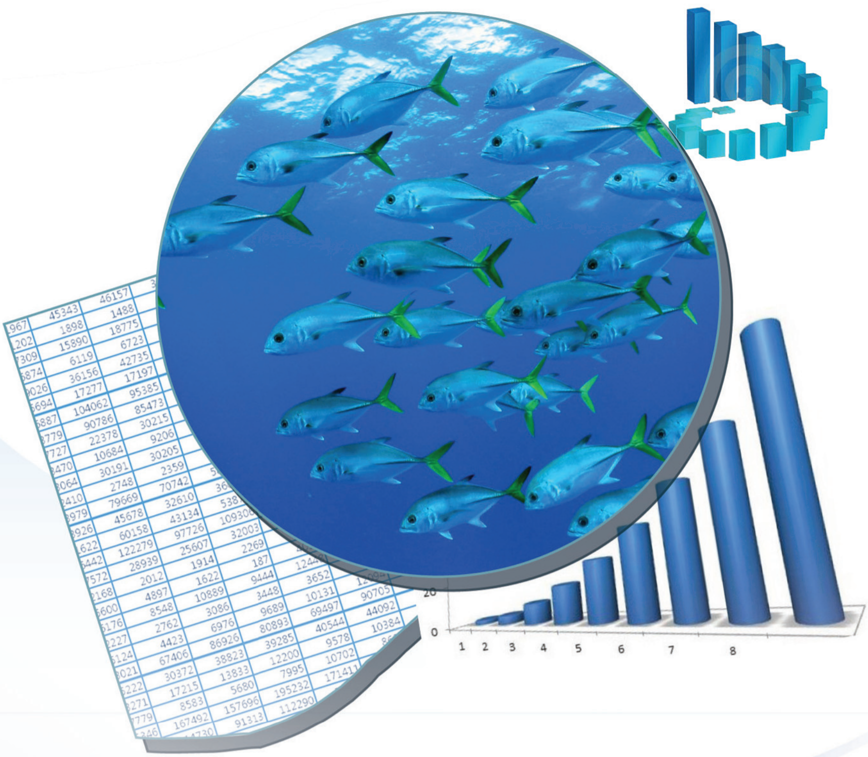
Training Manual

Fish Stock Assessment and Management



Fishery Resources Assessment Division

ICAR - Central Marine Fisheries Research Institute

kochi | Kerala | India

Training Manual

Fish Stock Assessment and Management

Editors
T.V.Sathianandan
Mini K.G
Somy Kuriakose

Technical Assistance
Manjeesh R

Published By
Dr.A.Gopalakrishnan
Director
Central Marine Fisheries Research Institute

Fishery Resources Assessment Division
ICAR - Central Marine Fisheries Research Institute
kochi I Kerala I India

Training Manual

Fish Stock Assessment and Management

Fishery Resources Assessment Division
ICAR - Central Marine Fisheries Research Institute
(Indian Council of Agricultural Research)
Post Box No. 1603, Ernakulam North P.O., Kochi-18
www.cmfri.org.in

Contents

Contents

01

Importance of Data Collection

J. Jayasankar

Introduction

Information is both the input as well as output of any scientific endeavour. Research investigations get initiated with requirements of certain types viz. estimating the unknown of confirming certain postulates etc. and these are borne out of information and to accomplish these requirements experiments have to be laid out which in turn again would lead to information. Hence it can be easily pronounced that research investigations are basically stemmed by information generation and processing. The tangible pieces of information are usually referred to as data. The details like the morphology of an animal, rate of growth, death rate and other facts collated during the course of the trailing a resource constitute the data pertaining to the resource. The data collected on a purpose are popularly termed as experimental data. For example if an investigator wants to estimate the rate of increase in size or weight of an animal over a period of time, he has to repeatedly take the measurement of interest at various pre-determined time periods. Each such attempt results in data and an objective analysis of thus collated data forms the crux of the investigation ultimately leading to better information on the aspect of interest pertaining to the animal under focus.

Types of data

The data thus collected can be of many a type. The nature of information can lead to one classification viz., qualitative and quantitative. The qualitative data are those which are statements of existence of the subjects under focus. For example the colour, state, presence or absence etc. of an item under focus can be treated as qualitative data. On the contrary numeric measurements pertaining to the state of the object constitute the quantitative data. For example length of fish, its weight, truss measurement, surface temperature etc. fall into this category. Another way of classifying data is based on the source of the data. If the data is freshly collected by the investigator for the experiment, then the data is termed as primary data. If for some investigative purposes, the historic data or data published in reports and journals are used, the information thus utilised is termed as secondary data. Yet another type of classification is based on the tagging of the data. If the data collated is based on the information obtained on a particular time period identified by day, week, month or year, such data are termed as time series data. For example, the annual landing figures of India from the year 1990 to 2010 can be treated as time series of landings. Another type of tagging pertains to the geographic location of the data. For example the sea surface temperature off a place in west coast of India which could be uniquely identified by its latitude and longitude can be termed s spatial data or geo-referenced data. Among the quantitative data further classification could be interval data, ordinal data and nominal data. Interval data is one which indicates an absolute measure of a trait of an individual, for example weight of fish whereas nominal data indicates numbers given for various categories of states of the individual (for example 1 for the trait present and 0 when not) and the ordinal data pertains to numbering given to states of objects with relational increase or decrease in intensity (for example 1 for less frequent, 2 for moderately frequent and 3 for highly frequent occurrences). From the sheer computational point of view data can be categorised into text or string, numeric, Boolean and date. The string data pertains to names of species, locations etc. and the numeric data could be any quantitative measure of the animal whereas Boolean is the recording of yes or no trait and date format indicated the date and time of recording.

Role of data collection in analytical research

As mentioned earlier data collected on purpose should adhere to basic premises of the underlying investigation. Be it a complicated research endeavour or a small survey, the data collected for a particular purpose should be perfectly suited to various analytical methodologies and tools existing for the fulfilment of the objectives of the experimentation. A research plan is to be adopted during the course of scientific investigations and the type of data to be collected and the checklist to be prepared before venturing into data collection.

