

7. SAMPLING

The ideal basis for fish stock assessment is data that fully represent a stock, at least from the moment that it has recruited to the fishery, without any systematic errors or biases. Although it may not be possible in practice to obtain data of such quality, it should be the aim of any programme for the collection of data on a fishery, to obtain samples that fully represent the population under investigation, to know the sources of bias and to find means to correct for those biases.

The full theory of sampling is dealt with in many textbooks, which are widely available, such as Raj (1968), Som (1973) or Cochran (1977).

In the first part of this Chapter, some basic aspects of sampling will be discussed, with emphasis on random sampling. The second part deals with an example of a typical tropical demersal fishery, where many species are landed, partly sorted, for human consumption and partly mixed in the form of by-catch (for fish meal production or other purposes). The aim of this example is to show how complex a sampling scheme may have to be in order to get a representative sample of a particular fishery and how raising factors should be applied to reflect the total fishery for a particular species.

This example was created on the basis of experience in FAO/DANIDA follow-up courses on fish stock assessment, where many data sets on tropical fisheries turned out to be incomplete or biased, due to errors in the design and/or execution of sampling schemes (see Venema *et al.*, 1988).

Good sampling requires large and long-term investments in terms of manpower and general expenses. Therefore, it is important that sampling programmes are designed in such a way that they provide the data needed for the assessment and management of important species and fisheries, and that these data meet the standards as formulated by international and national working groups.

The sampling principles dealt with in this Chapter also apply to the procedures for sampling catches on board of research vessels, for which additional specific details on deck-sampling are given in Section 13.4.

7.1 SIMPLE RANDOM SAMPLING

Let us go back to the problem of estimating the mean length of a cohort as addressed already in Sections 2.2 and 2.3. An estimate is said to be "*unbiased*" if replicate estimates deviate from the true value in a **random manner only**. The "*true value*" is the parameter value we would get from measuring all specimens in the total population (see Section 2.3). An estimate is "*biased*" if it deviates from the true value in a **systematic manner**. With an unbiased estimate we can approach the true value as closely as we want by increasing the sample size. With a biased estimate there will always be a deviation between the true value and the estimate, and the deviation is independent of the sample size.

To obtain an unbiased estimate of the mean length the sample should be a "*random sample*", i.e. any fish from the stock considered should have exactly the same probability of being sampled. Assuming that it is possible to obtain a random sample (this is usually very difficult in practice), how many fish, n , would we need in the sample to obtain a pre-specified accuracy?

Suppose that we require an estimate of the mean length not deviating more than 7% from the true mean length and that we want to be 95% certain of this. We then would require the upper and lower 95% confidence limits not to deviate more than 7% from the estimated mean, \bar{x} . So the deviation $t_{n-1} * s / \sqrt{n}$ should have a maximum value of $0.07 * \bar{x}$, or

$$t_{n-1} * s / \sqrt{n} = 0.07 * \bar{x}$$

or more general:

$$t_{n-1} * s / \sqrt{n} = \epsilon * \bar{x}$$

where ϵ stands for "maximum relative error" (in this case $\epsilon = 0.07$).

Solving this equation with respect to n gives:

$$n = \left[\frac{t_{n-1} * s}{\epsilon * \bar{x}} \right]^2 \tag{7.1.1}$$

In order to apply Eq. 7.1.1 for estimating the required sample size we must already know the standard deviation, s , from previous samples.

We can also solve Eq. 7.1.1 with respect to ϵ , using, for example, the estimates of $s = 2.20$ and $\bar{x} = 15.07$ from Table 2.1.2.

$$\epsilon = \frac{t_{n-1} * 2.20}{15.07 * \sqrt{n}} = 0.146 * \frac{t_{n-1}}{\sqrt{n}}$$

In Fig. 7.1.1 the maximum relative error, ϵ , is shown as a function of the sample size, n . Note that we gain relatively little by increasing n when $n > 50$. An increase from $n = 10$ to $n = 20$, however, produces a reduction in relative error from 10.0% to 6.8%.

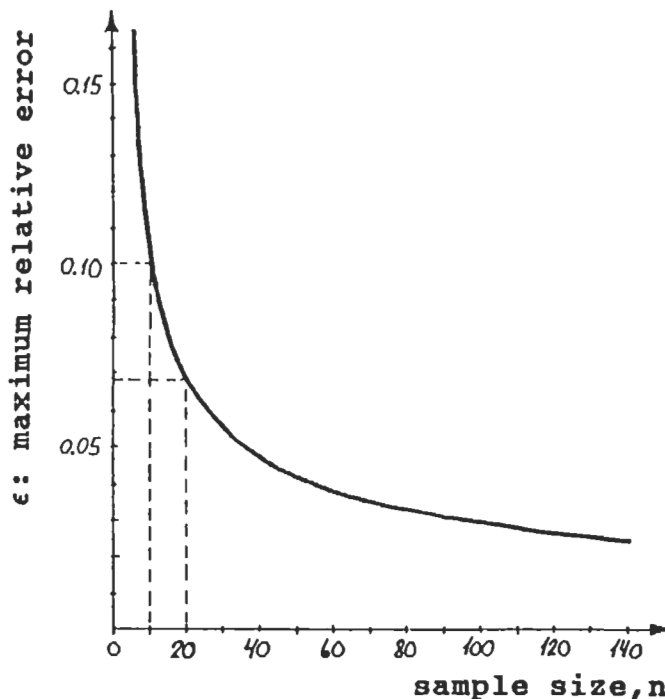


Fig. 7.1.1 Maximum relative error (ϵ) of the estimate of the mean length as a function of sample size (n) (using data from Table 2.1.2)

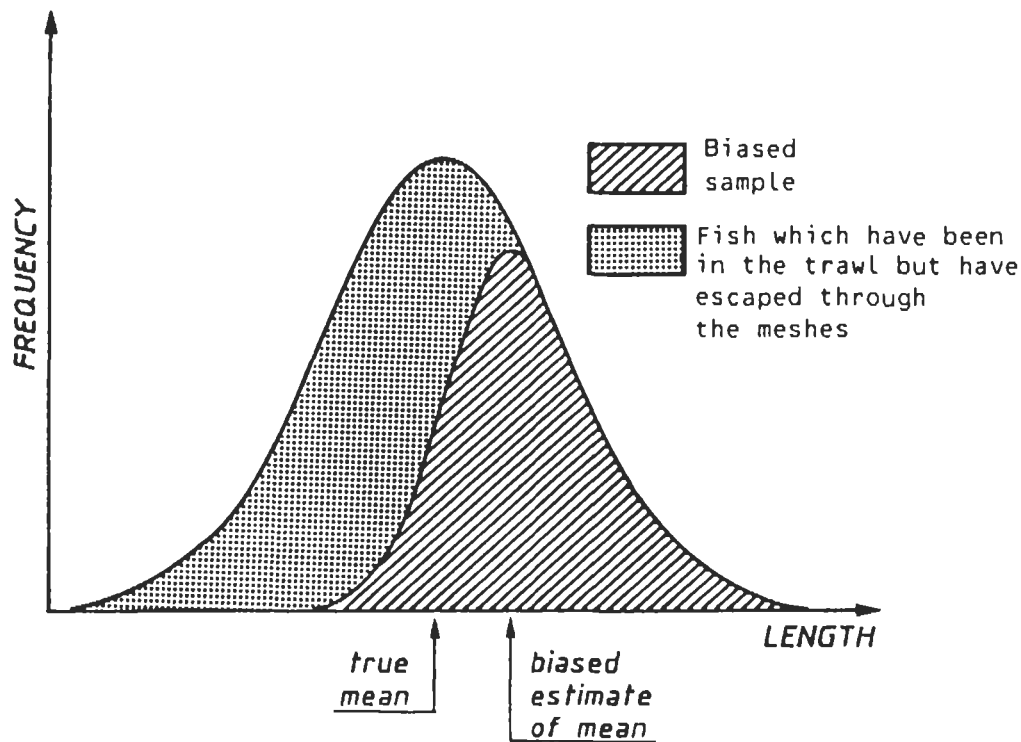


Fig. 7.1.2 Example of bias: bias caused by gear selection

If unbiased random samples can be obtained, there is no problem in estimating the required sample size for any pre-specified accuracy. However, samples are usually biased in one way or another. If small fish can escape through the meshes of a trawl we get an over-estimate of the mean length of the population of fish (see Fig. 7.1.2). This is one example of bias. If the larger fish can swim faster than the trawl is towed and thereby avoid being caught we have another type of bias.

