

## 4.1 Introduction and statistical basics

This chapter is designed to provide an overview of the data requirements for assessment and research within ICCAT.

To this end, it provides the reader with a basic understanding of how to design data collection programmes, through sampling of vessel catches, and how to ensure these are representative of the entire population (section 4.2). This theme is repeated throughout this chapter, but readers should refer to this section initially. One of the main forms of data collected routinely is length data. The collection of this data, and its use to estimate the age structure of catches, is detailed in section 4.3. Another source of stock status information used in assessments is catch per unit effort (CPUE), commonly derived from vessel logbooks. These data need standardisation over time between areas as well as between the different vessel categories or fishing gears, to ensure signals in the data are consistent. The issues involved in the use of CPUE data are detailed in section 4.4.

When managing stocks, a number of biological attributes are highly important. These include the geographic range and boundaries of a stock, its interaction with sub-stocks, and patterns of migration. A number of techniques are available to examine these factors, including genetic methods (section 4.5), and tagging (sections 4.6 and 4.7). Within defined stocks, knowledge on the reproductive patterns of large pelagics, as well as the characteristics of growth and mortality, will largely define the regenerative capacity of a population. Hence they are extremely important for management and conservation, and the construction of reliable models for effective stock assessment. Methods to investigate these biological attributes are detailed in sections 4.8 and 4.9. It should be noted that most of these approaches require invasive examination of the fish. As a result of the high value of most tuna species, length information is commonly the only data that can be collected without buying individuals or starting a fishery-independent research programme.

A key approach to collecting information on a wide range of fishery-related characteristics, including searching patterns, characterisation of fishing effort, bycatch and discard mortality, plus the collection of accurate biological information, is through scientific observer programmes. Section 4.10 discusses basic approaches to optimise observer coverage, the forms of information that can be collected, and the important issue of estimating bycatch from observer data.

### 4.1.1 Biostatistics

Much of this manual contains details of the statistical methods used within the areas of interest. To assist understanding, this section provides a simple and brief reminder of the basic statistical concepts. For further information, readers are urged to refer to other texts on biostatistics, including “Biometry” by Sokal and Rohlf (1995) that contains both the theory behind approaches and useful examples of their use on data, “Sampling techniques” by Cochran (1977) and “Sampling” by Thompson (1992). Sparre and Venema (1998) also provide an excellent manual for tropical fish stock assessment, from which much of this section is sourced.

#### *Mean value and variance*

Consider a sample of  $n$  fish of a single species, all caught in a single set, and let  $x(i)$  be the length of fish no.  $i$ ,  $i=1, 2, \dots, n$ . The mean length of the sample is defined as:

$$\bar{x} = \frac{[x_{(1)} + x_{(2)} + \dots + x_{(n)}]}{n} = \frac{1}{n} * \sum_{i=1}^n x_{(i)}$$

As an example, if 12 fish were sampled of lengths (cm) of 176, 175, 162, 174, 161, 156, 178, 158, 195, 171, 177 and 154, the mean length of this sample would be:

$$\bar{x} = \frac{[176 + 175 + \dots + 154]}{12} = \frac{1}{12} * 2037 = 169.75$$

The sample variance, a measure of the variability around the mean value, is defined as:

$$s^2 = \frac{1}{n-1} * \left[ (x_{(1)} - \bar{x})^2 + (x_{(2)} - \bar{x})^2 + \dots + (x_{(n)} - \bar{x})^2 \right] = \frac{1}{n-1} * \sum_{i=1}^n [x_{(i)} - \bar{x}]^2$$

Usually this is calculated as  $\sum x^2 - (\sum x)^2 / n / (n-1)$  to avoid rounding errors.