A concise summary of some of the key points we have tried to make in this document about research and research design. It is given in the table focuses on the relationship between research problems, research questions, and research aims and objectives. These are not self-evident matters. In order to clarify what our research is really going to be about we need to do a lot of analytic work in which we analyze our research problem to discover what components or dimensions it might contain. We need to analyze these components or dimensions in terms of the kinds of questions we can ask that will allow us to address them in ways we (and others) find satisfying. In addition, we need to know how to analyze both our problem and our questions in ways that allow us to identify what our research aims and objectives are. Without aims and objectives we will not be able to know when our tasks have been met or how well we have addressed these tasks.

Research Design Table as Questions to be addressed when Developing our Design

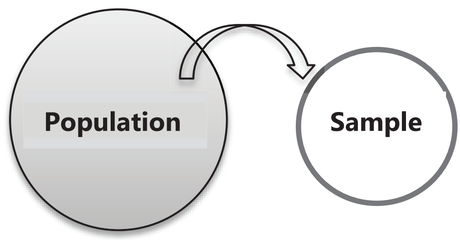
Research Purposes	Data to Collect	Data Analysis Approach	Informants
• What is my research problem? • What are its key dimensions? • What specific questions does it generate? • Which of these questions can I hope to address well? • How can I analyse my problem and questions to specify a succinct set of research aims and objectives, and a finite set of research questions?	• What kinds of research data must I collect? • How will this data help me address my purposes? • How will I collect this data? • How much data will I collect? • How will I validate this data or establish that it is good quality data? • How will I organize this data so it will be in good shape for being analysed?	• What forms of data analysis will I use? • What justifies these forms of analysis? • How will they help me achieve my research purposes? • How are these forms of analysis conducted? • What do I need to know, be able to do and have access to do in order to use these forms of analysis in an expert way? • What will these forms of analysis let me 'say' about the data (and not say)?	• Who helps us understand how to frame and refine research purposes? • What are some relevant books, articles, and chapters here? • Who provides theories and concepts relevant to our research problem and questions? • How do these guide us in deciding what kinds of data to collect and how to analyse it? • Who provides good advice on how to collect, validate and organize data? • Who provides good information about how to analyse data?

Of course, unless we are clear about our purposes and know what kinds of questions are involved we will lack a strong and clear base for deciding what kind of data to collect, how much of it to collect, and so on. Therefore, we need to be able to analyze and specify our questions and purposes in ways that guide our data collection. While this may seem obvious, it is surprising how often people fail to recognize the importance of the relationship between our purposes and decisions about data. This occurs, for example, every time someone collects their data before they have a clear research question. In terms of a logical research process it is simply absurd to collect data before one has a question. This is like buying ingredients for cooking before knowing what it is that one wants to cook. In addition, the data aspect draws attention to the fact that it is not rational to invest time in analyzing data unless we are confident that the data is trustworthy and of good quality. It doesn't matter how competently we analyze our data if the data itself is of poor quality. If the data is poor even excellent analysis won't be able to give us research outcomes we can have faith in. Once again, although this seems an obvious point it is surprising how often data is analyzed without researchers having checked to see if it really does pertain to their purposes and whether they are justified in believing the data is of good quality. In addition, it is important to recognize that data accumulates very quickly, and that analysis can become a very difficult task—if not impossible—unless we start organizing our data in appropriate ways from an early stage in the research process. Often the kinds of analysis we intend to conduct require us to arrange our data in one kind of format rather than another.

When we are thinking about collecting our data we need also to be thinking about how it will be analyzed. Some kinds of analysis require much more data than other kinds. Some require a lot more time than others. If we collect more data than we have the necessary time and other resources for analyzing, we are wasting time and effort—our own and other people's. Alternatively, if we find at the point of analysis that we have not collected enough data or enough good quality data it may be too late to go back and supplement the data we have. This could undermine the entire research activity.

The Fisheries Data Paradigm

In the field of natural resource investigations like fisheries studies, the generic paradigm presumed is the population- sample setup. The inferences about a population being studied, is drawn from the information captured by a sample taken from the population. The sites selected for sampling are a representation of the population and the mean abundance observed at these sites are used to estimate the mean abundance of the entire population by applying some raising factor. This paradigm can be denoted as in the diagram. A raising factor is a ratio between the total number of units and the sample number of units. Thus it is imperative

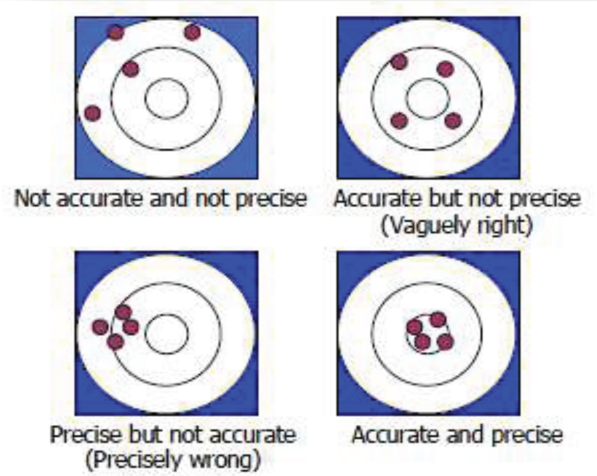


to ensure that the sample is reliable and truly representative of the population being studied before any data analysis is carried out. That is the sampling sites should be selected in an unbiased manner and this can be achieved through proper sampling and survey design. Bias and precision are two keys to take into consideration when designing a survey.

A sample is characterised by:

Variability (spread of observations) = Precision

Uncertainty (bias, deviation of observed means from true mean) = accuracy, which again in denoted in the diagram.



Surveys Vs Commercial Catch Statistics

Independent surveys of fisheries resources often appear to provide a more accurate prognosis of the status of a fish stock than commercial catch sampling. These surveys are normally carried out by fishery independent research vessels. Surveys were actively promoted after the conceptualisation of Exclusive Economic Zone (EEZ) by United Nations Convention on Law of the Sea (UNCLOS). Since then standardised scientific abundance survey data has become an important tool to estimate “indices of abundance” in the current situation and for long-term monitoring of most of the commercially important stocks. The advantage that survey based methods have over those based on commercial catch statistics is that the uncertainties associated with survey estimates can be studied and quantified and based on such research, survey methods and ultimately stock assessments can be improved. But one overwhelmingly looming deterrent to this methodology is the exorbitant cost involved which most of the developing nations can ill afford.