However, once you are aware of the bias you can often adjust for it, and in the example of mesh selection you can try to estimate how many fish there would have been if all had been retained by the gear (see Chapter 6).

Another type of bias occurs if the spatial distribution of the population of fish is size-dependent. For example, when juvenile fish concentrate in certain nursery grounds and gradually migrate to the fishing grounds this may introduce a bias, if this migration pattern is not well understood. Bias caused by migration will be discussed in Chapter 11.

Another case of possible bias concerns the situation where it would be inefficient to measure the complete catch, for example from a large trawl haul. After sorting the catch by species you would take a sub-sample of, say, 100 fish of one species. If you selected the 100 fish one by one by simple "hand-picking" this would create a biased sample, since there always is a tendency to select the larger specimens. The proper procedure would be to put the catch of that species into boxes of approximately equal weight and then select some of these boxes at random for the sub-sample.

So far we have considered a population consisting of a very large number of fish, so that the sample formed an insignificant part of the population. Some "populations" in fisheries data collection, however, consist of a small number of units, so that a sample constitutes a significant part of it. This is the case when the sampled "population" is not fish, but for example the landing places in a certain area.

Suppose the purpose of a sampling scheme is to estimate the mean landings per landing place during a certain period, say, a month, and suppose that the total number of landing places is 100. If we sampled all 100 landing places then the variance of our estimate of the mean landings would be zero. In that case we simply have the true population mean. If we have personnel for sampling only 50 landing places there would be a certain variance in our estimate. However, this variance would not be s^2 (as defined by Eq. 2.1.2), since only 50% of the population is left to produce the variance, as distinct from a length-frequency sample where practically 100% of the population would still be left.

We allow for this by applying the so-called "*finite population correction factor*":

$$(1 - n/N)$$

where N is the population size ($N = 100$ in the example) and n the sample size. Let $Y(i)$ be the landings at landing place no. i of the sample, $i = 1, 2, \dots, n$, and let \bar{Y} be the mean landings of all landing places, then the estimate of the mean landings is:

$$\bar{Y} = \frac{1}{n} * \sum_{i=1}^n Y(i)$$

The variance of the estimate then becomes (see Eqs. 2.3.2 and 2.1.2):

$$\text{VAR}(\bar{Y}) = \frac{1 - n/N}{n} * s^2 \quad (7.1.2)$$

where

$$s^2 = \frac{1}{n-1} * \sum_{i=1}^n (Y(i) - \bar{Y})^2 \quad (7.1.3)$$

The confidence interval of \bar{Y} is (see Eq. 2.3.1):

$$\left[\bar{Y} - t_{n-1} * s * \sqrt{(1/n - 1/N)}, \bar{Y} + t_{n-1} * s * \sqrt{(1/n - 1/N)} \right]$$

The estimate of the total landings is:

$$Y = N * \bar{Y} \quad (7.1.4)$$

and its variance is

$$\text{VAR}(Y) = N^2 * \text{VAR}(\bar{Y}) \quad (7.1.5)$$

Y is usually the quantity we are interested in. Eq. 7.1.5 follows from the general rule for a random variable, Eq. 2.3.3, where N is a constant.

Note that Eq. 7.1.2 is general in the sense that it also applies to large (in practice infinite) populations as the finite population correction factor, $(1-n/N)$, becomes 1.0 when N is infinite.

Often Eq. 7.1.4 is expressed as:

$$Y = \frac{N}{n} * \sum_{i=1}^n Y(i) \quad (7.1.6)$$

and we say that the sample has been raised to the total, Y , by application of the "*raising factor*" N/n .

7.2 STRATIFIED RANDOM SAMPLING

Consider again the problem of estimating the total landings, Y , from the 100 landing places during a certain month, as dealt with in Section 7.1. Assume that a sampling programme has been conducted in the previous years, based on which the 100 landing places have been divided into three categories as shown in Table 7.2.1. Such a division of the total population is called a "stratification" and the categories (large, medium, small) are called "strata". Table 7.2.2 shows a numerical example (from Gulland, 1966) corresponding to Table 7.2.1. To obtain an estimate of the standard deviation within each stratum, $s(j)$, $j = 1, 2, 3, \dots$ a survey covering all landing places was carried out during one month.

The standard deviation, $s(j)$, in Tables 7.2.1 and 7.2.2 is the square root of the corresponding variance:

$$s(j)^2 = \frac{1}{N(j)-1} * \sum_{i=1}^{N(j)} \{Y(j,i) - \bar{Y}(j)\}^2 \quad (7.2.1)$$

where the stratum mean is:

$$\bar{Y}(j) = \frac{1}{N(j)} * \sum_{i=1}^{N(j)} Y(j,i) \quad (7.2.2)$$

Note that Eqs. 7.2.1 and 7.2.2 produce the true parameters for the particular month in which the data were collected.

Table 7.2.1 An example of stratification (from Gulland, 1966)

landing place category or stratum	number of landing places by stratum	average landings by stratum	standard deviation within each stratum
1 large	$N(1)$	$\bar{Y}(1)$	$s(1)$
2 medium	$N(2)$	$\bar{Y}(2)$	$s(2)$
3 small	$N(3)$	$\bar{Y}(3)$	$s(3)$

Table 7.2.2 Numerical example of a stratification based on samples from one month (from Gulland, 1966)

$Y(j,i)$ = landing at landing place no. i in stratum j										
LARGE LANDING PLACES: $Y(1,i)$										
$N(1) = 10$ $Y(1) = 42.3$ $s(1) = 28.91$	45	59	87	41	71	25	9	69	10	7
MEDIUM LANDING PLACES: $Y(2,i)$										
$N(2) = 30$ $Y(2) = 15.2$ $s(2) = 8.57$	17	13	19	26	1	8	27	11	12	26
	5	8	10	16	16	4	16	16	13	29
	14	25	29	27	20	25	2	7	3	12
SMALL LANDING PLACES: $Y(3,i)$										
$N(3) = 60$ $Y(3) = 4.2$ $s(3) = 2.81$	2	6	7	0	1	2	1	5	4	7
	8	9	3	2	5	4	2	0	2	8
	5	3	8	9	8	9	1	6	5	3
	3	4	7	5	5	3	2	4	6	1
	6	2	5	1	0	3	8	0	4	3
	3	5	5	0	7	0	9	7	9	0

Often we are not in a position to sample all landing places. Let us assume that we only have manpower etc. available for a sample of n ($n < 100$) landing places. The sample size can be written as the sum:

$$n = n(1) + n(2) + n(3)$$

where $n(i)$ is the number of landing places sampled in stratum no. i . To "*design a sampling scheme*" means to decide which sample sizes, $n(1)$, $n(2)$ and $n(3)$, should be applied to each of the three strata.

One could now ask why we should complicate the sampling by the introduction of strata. The answer is that we (nearly) always obtain a more precise estimate of the population mean from stratified sampling than from simple random sampling. How much we gain depends on the choice of strata. If the observations within a stratum are approximately of the same size (as is the case in Table 7.2.2) we are likely to gain from the stratification. The less variation there is within a stratum, the more is gained by stratification. On the other hand, for practical reasons there is a limit to the number of strata that can be selected.

The basic rules of stratified sampling are that the sample size per stratum, $n(j)$, should be large when:

- 1) The stratum is large (if $N(j)$ is large)
- 2) The standard deviation $s(j)$ is large

To these rules we can add for economic reasons that samples should be large when:

- 3) Sampling is inexpensive

Mathematically the first two conditions can be expressed as

$$n(j) \text{ proportional to } N(j) * s(j) \tag{7.2.3}$$

or

$$n(j) = n * \frac{N(j) * s(j)}{\sum N(j) * s(j)} \tag{7.2.4}$$

where n is the total sample size. This formula is called the "*optimum stratified sampling equation*" (or "*Neyman allocation*").