The variance is therefore the sum of the squares of the deviations from the mean divided by the number (n) minus one. If all fish in the same had the same length, therefore, the variance would be zero. For the sample of lengths detailed above, the variance would therefore be:

$$s^2 = \frac{1}{12-1} * \left[ (176 - 169.75)^2 + (175 - 169.75)^2 + \dots + (154 - 169.75)^2 \right] = \frac{1}{12-1} * 1556.25 = 141.48$$

The square root of the variance, s, is the standard deviation. In the example, s=11.89. Variance can also be expressed relative to the size of the mean, as the coefficient of variation. For this, the standard deviation is relevant since it has the same unit as the mean. The coefficient of variation is:

$$\frac{s}{\bar{x}}$$

From the example, the coefficient of variation (CV) is:

$$\frac{11.89}{169.75} = 0.07$$

Much of statistics relies on the 'normality' of data. This essentially means that the data (and the population from which they are taken) conform to a normal distribution:

$$Fc(x) = \frac{n * dL}{s * \sqrt{2\pi}} * \exp \left[ - \frac{(x - \bar{x})^2}{2s^2} \right]$$

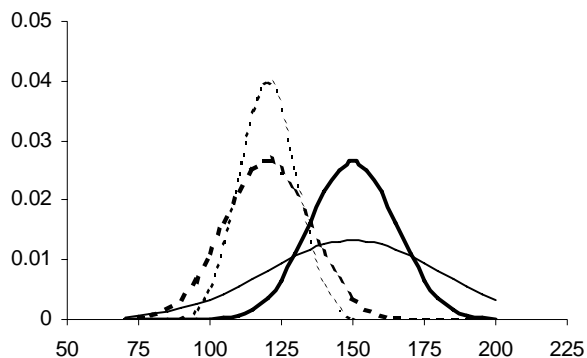
where Fc(x) is the 'calculated frequency', n is the number of observations, dL is the interval size (of the measurement in question), s=standard deviation,  $\bar{x}$  the mean length and  $\pi=3.14159\dots$

A normal distribution is often observed for older, larger fish (young, small ones would need some negative values to be normally distributed) when recording length frequencies of fish from a single cohort (i.e. fish of the same age), and the probability of a fish more or less than a given size in the sample can therefore be estimated. Other probability distributions exist (e.g. lognormal), where the distribution of measurements is skewed, rather than being centred around the mean as in the normal distribution.

The concepts of bias and precision arise from the consideration of means and distributions (**Figure 4.1.1**). An estimate from a sample is said to be unbiased if the average of many replicate estimates is the same as the true value (that would be achieved if all specimens in the total population were sampled). An estimate is biased if it deviates from the true value in a systematic manner. For example, if estimates of mean length from samples were always greater than the true length in the population. This could occur due to the selectivity of the gear. With an unbiased sample, the true value can be more closely approached by increasing the sample size. This is 'consistency'. With a biased estimate, there will always be a difference between true and estimated values.

To obtain an unbiased estimate, a random sample should be taken. In this case, any fish sampled from the stock (as an example) should have exactly the same probability of being sampled. True random samples are often difficult to achieve in practice, however.

Precision is a measure of whether samples or estimates are 'precise'. In this case, the variance around the mean value of the sample or estimate is low (**Figure 4.1.1**). This does not necessarily mean the sample or estimate is unbiased – they can be precise (clustered closely around a given value), but biased (that given value is not equal to the true mean, for example).



**Figure 4.1.1** Demonstration of bias and precision. The normal distribution indicated by the thick black solid line represents the population distribution. The thin solid line represents an unbiased but less precise distribution (the mean is the same as the population, but the spread is wider). The distribution represented by the thick broken line represents a biased sample - the distribution has identical variance to the true population but the mean value is lower than the true value. The distribution represented by the thin broken line is biased, but more precise.

#### 4.1.2 Further reading

COCHRAN, W.G. (1977). Sampling techniques. New York, J. Wiley & Sons, Inc.

SOKAL, R.R. and F.J. Rohlf (1995). Biometry: the principles and practice of statistics in biological research. W.H. Freeman and Company, New York.

SPARRE, P. and S.C. Venema (1998). Introduction to fish stock assessment. Part 1. Manual. FAO Fish. Tech. Pap. 306(1), FAO, Rome.

THOMPSON, S. K. (1992). Sampling. John Wiley & Sons, Inc. 343p.