Basics of data collection

Preferably data should be collected at the highest resolution possible. Some of the basic data needed for fisheries resource assessments include:

(i) Catch or landed catch (numbers or weight of fish) recorded by sampling area

(ii) Location of catch; latitude, longitude, provincial boundaries, statistical rectangles

(iii) Species of fish (catch and stock composition); especially tropical coral reef systems are multi species and the landed catch comprises significant number of species

(iv) Effort which is a measure of fishing activity such as number of boats, number of days fished, number of hours fished

(v) Measurement of individual fish in sample

a. Length frequency measurements: fish are normally measured on a measuring board and counts are tallied. E.g.

cm	
10	
11	I
12	II
13	I
14	III
15	II
16	
17	III
18	II

a. Individual measurements can be recorded and tabulated in which case together with length, weight, sex, maturity stage etc. is also recorded for each fish (Example given in table). If the fish can be aged through readings of hard structures such as otoliths/ scales then this information can also be added.

No.	Length	Weight	Sex	Mat	age
1	14.5	12.5	0	4	2
2	15.0	13.0	0	5	3
3	14.5	12.0	1	4	3

b. Other detailed biological measurements include, liver weight, fat content, RNA / DNA ratios, size and number of eggs.

(vi) A measure of discarded catch

(vii) For economic analysis: cost of fishing, number of jobs

Data collection can be a costly affair. Therefore it becomes necessary to prioritise and compromise under funding constraints. More so it is necessary to optimise through information sharing. Such as government agencies, NGOs, private sectors and fishermen can team up in information gathering.

Conclusion

Any investigation into a vibrant system like fisheries resources which are always in a state of flux timely data collection with due objectivity becomes mandatory. With a clear understanding of the mandate and the objectives and a comprehensive picture of resources at disposal one has to do the balancing between cost effectiveness and precision. Often one could lead to trade off with the other and hence a clear delineation of priorities is a must before even planning an experiment. Data hence collected to a plan under watchful supervision is bound to reflect the true story of the animals eternally covered under a sheet of water. So precise data collection, at a prompt occasion would certainly propel the investigator to a peerlessly precise assessment of resources under focus.

References

Colin Lankshear, Some Notes on the Nature and Importance of Research Design within Educational Research,Faculty of Education, Monash University.

02 Statistical Tools

T.V. Sathianandan

Statistics is a branch of mathematics that provides us with the techniques that enable us to classify, summarize and draw valid inferences from data according to basic rules of evidence.

Branches of statistics

Descriptive	Inferential
• the goal of descriptive statistics is to organize and to summarize data so that they are more readily understood.	• the goal of inferential statistics is to take into account the chance factors associated with sampling and subsequently drawing valid conclusions.

Terminology

Population

The total group of observations about which one wants to draw a conclusion.

Parameter

A measurable characteristic of a population.

Example - mean, standard deviation, correlation, proportion

Sample

Any subset or subgroup of a population . Random samples occur when every element of the population has an equal chance of being selected or drawn

Statistic

A measurable characteristic of a sample.

Example - mean, standard error, correlation, proportion

Variables are characteristics we wish to observe; properties whereby members of a group differ from one another

There are different ways to categorize variables

1. Quantitative vs. Qualitative
2. Discrete vs. Continuous
3. Dependent vs. independent

Qualitative or nominal variables differ in kind rather than amount (e.g., ethnicity, eye color).

Ungrouped data – a table in which scores are listed from smallest to largest and the frequency of occurrence is noted for each possible score.

Grouped data – scores are placed into class intervals and the frequency of scores for each class interval is noted. The lowest and highest possible scores for each interval are known as the score limits. Guidelines for creating grouped data include:

1. Intervals cannot overlap
2. All intervals have the same width
3. Try to have between 10 and 20 intervals

f	= number of scores in interval
Cum f	= number of scores accumulated at the end of the interval
%f	= (f / N) x 100
Cum % f	= (Cum f / N) x 100

Frequency Distribution – a table which shows scores and their frequency of occurrence Questions one can answer with frequency distributions:

1. What is the most frequently occurring score?

2. How spread out or concentrated are the scores?

3. Is there anything unusual about the pattern (or shape) of the score distribution?

4. What is the middle or average score?

5. Where is the highest concentration of scores?

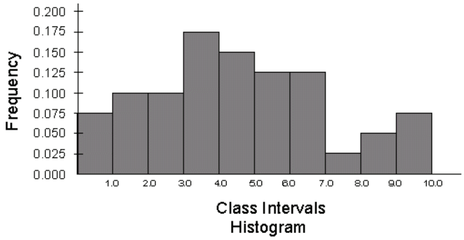
4. Intervals must be continuous through the distribution (there should be an interval for each observable score)

Note: when data are grouped you lose information.

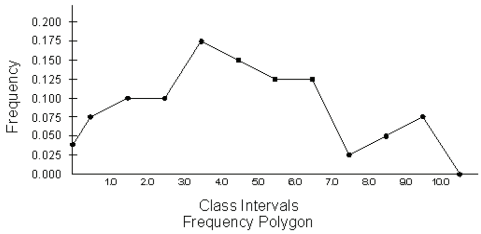
A set of data does not result in a unique set of grouped intervals. Two different sets of raw scores could yield the same grouped data frequency distribution.

Graphically displaying data

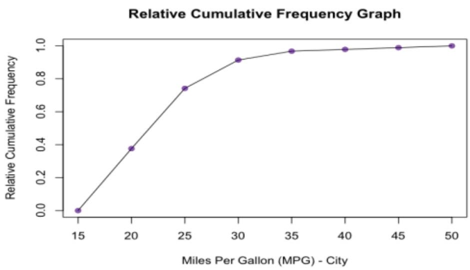
Histogram – a graphical way to display frequencies using a bar graph. The height of each bar represents the frequency of scores for the interval.