The first two conditions will be dealt with below in Example 25 and the third condition in Example 26.

Example 25: Stratified random sampling

Let us assume that we have funds available to collect 100 observations, that there are two strata and that their sizes, $N(j)$, and standard deviations, $s(j)$, are:

	Stratum 1	Stratum 2
$N(j)$	1000	2000
$s(j)$	50	10

Then, the 100 observations should be allocated to the two strata according to Eq. 7.2.4 as follows:

$$\text{Stratum 1: } \frac{1000 \cdot 50}{1000 \cdot 50 + 2000 \cdot 10} * 100 = 71.4 \text{ say } 71$$

$$\text{Stratum 2: } \frac{2000 \cdot 10}{1000 \cdot 50 + 2000 \cdot 10} * 100 = 28.6 \text{ say } 29$$

Usually the budget available for the execution of a sampling programme is limited. The cost of taking a sample will differ from stratum to stratum and it is therefore possible also to take the cost of sampling into account, when designing a sampling scheme.

Let $c(j)$ be the cost of taking a sample unit in stratum j and let C_0 be the additional fixed cost of the whole sampling programme. The total cost, C , for m strata then becomes:

$$C = C_0 + \sum_{j=1}^m c(j) * n(j) \quad (7.2.5)$$

It can be shown that for an optimum allocation the sample size should be proportional to $1/\sqrt{c(j)}$.

An optimum use of the available resources is then achieved (see Eq. 7.2.3) when the sample size is proportional to

$$N(j) * s(j) / \sqrt{c(j)}, \text{ or if}$$

$$n(j) = n * \frac{N(j) * s(j) / \sqrt{c(j)}}{\sum N(j) * s(j) / \sqrt{c(j)}} \quad (7.2.6)$$

where n is the total sample size. Using the criterion Eq. 7.2.6, we minimize the variance of the estimate of the total landings, Y .

If the total budget available is C and the fixed cost is C_0 then the total number of sample units in all strata is given by:

$$n = [C - C_0] * \frac{\sum N(j) * s(j) / \sqrt{c(j)}}{\sum N(j) * s(j) * \sqrt{c(j)}} \quad (7.2.7)$$

and the number of units sampled in stratum j is given by:

$$n(j) = [C - C_0] * \frac{N(j) * s(j) / \sqrt{c(j)}}{\sum N(j) * s(j) * \sqrt{c(j)}} \quad (7.2.8)$$

Example 26: Stratified random sampling, considering costs

Let the total budget for the sampling programme given in Example 25 be 1800 (money units) while the fixed cost is 444 and the prices per sample unit are:

Stratum 1: 16
Stratum 2: 9

Then the total number of samples, n , that can be taken with the available budget is determined by Eq. 7.2.7:

$$n = (1800-444) * \frac{1000*50/\sqrt{16} + 2000*10/\sqrt{9}}{1000*50*\sqrt{16} + 2000*10*\sqrt{9}} = 99.96, \text{ say } 100$$

These 100 sample units should then be allocated to stratum 1 and 2 respectively as follows:

$$n(1) = (1800-444) * \frac{1000*50/\sqrt{16}}{1000*50*\sqrt{16} + 2000*10*\sqrt{9}} = 65.19, \text{ say } 65$$

$$n(2) = (1800-444) * \frac{2000*10/\sqrt{9}}{1000*50*\sqrt{16} + 2000*10*\sqrt{9}} = 34.77, \text{ say } 35$$

Note that now 35 sample units are allocated to the inexpensive stratum 2, compared to 29 in Example 25, where the price per sample unit was not taken into account (or assumed to be the same for both strata).

Hereby we conclude the theory for optimum allocation and turn to the question of how the estimates of mean values, the totals and their variances can be calculated. Although the theory given below is general, you may again think of the example given in Table 7.2.2. Having determined the sample sizes, $n(j)$, we obtain an estimate of the mean landings of each stratum, $\bar{Y}(j)$, by:

$$\bar{Y}(j) = \frac{1}{n(j)} * \sum_{i=1}^{n(j)} Y(j, i) \tag{7.2.9}$$

Let m be the total number of strata ($j = 1, 2, \dots, m$), then the estimate of the total population mean i.e. the mean landings in all strata, is:

$$\bar{Y} = \sum_{j=1}^m \frac{N(j)}{N} * \bar{Y}(j) \tag{7.2.10}$$

and finally, the estimate of the total landings, Y , (see Eq. 7.1.4) is:

$$Y = N * \bar{Y}$$

Let the symbol "VARst" stand for the variance of an estimate obtained from stratified ("st") random sampling. The estimate of the variance of the total population mean, \bar{Y} , is:

$$\text{VARst}(\bar{Y}) = \sum_{j=1}^m \left[\frac{N(j)^2}{N^2} * \text{VAR}(\bar{Y}(j)) \right] \tag{7.2.11}$$

where the estimate of the variance of the estimate of the mean for each stratum Y , $\text{VAR}(\bar{Y}(j))$, is defined by Eq. 7.1.2:

$$\text{VAR}(\bar{Y}(j)) = \frac{1 - n(j)/N(j)}{n(j)} * s(j)^2 \quad (7.2.12)$$

Note the finite population correction factor: $1 - n(j)/N(j)$

Eq. 7.2.11 follows from the general rules for the variance of a sum of independent random variables given in Eqs. 2.3.3 and 2.3.4.

The variance within stratum j as defined by Eq. 7.1.3 is:

$$s(j)^2 = \frac{1}{n(j)-1} * \sum_{i=1}^{n(j)} [Y(j,i) - \bar{Y}(j)]^2 \quad (7.2.13)$$

Inserting Eq. 7.2.12 into Eq. 7.2.11 the latter may also be written:

$$\text{VAR}_{st}(\bar{Y}) = \frac{1}{N^2} * \sum_{j=1}^m \frac{N(j) * [N(j) - n(j)]}{n(j)} * s(j)^2 \quad (7.2.14)$$

Eq. 7.2.14 is more convenient from a calculation point of view (see Exercise 7.2).

The variance of the total landings in stratum j is:

$$\text{VAR}(Y(j)) = N(j)^2 * \text{VAR}(\bar{Y}(j)) \quad (7.2.15)$$

and the variance of the total catch (all strata) is:

$$\text{VAR}_{st}(Y) = N^2 * \text{VAR}_{st}(\bar{Y}) \quad (7.2.16)$$

In the examples given above the strata consisted of different sizes of landing places in terms of landed weight. Stratifications can also be based on other criteria, for example:

- Gear types
- Boat types
- Fishing seasons
- Fishing grounds
- Species or species groups
- Commercial size categories

Usually one can take advantage of whatever stratification is available. Sometimes we are forced to stratify the sampling, as might be the case when fish are landed already sorted into commercial size categories. Sometimes we will have to do extra work to obtain the stratification. This may apply to a stratification by boat type. If we sample from an auction hall at the time the fish are sold, we may have to trace the boat which caught the sample. If we are to invest extra resources (time, funds and manpower) in obtaining a stratification the increased precision obtained from that stratification should be considered relative to the cost.

7.3 PROPORTIONAL SAMPLING

In some cases (e.g. when starting up) we may not know the variance within strata but only the stratum size, $N(j)$. In that case it is recommended to use "*proportional sampling*", i.e. to allocate samples in proportion to the stratum size. Applying proportional sampling to the 100 observations of Example 25 given above, the allocation to the two strata should be in the proportions:

$$\text{Stratum 1: } \frac{1000}{1000+2000} * 100 = 33.3 \text{ say } 33$$

$$\text{Stratum 2: } \frac{2000}{1000+2000} * 100 = 66.7 \text{ say } 67$$

Compare with the values obtained by optimum allocation in Example 25, respectively, 71 for stratum 1 and 29 for stratum 2.

Note that proportional sampling is only identical to optimum stratified sampling in the exceptional case that the variances of all strata are equal.

(See Exercise(s) in Part 2.)

