample
Sum of landings by species (control total)

Part E – Supplementary information and remarks

IOTC Species identification and sampling workshop, 2005

Landing survey – collect catch data

- Taking the weight of the landings by species at the landing sites is the most common way to collect landings of small-scale fisheries.
- Data will be collected by enumerators at the landing sites. In the design of the landing surveys great care should be given not to overburden the enumerators, keep things simple!.
- The more complicated the survey, the higher the risks of sample errors, non-co-operation of the fishers etc. For example, Is it necessary to cover all species, or do we focus on species for which are covered by a fisheries management plan only?

Fishing activity survey

- Fishing activity survey consist in fisher interview
- “How many days have you fished during the last past x days ?”
- X depend on the fishing activity
  - 5 to 7 for short trips (one day trips)
  - 30 for long trips (20 days trips)

- Can be easily be included in landing survey – Data collector ask the question to the captain when collecting landings data

Boat activity survey

Table 8.10 Comparison of Results of Frame and Boat Activity Surveys

Sampling days	7	8	11	13	22	29
Homeport A						
Frame data		**10**		**10**	**10**	
Active		**7**		**3**	**5**	
Homeport B						
Frame data	**20**		**20**			**20**
Active	**4**		**6**			**20**

Based on the total number of active boats and the number of boats assumed to be present at each site, the standard approach for formulating BAC is as follows:

$$BAC = [(7+3+5)+(4+6+20)] / [(10+10+10)+(20+20+20)] = 45/90 = 0.5.$$

Other way to collect effort

- Household survey
- Telephone survey

Thank you for your attention

Questions?