Frequency polygon – a line graph in which the frequency for each interval is denoted by a single point at the midpoint of the interval. The points are then connected forming a “polygon” with the x-axis. Points are plotted at the midpoint of the interval. The graph starts and ends at the X-axis.



Ogive Curve – a line graph displaying cumulative frequencies. The y-axis represents cumulative frequencies. Values are plotted at the upper limit of each interval.



Vocabulary associated with the shapes of distributions

- **Symmetric** – a distribution which has the same shape on both sides of a center (e.g., a normal distribution)
- **Modality** – the number of “peaks” or modes in the distribution (e.g., unimodal, bimodal)
- **Normal distribution** – a “bell-shaped” distribution that is symmetric and unimodal. Note that there is a specific formula that describes this unique shape. Technically, no observed distribution is exactly normal.
- **Skewness** – the degree of asymmetry of a distribution. If the distribution tapers off to the left it is said to be negatively skewed. If it tapers off to the right it is positively skewed. The skewness value for a symmetric distribution is zero.

$$\text{Skewness} = \frac{\sum z^3}{N} \text{ where } z = \frac{X - \overline{X}}{S}$$

- **Kurtosis** – is a measure of the "peakedness" of a distribution. Severely peaked distributions are called "leptokurtic". Moderately peaked distributions such as a normal curve are called "mesokurtic". Flat distributions are called "platykurtic". The kurtosis value for a normal distribution is 3.

$$\text{Kurtosis} = \frac{\sum z^4}{N} \text{ where } z = \frac{X - \overline{X}}{S}$$

Measures of central tendency

Mean	Most common measure ; is computed by summing all the scores and dividing it by the number of scores . It does not have to be an observed value. It tends to be the most stable estimate of the population mean from sample to sample.
Mode	the value which is most frequent. It represents the most common response. It is used most often with nominal data.
Median	the value below which lies 50% of the values in a distribution. The median does not have to be an observable value. It is the preferred measure of central tendency with "opened data" or with data that are severely skewed.

Measures of variability

In general we are often interested in how the values of a variable are dispersed. This dispersion is usually in reference to the mean. The range is the easiest measure of dispersion to compute and is simply the difference between the largest value and the smallest value. The range does not really indicate how the scores are concentrated along the distribution.

The interquartile range represents the difference between the 25% value and the 75% value. This range indicates the spread of the middle 50% of scores. It is used most often in skewed distributions.

The variance is the most commonly accepted measure if dispersion. It represents the average of the squared deviations about the mean. It is in squared units! The formula for variance depends on whether you are working with a population or sample:

$$\text{Population: } \sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

$$\text{Sample: } S^2 = \frac{\sum (X - \overline{X})^2}{n - 1}$$

To return to the original metric we take the square root of the variance. This is called the standard deviation. Thus,

$$\text{Population: } \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(X-\mu)^2}{N}}$$

$$\text{Sample: } S = \sqrt{S^2} = \sqrt{\frac{\sum(X-\overline{X})^2}{n-1}}$$

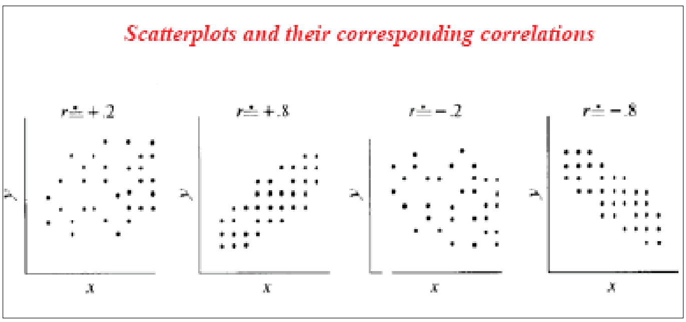
One measure of dispersion that is helpful in comparing standard deviations is the coefficient of variation. Because the size of the standard deviation could be effected by the magnitude of the mean, we divide the standard deviation by the mean to form the coefficient of variation (CV). $CV = \frac{S}{\overline{X}}$

Estimators – In some cases because we do not have access to the entire population we have to estimate the population parameter based upon a sample. The mean of a sample is an estimate of the population mean.
If the mean (also called the expected value) of the sampling distribution of a given statistic is equal to the population value then we say that statistic is unbiased. This is why we compute variance by dividing the sum of the deviations squared by n-1, it makes the variance unbiased.

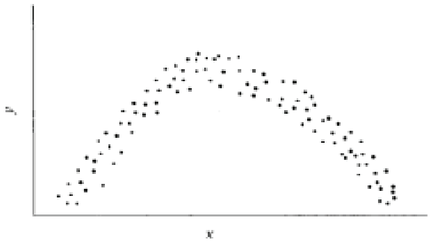
Correlation – In some cases we want to compare performance on two interval variables (e.g., two sets of scores, say EOG math and EOG reading). Visually we do this by creating a scatter plot in which each subjects' pair of scores is represented by a single point. The more linear the distribution of points is in this two-dimensional plot, the greater the value of the correlation.

Creating a scatterplot

Correlation coefficient of a sample is denoted by the letter r or r_{xy} (the subscript x and y denote the variables that are being correlated). r ranges from -1.0 to +1.0. The sign of the correlation indicates the direction of association not the strength. A positive correlation indicates people who scored high on one test tended to score high on the other test. A negative correlation indicates an inverse relationship: people who scored high on one measure tended to score low on the other.



One evaluates the strength of the correlation by comparing r^2. When computing the correlation one should always create a scatterplot to check for anomalies. If there are identifiable subgroups, they should be plotted in different scatterplots. If the range of one or both of the variables is restricted the correlation value will be lowered. Always plot subgroups separately to make sure the correlation is the same for each subgroup! Also check for outliers!



If your scatterplot is not linear, (i.e., it is curvilinear) then Pearson's correlation is not appropriate and will underestimate the relationship between the two variables.

There are several formulas for correlation including:

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\text{Covariance}}{\text{Product of the Standard deviations}}$$

where $$S_{xy} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{n-1}$$

If your scatterplot is not linear, (i.e., it is curvilinear) then Pearson's correlation is not appropriate and will underestimate the relationship between the two variables.