7.4 SAMPLING COMMERCIAL CATCHES

In order to be able to carry out assessments of exploited stocks we must have adequate data for each species under investigation. We must know the total weight of the catch and the length and/or age composition of the catch of each stock. In order to obtain this kind of data it will be necessary to sample the landings of the commercial fisheries, according to a pre-determined scheme. Such a sampling scheme should take into account the following factors:

- 1) the total area of distribution of the stock of a species and
- 2) all the fishing activities taking place in that area, which are catching that particular species. These may include different types of boats (fleets) and different gears. The fleets may either be completely national, or include also those of other countries exploiting the same stock.

Since stocks of fish may occupy areas across international boundaries, the data collection should also be set up in such a way that merging of data collected by different countries is possible. In such cases it is essential that agreements on data requirements are made between countries, for example, through international fisheries bodies, and that these agreements are put into practice through the creation of international working groups. The same criterion applies in the case of large countries where several institutes or sub-institutes are covering the landings of a particular stock.

The data collected should be verified and entered in a computerised data base, which should be accessible to all scientists who have an interest in the data and who are authorised to use them. Data collected through such sampling schemes should never be considered as the private property of the scientists who were responsible for their collection.

Because of the complexity of many fisheries, in particular in the tropics, where many different gears are used to catch a mixture of species and where the landings are often spread over many different sites, it is important to design a sampling scheme that is based on an in-depth knowledge of the fisheries on a particular species. When more than one species has been selected for sampling, it may be possible to combine sampling schemes, if there is sufficient overlap in landing sites, fleets etc.

Sampling schemes, once set up, and in particular when international agreements are involved, should have secured funding over very long periods, in some cases practically forever.

Sampling for stock assessment purposes is very closely connected to sampling or total enumeration systems set up by fishery statisticians. While the fishery statistician is mainly interested in obtaining an estimate of the catches of all important species, usually by type of gear and boat, the fishery scientist will usually concentrate on a smaller number of species. Biological sampling is time-consuming and it is therefore not possible in most cases to sample all the landings in a particular place. However, also for stock assessment purposes it is necessary to estimate how much fish of a particular species has been landed by the vessels that were not sampled, both in the landing places where biological sampling did take place and in all other places. The above makes it obvious that the work of the fishery statisticians is extremely important for the biologists, since the general statistics will be needed to determine the so-called "*raising factors*", which consist of the ratio between the total number of units and the sampled number of units. Raising factors are therefore used to raise the data (e.g. length-frequencies) obtained from the sample to the level of the total catches. When the samples are small relative to the total landings, the raising factor may become very large and if the samples are biased for one reason or another this will be reflected at a larger scale in the totals.

Therefore, it is very important that the samples are random samples and that they represent a reasonable proportion of the total catch. Rather than thinking in terms of measuring hundreds of fish, plans should be made to measure many thousands of fish. Only then the data may provide a reliable basis for stock assessments. In the example of whiting given in Table 4.4.3.1, the total catches of whiting of all year classes (cohorts) were estimated to be 2,021,800,000 fish, while the number of whiting otoliths sampled in all the countries exploiting this resource and used by the international working group was approximately 10,000. The overall raising factor from samples to total catch was therefore approximately 200,000.

In Table 5.1.1, the total estimated number of survivors from the 1974-cohort is 9,856,600,000 whiting, or about 3.4 times more than the numbers caught, so the number actually sampled is only a very small fraction of the real population in the sea, despite enormous efforts.

The basis for all sampling schemes is a decision on the species to be sampled. This decision will usually be determined by the economic importance of the species (or species groups) and in the case of resources being exploited by more than one country, on the requirements of international working groups.

Assuming that the general fishing pattern of the commercial fisheries is known for each of the selected species, it will be possible to determine which landing places and fleets should be sampled each season.

On that basis, and on that of the availability of funds and personnel, the system can be designed and tested.

The complexity of sampling schemes and the various raising factors involved are illustrated in the following theoretical example, which has many features of a tropical fishery for shrimp and demersal fish.

Example 27: Sampling scheme for a tropical demersal fishery

The sampling scheme illustrated in the example has two major objectives:

- 1) To determine the total catch in weight of one species "s" (Section 7.5) and
- 2) To determine the length composition of the total catch of species "s" (Section 7.6).

It is assumed that all specimens of species "s" caught originate from one stock and that all the landings from this stock are covered by the sampling scheme.

The physical features of the example are shown in Fig. 7.4.1. We assume the stock to be confined to one fishing ground exploited only by the boats from three landing places denoted by the index h and labelled, I, II and III. There are three different types of boats, for example:

- a) large trawlers,
- b) gill netters and
- c) "baby"-trawlers

Each landing place is the home port to a number of boats of each type. A batch of similar boats is called a fleet (index f). Because locations of landing places are factors of practical importance in a sampling scheme, each fleet is also distinguished according to its landing place, so that in this case we actually operate with nine sets of boats, which are distinguished by combinations of f and h, for example, aI, bII, bIII etc. In Fig. 7.4.1 the fleets have been assigned to their respective landing places, while in Table 7.4.1 they have been arranged by type of boat in order to facilitate the calculation of catch per unit of effort (CPUE).

Table 7.4.1 The procedure for estimating total catch in weight for one species, s, during one period, t (compare Fig. 7.4.1)

TOTAL CATCHES OF SPECIES "s" DURING TIME PERIOD "t" IN WEIGHT UNITS							
fleet	landing place	effort	observed catch	CPUE by fleet	estimated catch	total catch by fleet and landing place	total catch by fleet
f	h	E (t, f, h)	W (s, t, f, h)	$\frac{\Sigma W}{\Sigma E}$	$\overline{CPUE} * E$	W(s, t, f, h)	W(s, t, f, *)
a	I	15	57	$\frac{57+83}{15+20} = 4$	-	57	172
a	II	20	83		-	83	
a	III	8	-		$4 * 8 = 32$	32	
b	I	25	55	$\frac{55+105}{25+55} = 2$	-	55	184
b	II	55	105		-	105	
b	III	12	-		$2 * 12 = 24$	24	
c	I	20	-	$\frac{25}{25} = 1$	$1 * 20 = 20$	20	75
c	II	25	25		-	25	
c	III	30	-		$1 * 30 = 30$	30	
estimated grand total catch: W(s, t, *, *) =							431