Comparing frequencies of two or more groups

With a qualitative or nominal variable the only type of data that we can collect is usually frequency data. If one wanted to compare the frequencies of subjects in the different categories of one variable or one wanted to compare say the frequencies of males and females across the different categories of another variable one would use a chi-square analysis.

If the same subjects are used and one wants to compare their performance at time one and time two (i.e., pretest, posttest) then one would use a dependent t-test. If data are collected at three or more time points on the same subjects, then one would use a repeated measures ANOVA.

Comparing means of two or more groups

To compare the mean performance on a test between two groups is done using inferential statistics. If there are only two groups, and the groups contain different subjects one would use an independent t-test. Note that the n's need not be the same for each group. If there are more than two groups one would perform an analysis of variance.

03

Correlation

Somy Kuriakose

Introduction

Correlation is a statistical technique used to assess a possible linear association between two continuous variables. Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. If there is any relation between two variables i.e. when one variable changes the other also changes in the same or in the opposite direction, we say that the two variables are correlated. If it is proved that in a large number of instances two variables, tend always to fluctuate in the same or in the opposite direction then it is established that a relationship exists between the variables. This is called a "Correlation." It means the study of existence, magnitude and direction of the relation between two or more variables.

Types of Correlation

1. Positive and negative correlation

If two variables change in the same direction (i.e. if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation. If two variables change in the opposite direction (i.e. if one increases, the other decreases and vice versa), then the correlation is called a negative correlation.

2. Linear and non-linear correlation

If the ratio of change between two variables is uniform, then the correlation is said to be linear. If the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable, then the correlation is said to be non-linear or curvilinear. The nature of the graph gives us the idea of the linear type of correlation between two variables. If the graph is in a straight line, the correlation is called a "linear correlation" and if the graph is not in a straight line, the correlation is **non-linear** or **curvi-linear.**

Degrees of Correlation

Through the coefficient of correlation, we can measure the degree or extent of the correlation between two variables. On the basis of the coefficient of correlation we can also determine whether the correlation is positive or negative and also its degree or extent.

1. **Perfect correlation:** If two variables changes in the same direction and in the same proportion, the correlation between the two is **perfect positive**. According to Karl Pearson the coefficient of correlation in this case is +1. On the other hand if the variables change in the opposite direction and in the same proportion, the correlation is **perfect negative**. its coefficient of correlation is -1. In practice we rarely come across these types of correlations.

2. **Absence of correlation:** If two series of two variables exhibit no relations between them or change in variable does not lead to a change in the other variable, then we can firmly say that there is **no correlation** or **absurd correlation** between the two variables. In such a case the coefficient of correlation is 0.

3. **Limited degrees of correlation:** If two variables are not perfectly correlated or is there a perfect absence of correlation, then we term the correlation as Limited correlation. It may be positive, negative or zero but lies with the limits • 1.

High degree, moderate degree or low degree are the three categories of this kind of correlation. The following table reveals the effect (or degree) of coefficient or correlation.

Degrees	Positive	Negative
Absence of correlation	Zero	0
Perfect correlation	+ 1	-1
High degree	+ 0.75 to + 1	- 0.75 to -1
Moderate degree	+ 0.25 to + 0.75	- 0.25 to - 0.75
Low degree	0 to 0.25	0 to - 0.25

Methods of Determining Correlation

The following are the most commonly used methods of determining correlation..

(1) Scatter Plot

(2) Kar Pearson's coefficient of correlation

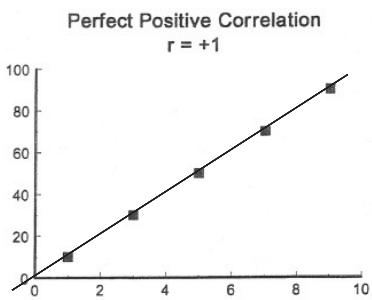
Scatter Plot (Scatter diagram or dot diagram)

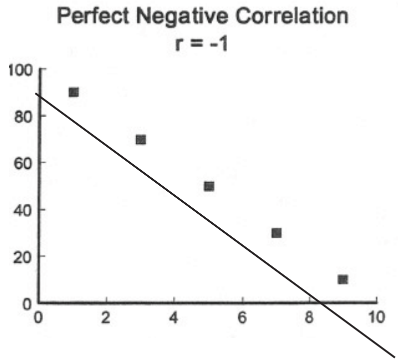
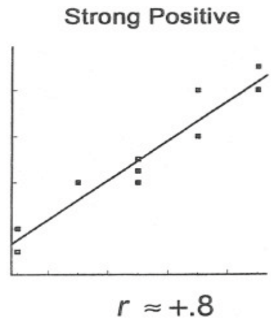
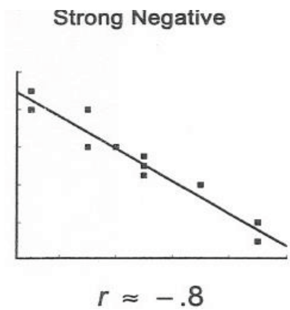
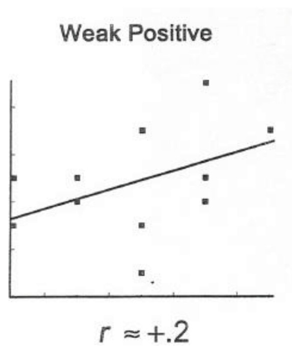
The scatter diagram may be described as the diagram which helps us to visualize the relationship between two phenomena. This is the simplest method for finding out whether there is any relationship present between two variables. In this method the values of the two variables are plotted on a graph paper. One is taken along the x-axis and the other along the y-axis. By plotting the data, we get points on the graph which are generally scattered and hence the name 'Scatter Plot'. The manner in which these points are scattered, suggest the degree and the direction of correlation. The grater the scatter of the points on the chart, the lesser is the relationship between the two variables. The more closely the points come to a straight line, the higher the degree of relationship. The degree of correlation is denoted by ' r ' and its direction is given by the signs positive and negative. Scatter diagrams will generally show one of six possible correlations between the variables:

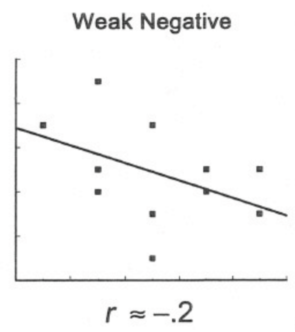
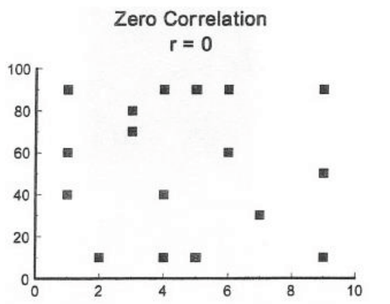
- *Strong Positive Correlation:* The value of Y clearly increases as the value of X increases.
- *Strong Negative Correlation:* The value of Y clearly decreases as the value of X increases.
- *Weak Positive Correlation:* The value of Y increases slightly as the value of X increases.
- *Weak Negative Correlation:* The value of Y decreases slightly as the value of X increases.
- *Complex Correlation:* The value of Y seems to be related to the value of X, but the relationship is not easily determined.
- *No Correlation:* There is no demonstrated connection between the two variables.

Though this method is simple and provide a rough idea about the existence and the degree of correlation, it is not reliable. As it is not a mathematical method, it cannot measure the degree of correlation.

Illustrations

1	If all points lie on a rising straight line the correlation is perfectly positive and r = +1	

2.	If all points lie on a falling straight line the correlation is perfectly negative and r = -1	
3	If the points lie in narrow strip, rising upwards, the correlation is high degree of positive	
4	If the points lie in a narrow strip, falling downwards, the correlation is high degree of negative	
5	If the points are spread widely over a broad strip, rising upwards, the correlation is low degree positive	

6	If the points are spread widely over a broad strip, falling downward, the correlation is low degree negative	
7	If the points are spread (scattered) without any specific pattern, the correlation is absent. i.e. r = 0.	

Karl Pearson's coefficient of correlation

The most widely-used type of correlation coefficient is *Pearson r*, also called *linear* or *product- moment* correlation. It gives the numerical expression for the measure of correlation. The value of ' r ' gives the magnitude of correlation and sign denotes its direction. It is defined as

$$r = \frac{\sum XY}{n\sigma_x\sigma_y}$$

where $X = (X_i - \overline{X}), \quad Y = (Y_i - \overline{Y}),$

$$\sigma_x = s.d.of\ X, \sigma_y = s.d.of\ Y$$

and n is the number of pairs of observations

Properties of Correlation coefficient

1. The value of correlation does not depend on the specific measurement units used; for example, the correlation between height and weight will be identical regardless of whether *inches* and *pounds*, or *centimeters* and *kilograms* are used as measurement units.
2. The value of correlation coefficient lies between -1 and +1, -1 means perfect negative linear correlation and +1 means perfect positive linear correlation.

3. The correlation coefficient r only measures the strength of a linear relationship. There are other kinds of relationships besides linear.
4. If the two variables are independent, then the value of the correlation coefficient is zero. If the value of the correlation coefficient is zero, it does not men that there is no correlation, but there may be non-linear correlation.
5. The value of r does not change if the independent (x) and dependent (y) variables are interchanged.
6. The correlation coefficient r does not change if the scale on either variable is changed. You may multiply, divide, add, or subtract a value to/from all the x-values or y-values without changing the value of r.
7. The correlation coefficient r has a Student's t distribution.

Assumptions to use the Pearson product-moment correlation

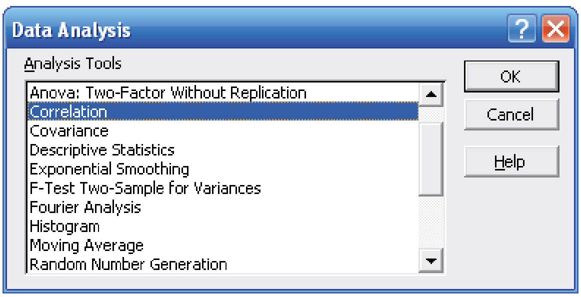
1. The measures are approximately normally distributed
2. The variance of the two measures is similar (homoscedasticity)
3. The relationship is linear
4. The sample represents the population
5. The variables are measured on a interval or ratio scale

Using Excel for Calculation of Correlation

Method I

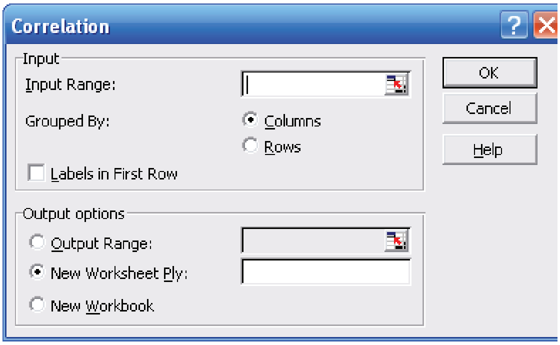
To do correlation in Microsoft Excel, choose *Data- Data Analysis--Correlation.*

In the Correlation popup, you will have to specify:



- the range containing the observed values; if the range includes the cells containing the labels and if you check the box next to Labels in the popup, your results will include those labels.
- whether the grouping is by columns or rows; if, as recommended, you have entered the values of each variable into a column, you need to select the radio button for columns.
- the upper left-hand corner of the range in which you want the output to appear (or you can choose a different sheet in the workbook, or a different workbook).

Excel displays the results in a 2x2 table (for the 2-variable case), showing the correlation of each variable with itself (viz., 1) and with the other variable. Only the cells along the diagonal and in the lower half of the table are filled in, since the table is symmetric with respect to the diagonal. The sample Pearson correlation coefficient thus appears in the cell in the lower left corner.