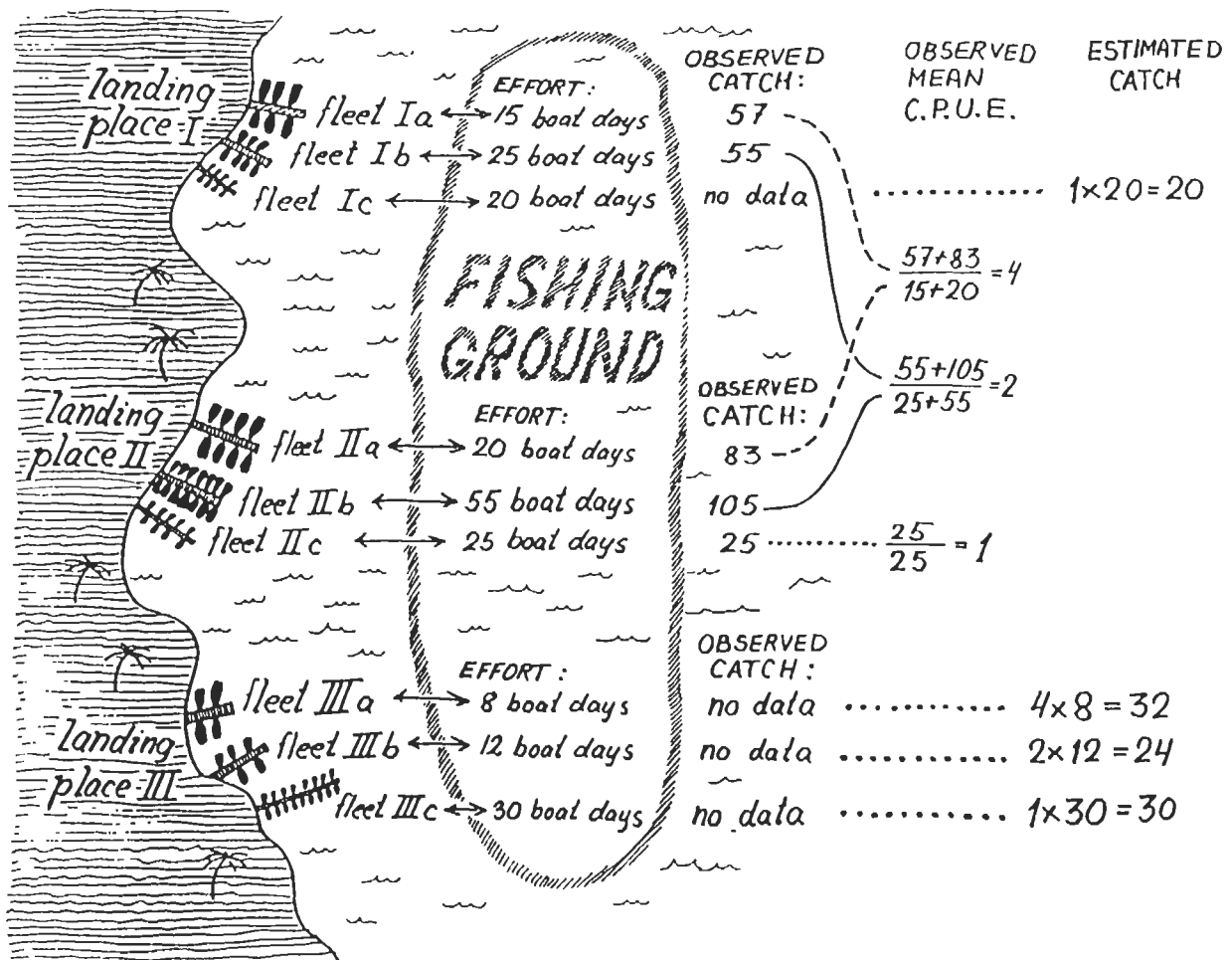


Fig. 7.4.1 Hypothetical example to illustrate the estimation of the total catch in weight of species, s , during a time period, t , say a quarter of a year

The next important assumption we make is that data are available on the total fishing effort, E , of each fleet in each landing place. In the present example effort has been measured as the number of boat-days during the period of time considered. (Note that a number of alternative measures of effort could have been used. In the worst case only the number of boats in each fleet is known. In that case one must assume an average number of effort units per boat per time unit. If we know, for example, the average number of boat-days per fleet per month for one landing place, then this figure can be applied also to landing places for which this information is not available.)

The number of effort units expended during time period t by fleet f from landing place h is denoted:

$$E(t, f, h) = \text{effort}$$

The observed (or estimated) values of the effort $E(t, f, h)$ of the nine sets of boats are shown in Table 7.4.1 and Fig. 7.4.1. For the moment we confine the description of the sampling scheme to a single time period, for example the second quarter of 1978, which makes the index " t " a constant. Later, data from different time periods will be combined and then t becomes a variable index.

7.5 ESTIMATION OF THE TOTAL CATCH IN WEIGHT OF A CERTAIN SPECIES

Assume that due to shortage of funds and personnel the total catches have not been recorded for 1) any of the fleets of landing place III, 2) for fleet c of landing place I. The observed total catches for the remaining components of the fishery are shown in Table 7.4.1 and Fig. 7.4.1. We denote by

$$W(s, t, f, h)$$

the total weight of species "s" landed during period "t" by fleet "f" at landing place "h".

The available data on effort and catch can now be combined to calculate the mean CPUE per boat type and these figures can then be used to make a complete estimate of the total catch of all fleets. (Note that CPUE is measured in weight and not in numbers of fish, compare Section 4.3.)

The formula is simple:

$$\text{effort} * \text{CPUE} = \text{estimated catch}$$

the calculations and the results are presented in Fig. 7.4.1 and Table 7.4.1.

To avoid too many summation signs in the following calculations we introduce a more convenient notation. Whenever an index is replaced by "*" it means the sum over the index in question. For example, the total weight of the three landing places "h" of species "s" in period "t" by fleet "f" can be denoted as:

$$W(s, t, f, *) \text{ which is the same as } \sum_{h=1}^3 W(s, t, f, h)$$

$$\text{and } \sum_{f=1}^3 \sum_{h=1}^3 W(s, t, f, h) = \sum_{f=1}^3 W(s, t, f, *) = W(s, t, *, *)$$

which is the total catch of species s during time period t landed by all fleets to all landing places (Table 7.4.1).

7.6 ESTIMATION OF THE LENGTH COMPOSITION OF A CERTAIN SPECIES IN THE TOTAL CATCH

We now turn to the main objective, the derivation of the length composition of species s, i.e. the derivation of inputs to the age-based or length-based cohort analyses. First some general features of sampling species s for length composition. Like in most sampling procedures we have to raise the outcome of the sampling to the total catch of the boat and eventually to the total catch of all the fleets. Therefore it is useful to introduce the suffix "m" for quantities that have been sampled. Another useful thing is to distinguish quantities of the sample from those of the total sampled catch. This is done by using capital letters for the total catch and small letters for the sample, e.g.

W_m = the **total weight** of the **sampled catch**

w_m = the **weight** of the **sample**

C_m = the total number of fish in the sampled catch

c_m = the number of fish in the sample.

A sampling procedure usually starts with sorting of the catch by species and weighing or estimating the total weight caught of each species. The next step is to select a sample at random, to weigh the sample, and after that to measure the length of all the fish in the sample.

In some cases it will be necessary to take sub-samples before it is possible to measure the length of a particular species. This is the case with the by-catch category as will be demonstrated below.

In the present example the basic sample for estimation of length compositions is associated with a "trip". Each fleet from each landing place makes a number of trips during the time period considered. In Fig. 7.6.1 fleet a in landing place I is considered as an example. During period t , 15 trips were carried out and at the end of four of those 15 trips samples were collected on the jetty when the catch was landed. (Note, usually, it is impossible to take samples from all trips, in this case only four out of the 15 trips were sampled. Care should be taken that these are selected at random, see the discussion in Section 7.1.)

Let the number of samples be denoted:

$$n(t, f, h)$$

In the example of Fig. 7.6.1: $n(t, a, I) = 4$. Each sample is given a number, or an index, j , for our internal book-keeping:

$$j = 1, 2, \dots, n(t, f, h) \quad (\text{see Fig. 7.6.1})$$

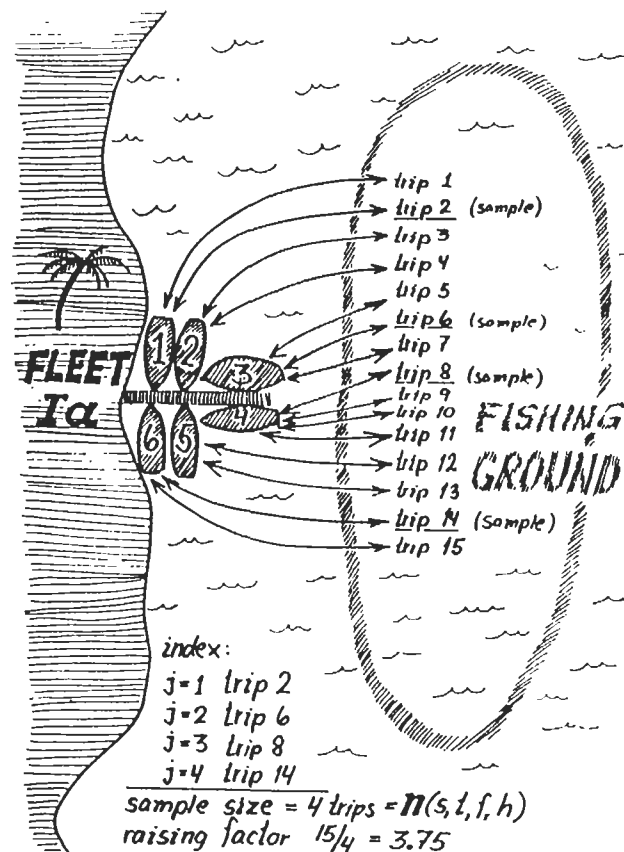


Fig. 7.6.1 Illustration of sampling for the estimation of the length composition of the catch of one fleet from one landing place

The often complex situation of sampling a single vessel after a trip is illustrated in Fig. 7.6.2 In this case the catch is assumed to consist of two major categories:

1. **FISH FOR DIRECT HUMAN CONSUMPTION or CONSUMPTION FISH.** This part is sorted into species (or species groups) and contains the marketable sizes.
2. **BY-CATCH.** This part is not sorted by the fishermen. It contains fish not used for human consumption, including valuable fish species below marketable sizes. The quantities referring to the by-catch category are distinguished from those of the consumption fish category by the suffix "b". The total weight of all the by-catch of a sampled trip is denoted by $W_{bm}(*,t,f,h,j)$.