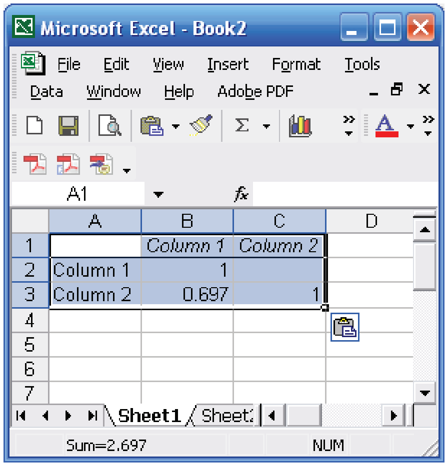


Method II

The function **CORREL** will calculate the correlation between the variables

= CORREL(array *x*, array *y*)

There are several common pitfalls in using correlation. Correlation is symmetrical, not providing evidence of which way causation flows. If other variables also cause the dependent variable, then any covariance they share with the given independent variable in a correlation may be falsely attributed to that independent. Also, to the extent that there is a nonlinear relationship between the two variables being correlated, correlation will understate the relationship. Correlation will also be attenuated to the extent there is measurement error, including use of sub-interval data or artificial truncation of the data. Correlation can also be a misleading average if the relationship varies depending on the value of the independent variable.



04

Regression Analysis

Somy Kuriakose

Introduction

Correlation gives us the idea of the measure of magnitude and direction between correlated variables. Now it is natural to think of a method that helps us in estimating the value of one variable when the other is known. The fact that the variables x and y are correlated does not necessarily mean that x causes y or vice versa. Regression analysis is a statistical tool for the investigation of relationships between variables. It is a powerful technique used for predicting the unknown value of a variable from the known value of another variable. When there is only one independent variable then the relationship is expressed by a straight line. This procedure is called simple linear regression. More precisely, if X and Y are two related variables, then linear regression analysis helps us to predict the value of Y for a given value of X or vice verse. Multiple regression is an extension of bivariate regression in which several independent variables are combined to predict the dependent variable. In multilple regression analysis,the value of Y is predicted for given values of X_1, X_2, ..., X_k.

Dependent and Independent Variables

By simple linear regression, we mean models with just one independent and one dependent variable. The variable whose value is to be predicted is known as the dependent variable and the one whose known value is used for prediction is known as the independent variable. Similarly for Multiple Regression the variable whose value is to be predicted is known as the dependent variable and the ones whose known values are used for prediction are known independent (exploratory) variables.

The Regression Model

The line of regression of Y on X is given by Y = a + bX where a and b are unknown constants known as intercept and slope of the equation. This is used to predict the unknown value of variable Y when value of variable X is known. The Simple Regression model is

$$Y = a + bX$$

The coefficient of X in the line of regression of Y on X is called the regression coefficient of Y on X. It represents change in the value of dependent variable (Y) corresponding to unit change in the value of independent variable (X).

In general, the multiple regression equation of Y on X1, X2, ..., Xk is given by:

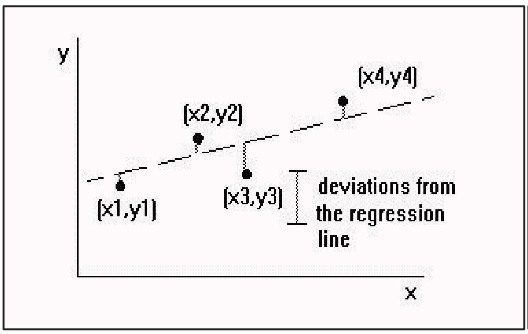
$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots\dots\dots\dots\dots\dots + b_k X_k$$

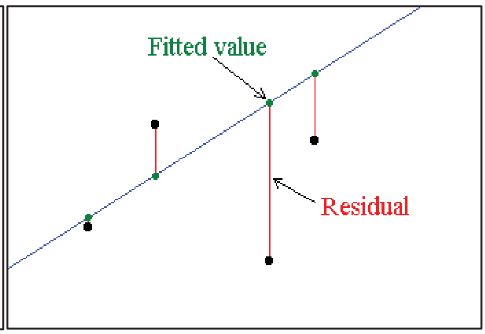
Here b_0 is the intercept and b_1, b_2, b_3, ..., b_k are analogous to the slope in linear regression equation and are also called regression coefficients. They can be interpreted as the change in the value of dependent variable (Y) corresponding to unit change in the value of independent variable X_i.

Fitting of regression line

In scatter plot, we have seen that if the variables are highly correlated then the points (dots) lie in a narrow strip. If the strip is nearly straight, we can draw a straight line, such that all points are close to it from both sides. Such a line can be taken as an ideal representation of variation. This line is called the line of best fit if it minimizes the distances of all data points from it and also called as the line of regression. Now prediction is easy because all we need to do is to extend the line and read the value. Thus to obtain a line of regression, we need to have a line of best fit.

The problem of choosing the best straight line then comes down to finding the best values of a and b. By 'best' we mean the values of a and b that produce a line closest to all n observations. This means that we find the line that minimizes the distances of each observation to the line. Choose the a and b values that give the line such that the sum of squared deviations from the line is minimized. This method of estimation of parameters is called least square method. The best line is called the regression line, and the equation describing it is called the regression equation. The deviations from the line are also called residuals.