Of the species selected for a sampling programme, samples have to be taken from both categories, since from a biological point of view they are of equal importance. We will first deal with the sampling of the category consumption fish, then with the by-catch and then combine the two sets of data. The whole procedure has been illustrated in Figs. 7.6.2, 7.6.3 and 7.6.4.

Sampling the catch for human consumption from one trip

The general sampling procedure for length-frequency data is given below, based on the example of the consumption category of species s :

- 1) The weight of the total sampled catch of species s , of the consumption fish category, is recorded:

$W_{m}(s,t,f,h,j)$ = total weight of consumption catch of species s of sampled trip j .

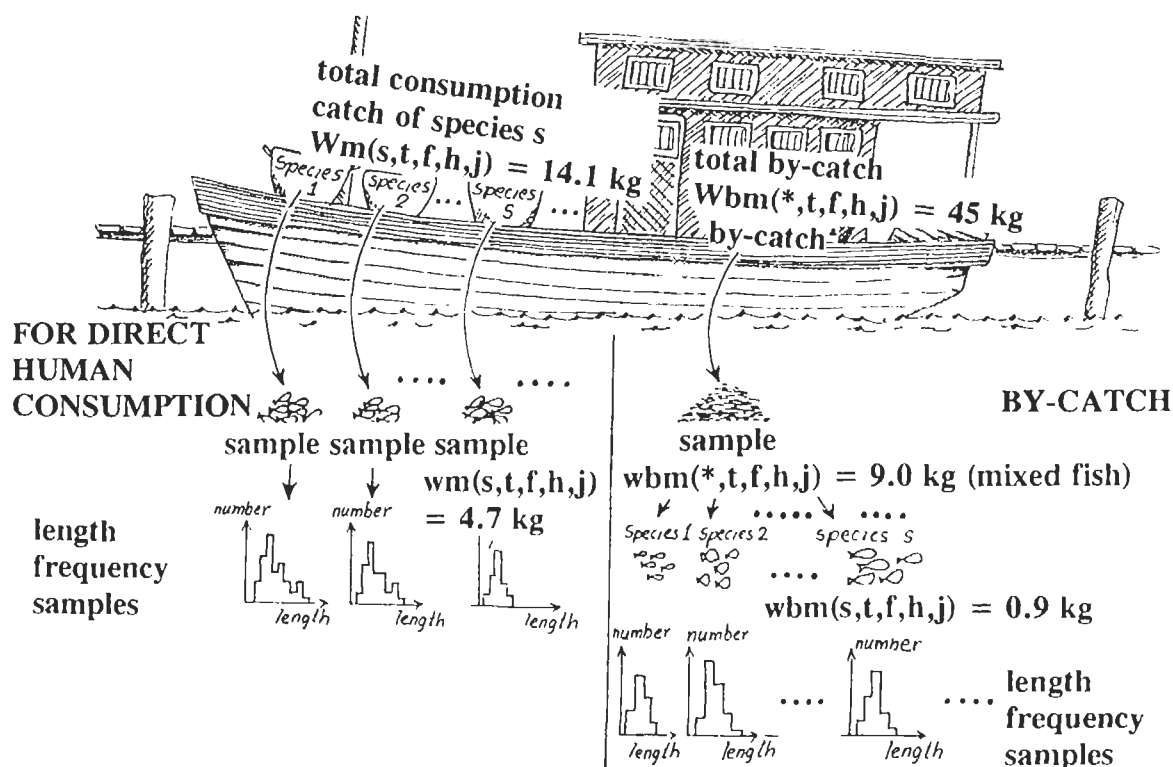


Fig. 7.6.2 Sampling from a single trip (for further explanation, see text)

Consumption fish:

Total weight of sampled catch: $W_m = 14.1 \text{ kg}$
 Total weight of sample: $w_m = 4.7 \text{ kg}$
 Raising factor: $W_m/w_m = 14.1/4.7 = 3$

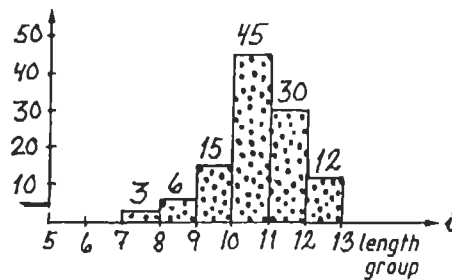
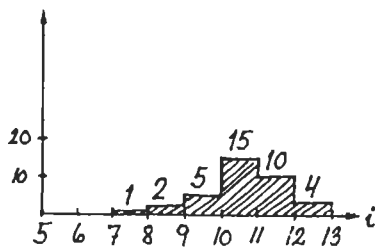
By-catch:

Total weight of sampled by-catch: $W_{bm} = 45 \text{ kg}$
 Total weight of sample (all species): $w_{bm} = 9 \text{ kg}$
 Raising factor: $W_{bm}/w_{bm} = 45/9 = 5$

Consumption fish (species s)

length-frequency sample
 $cm(s,t,f,h,j,i)$
 sample size = 37

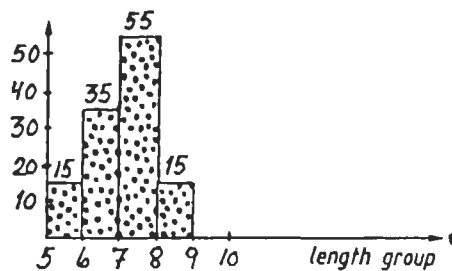
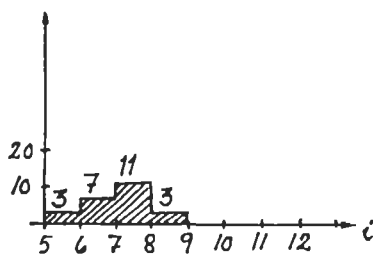
raised length-frequency sample
 $cm * W_m/w_m = cm * 3$
 $37 * 3 = 111$



By-catch (species s)

length-frequency sample
 $cbm(s,t,f,h,j,i)$
 sample size = 24

raised length-frequency sample
 $cbm * W_{bm}/w_{bm} = cbm * 5$
 $24 * 5 = 120$



Total catch (species s)

estimated length-frequency of species s in the total catch of the sampled trip
 $C_m(s,t,f,h,j,i)$

combined length-frequency sample
 $C_m = cm * 3 + cbm * 5 = 111 + 120 = 231$

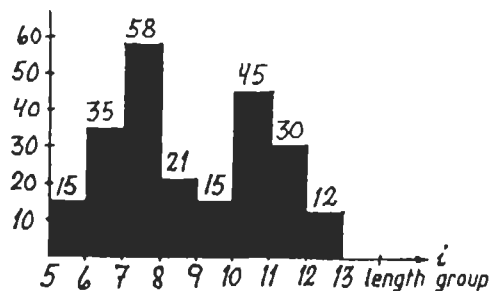


Fig. 7.6.3 Combining length-frequency samples of species s encountered in the consumption fish and by-catch categories of the landings of a single trip (see also Fig. 7.6.2)

2) A random sample is taken and the weight of the sample is recorded as follows:

$wm(s,t,f,h,j)$ = weight of all specimens of species s in sample j .

3) The sample is then measured for length. Let i be the index of a length group, then we denote by

$cm(s,t,f,h,j,i)$

the number of fish of species s in length group i , of sample j , from landing place h , caught by fleet f , in time period t .

The total number of fish of all length groups in the sample is denoted by

$cm(s,t,f,h,j,*)$

In Fig. 7.6.2 is shown that of species s a sample of 4.7 kg = $wm(s,t,f,h,j)$ was taken from a total catch of 14.1 kg = $Wm(s,t,f,h,j)$.

The sample was measured for length and frequencies were obtained as depicted in Fig. 7.6.3, with a total sample size of 37 specimens = $cm(s,t,f,h,j,*)$.

This length-frequency sample has to be raised to the total catch of the boat, by raising each frequency by a "raising factor", which is simply the total weight of the catch of species s divided by the weight of the sample

$$\frac{Wm}{wm}$$

In the case of species s the total estimated number caught, in the consumption category, is

$$\frac{Wm(s,t,f,h,j)}{wm(s,t,f,h,j)} * cm(s,t,f,h,j,*) = \frac{14.1}{4.7} * 37 = 111$$

Sampling the by-catch from one trip

The sampling procedure for the by-catch includes an extra step, namely sorting into species. Let the total weight of the by-catch be

$Wbm(*,t,f,h,j)$ kg

(Index "s" is replaced by "*" because the by-catch is not sorted into species.)

From the Wbm kg a sample is taken of

$wbm(*,t,f,h,j)$ kg

This sample is then separated into species (see Fig. 7.6.2). We are only interested in species s . The weight of species s from the by-catch sample is

$wbm(s,t,f,h,j)$ kg

(Note that this is not the weight to be used to raise the sample to the total catch.)

These fish are measured and the lengths are recorded (see Fig. 7.6.3). The number of fish in length group i is denoted as:

$$cbm(s, t, f, h, j, i)$$

and the total number of species s in the sample is

$$cbm(s, t, f, h, j, *)$$

In this case the number is 24 (see Fig. 7.6.3).

These numbers are raised to account for the total by-catch of the boat trip by a raising factor, consisting of the total weight of the by-catch and the weight of the total sample of by-catch (and **not** the weight of species s only):

$$\frac{W_{bm}}{w_{bm}}$$

$$\frac{W_{bm}(*, t, f, h, j)}{w_{bm}(*, t, f, h, j)} * cbm(s, t, f, h, j, i)$$

gives the length-frequencies of species s that represent the entire by-catch.

In the example (see Fig. 7.6.2) $W_{bm}(*, t, f, h, j) = 45$ kg and the sample weight $w_{bm}(*, t, f, h, j) = 9$ kg so that the raising factor becomes $W_{bm}/w_{bm} = 45/9 = 5$, and the total number of specimens of species s in the by-catch category is $5 * 24 = 120$.

Combining consumption fish sample and by-catch sample from one trip

We now have to combine the raised length-frequencies of the two categories, in order to obtain a complete picture of the length-frequencies of species s in the catch of the sampled trip. This is done by a simple summation of the two raised frequencies.

An estimate of the **total catch in numbers by length group**, C , is obtained by simple addition of consumption and by-catch estimates:

$$\frac{W_m}{w_m} * cm(s, t, f, h, j, i) + \frac{W_{bm}}{w_{bm}} * cbm(s, t, f, h, j, i) = C_m(s, t, f, h, j, i)$$

(see Fig. 7.6.3).

For example, the estimated total number for length groups 8-9 cm becomes

$$\frac{14.7}{4.7} * 2 + \frac{45}{9} * 3 = 3 * 2 + 5 * 3 = 21$$

Summation of samples from several trips

Again a simple summation will do. The estimated number caught by length group in all $n(t, f, h)$ sampled trips is:

$$\sum_{j=1}^{n(t, f, h)} C_m(s, t, f, h, j, i) = C_m(s, t, f, h, *, i)$$

In Fig. 7.6.4, for example, the estimated total number of all four sampled trips for length group 8-9 cm becomes:

$$21+22+20+26 = 89$$

Raising the sampled trips to the total catch of the fleet in one landing place

The total length distribution of the sampled trips $C_m(s,t,f,h,*,i)$ can be raised to account for the entire catch in period t , by fleet f in landing place h , by using a raising factor based on the number of trips:

$$\frac{\text{total number of trips}}{\text{number of trips sampled}} * C_m(s,t,f,h,*,i) = CR(s,t,f,h,*,i)$$

where the suffix "R" stands for "raised". R has only been used here to indicate raising procedures that include quantities that have not been sampled (in this case 11 out of the 15 trips). In the example (Fig. 7.6.1) the raising factor is $15/4 = 3.75$ and the result is shown in Table 7.6.1. The estimated total number of specimens in length group 8-9 cm for all trips by fleet f in landing place h becomes:

$$89 * 3.75 = 333.75$$

This raising procedure is reasonable only if a "trip" is a well-defined unit of effort. If some trips are of a duration of, say, one fishing day and others are of a duration of five fishing days, it is better to use a "fishing day" as unit of effort. Either unit, "trip" or "fishing day" makes sense only if the boats of a fleet are fairly similar, in the sense that they have the same "fishing power". Other possible units of effort are "number of man days", "number of trawling hours", "number of gill net sets", etc.

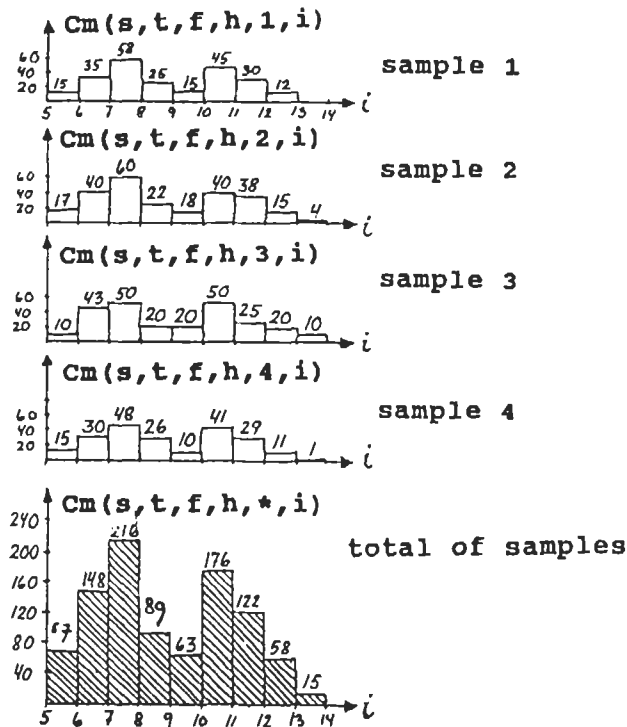


Fig. 7.6.4 Adding up length compositions of species s of sampled trips. The sample used in the example of Fig. 7.6.3 appears as sample no. 1

Table 7.6.1 Raising the sampled trips of fleet *f*, in landing place *h* to the total catch of all trips that fleet (see Figs. 7.6.1 and 7.6.4)

length group (<i>i</i>)	total sampled trips (from Fig. 7.6.4) $C_m(s, t, f, h, *, i)$	raised to total number of trips trips in the sampled landing place $CR(s, t, f, h, *, i)$
5-6	57	213.75
6-7	148	555.00
7-8	216	810.00
8-9	89	333.75
9-10	63	236.25
10-11	176	660.00
11-12	122	457.50
12-13	58	217.50
13-14	15	56.25
total	944 $= C_m(s, t, f, h, *, *)$	3540.00 $= CR(s, t, f, h, *, *)$

Summation of sampled landing places for one fleet and raising to all landing places

The total length distribution of all sampled landing places (see Fig. 7.4.1) is obtained by simple addition:

$$\sum_h CR(s, t, f, h, *, i) = CR(s, t, f, *, *, i)$$

This figure can be raised to the total for all landing places by applying a raising factor based on the effort expended in all the landing places:

$$\frac{\text{total effort of all landing places}}{\text{effort of sampled landing places}} * CR(s, t, f, *, *, i) = CRR(s, t, f, *, *, i)$$

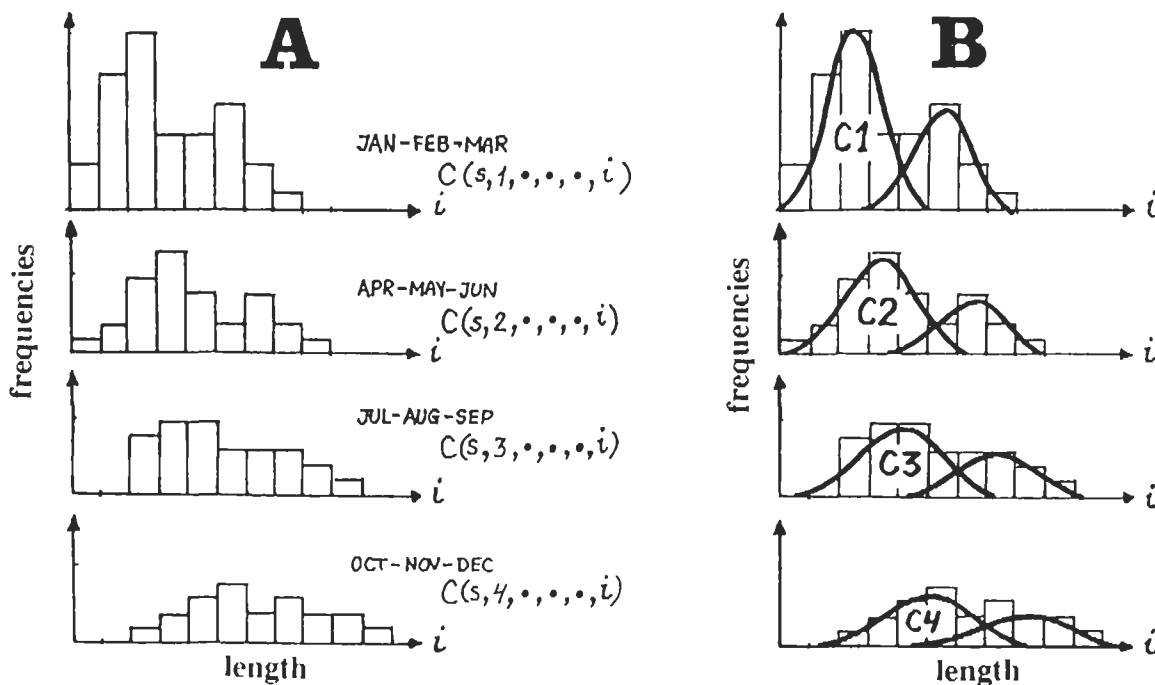


Fig. 7.6.5 A: Total length-frequencies by quarter year.
B: Length-frequencies resolved into normally distributed cohort components (input to Pope's cohort analysis)

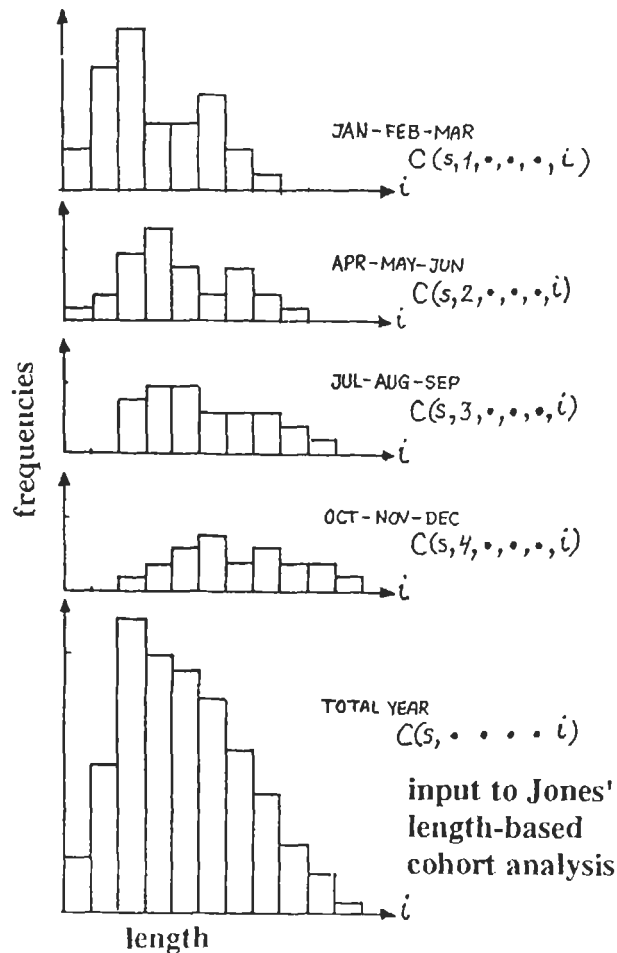


Fig. 7.6.6 Summation of total length compositions of all time periods (total length composition as input for Jones' length-based cohort analysis)

Summation of fleets

This too is a simple summation. The total length distribution of species s caught during time period t is:

$$\sum_f \text{CRR}(s, t, f, *, *, i) = \text{CRR}(s, t, *, *, *, i)$$

We have now obtained a complete picture of the length-frequency distribution of all the landings of species s in one quarter of the year, t . Depending on the type of analysis required we may stop the processing of catch data at this level. In that case the results after one year are the four quarterly length-frequencies as shown in Fig. 7.6.5A. These may be resolved into cohort components by, for example, the Bhattacharya method (Section 3.4) as shown in Fig. 7.6.5B. The numbers caught from the same cohort in different quarters of the year (e.g. the numbers C_1, C_2, C_3 and C_4 in Fig. 7.6.5B) form the inputs to Pope's cohort analysis (compare Section 5.2).

Alternatively, we may choose to apply Jones' length-based cohort analysis (compare Section 5.3). In this case we do not separate the cohorts, but proceed as follows.

Summation of time periods

This is the final step, which gives the length-frequencies of species s for the whole year. The summation is simple:

$$\sum_t CRR(s, t, *, *, *, i) = CRR(s, *, *, *, *, i)$$

Fig. 7.6.6 shows an example in which quarterly length compositions are summed to an annual length composition. These final C-values can be used as inputs to Jones' length-based cohort analysis. We may also use the average values for a range of years (see Section 5.3).

Data analysis

The resolution of length-frequency samples into normally distributed components, as shown in Fig. 7.6.5, becomes more problematic the longer the sampling period is. The quarterly samples show a certain cohort structure, whereas in the length-frequency distribution for the whole year the cohorts cannot be distinguished (see Fig. 7.6.6). This example illustrates that for age-based cohort analyses we must work with relatively short time periods, otherwise we are unable to identify the cohorts. Pope's age-based cohort analysis deals with the numbers caught per cohort.

On the other hand, for a Jones' length-based cohort analysis we are interested in the right hand slope of the combined length-frequency distribution, because this slope is a reflection of total mortality. Therefore the combined length-frequencies should represent a long period, so that the individual slopes of length-frequencies of single cohorts are levelled out.

Finally, it is emphasized that the procedure explained above is an example which may not fit to all fisheries. Especially the definition of the effort unit ("trip") may be inappropriate for many fisheries. Also the assumption that the samples are taken from the boats at the time the catch is unloaded may not fit to all cases. In the example it has been assumed that the total catch has been landed, in the form of consumption fish and unsorted by-catch. In this case there were no "discards".

Discards are fish caught but not landed, that is, thrown back into the sea. Discards are believed not to survive the encounter with the fishing gear. From a biological point of view, discards are as important as landings, as the important biological point is that the fish were killed by fishery. Some fisheries, notably shrimp trawling, discard up to 90% and sometimes even more of the weight caught. The discards may well contain good fin fish for human consumption, but which compared to shrimps are of relatively low value. Actually, one should carefully distinguish between "landings" and "catches", the latter including both landings and discards. Discards are difficult or expensive to sample, as reliable estimates would require observers to be placed on board of the commercial vessels. However, if discards are important quantities, attempts to sample them should be made.

Computer programs

The FiSAT and LFSA packages of microcomputer programs contain programs for data manipulations as described in this chapter, i.e. various types of summations and raising procedures for length-frequency samples.