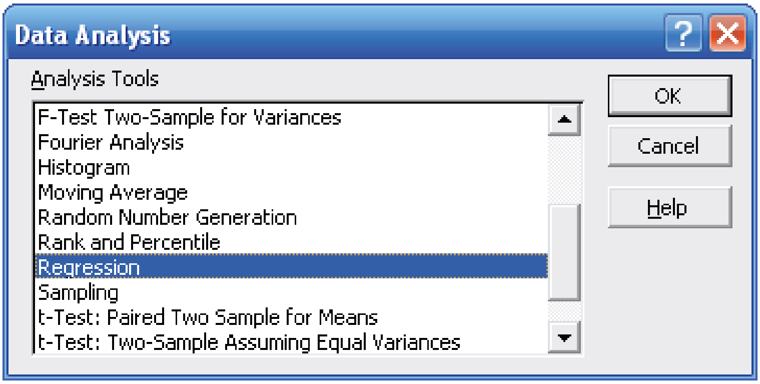


R^2 - coefficient of determination

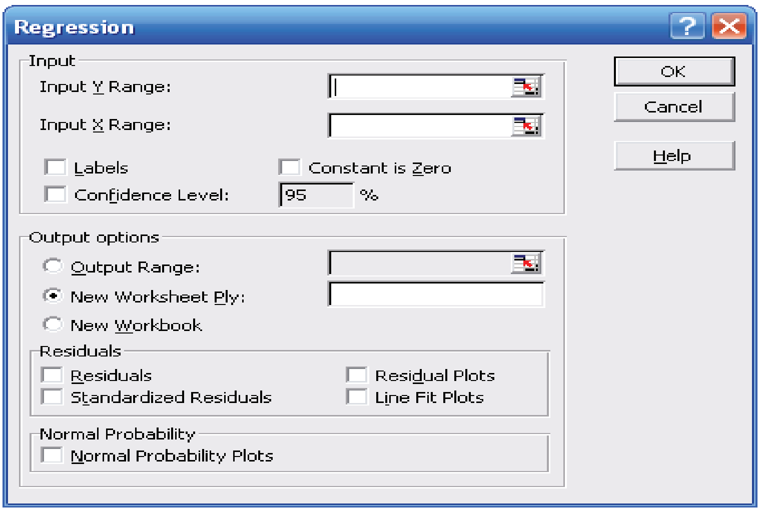
Once a line of regression has been constructed, one can check how good it is (in terms of predictive ability) by examining the coefficient of determination (R^2), which is defined as the proportion of variance of the dependent variable that can be explained by the independent variables. The coefficient of determination is a measure of how well the regression equation y = a + bx performs as a predictor of y. R2 always lies between 0 and 1. Higher values of this are generally taken to indicate a better model. The closer R2 is to 1, the better is the model and its prediction.

Regression Analysis using Microsoft Excel

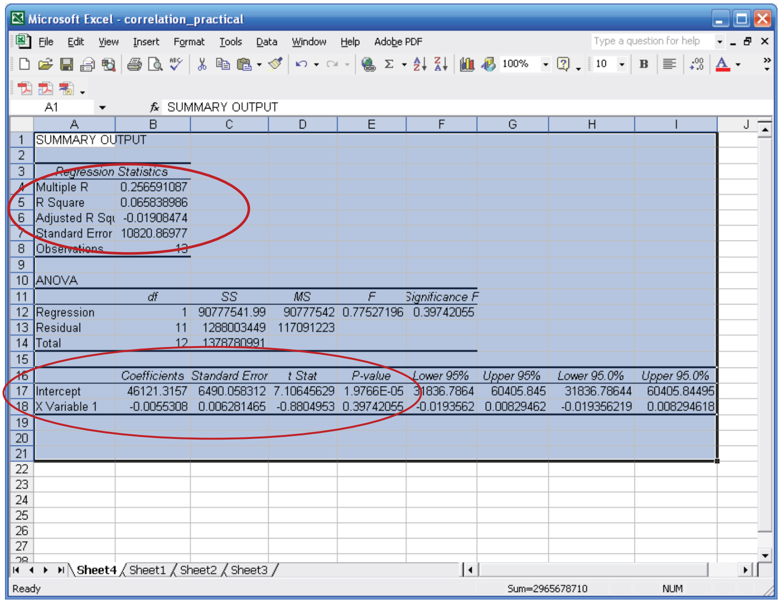
To do Regression in Microsoft Excel, choose **Data>Data Analysis>Regression**



- Enter the variable data, y as the dependent and x as the independent



- Check labels, if including column labels
- Check Residuals, Confidence levels to displayed them in the output
- The **SUMMARY OUTPUT** is displayed below



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.256591087
R Square	0.065838986
Adjusted R Squ	-0.01908474
Standard Error	10820.86977
Observations	13

ANOVA

	df	SS	MS	F	Significance F
Regression	1	90777541.99	90777542	0.77527196	0.39742055
Residual	11	1288003449	117091223		
Total	12	1378780991			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	46121.3157	6490.058312	7.10645629	1.9766E-05	31836.7864	60405.845	31836.78644	60405.84495
X Variable 1	-0.0055308	0.006281465	-0.8804953	0.39742055	-0.0193562	0.00829462	-0.019356219	0.008294618

The coefficients of the regression line can be obtained from the summary output. The slope of the line (b) is the coefficient corresponding to *X variable1 and* a value is the coefficient corresponds to intercept. To check whether your results are statistically significant, look at Significance F. If this value is less than 0.05, the model is OK. If Significance F is greater than 0.05, it's probably better to stop using this set of independent variables. For testing the significance of the regression coefficients you can check the corresponding p-value.

05

Introduction to Sampling Theory

Mini K. G.

Basic of Sampling

A "sample" is a miniature representation of a larger group or aggregate. In other words, the sample provides a specimen picture of a larger whole. This larger whole is termed as the “population” or “universe”. The procedure of collecting data for each and every unit belonging to the population, which is the aggregate of all units of a given type under consideration, is termed as census. The effort, money and time required for carrying out census will be enormously large. The census is necessary only if information is required for each and every unit in the population under study. Most of the time, it is impossible, difficult or expensive to observe all the elements of a population in order to arrive at a valid conclusion. Moreover, the sizes of populations are often so large that the study of all the units would not only be expensive but also cumbersome and time consuming, for example, the fishery resources in Indian Seas.

For fisheries research, it is impossible to collect information about the all the fishery resources in the sea. So, a researcher will have to select a representative few, i.e., a sample from the population of the survey. This process is known as sampling. In fisheries research, the sampling surveys are an essential ingredient of fish stock assessment. In sample survey, data is collected from a part or a sample of the population under study in a suitable manner and inference is drawn about the population on the basis of observations made on the selected units. If the samples are selected with an adequate criterion, it is possible to measure the precision of the conclusions or inferences about that population.

The objective of a sample survey is to estimate the unknown population parameters like total, ratio or median based on a random sample drawn by some specified rule from the population. A sample is considered as a subset of population. The principal advantages of sampling compared to census are reduction in cost, greater speed, wider scope and higher accuracy and more importantly the estimate error. Most fisheries research requires some planning and sampling designs before the research is undertaken. The objective of conducting sample surveys in fisheries is to collect data from the stocks and their exploitation level, to study the characteristics of the resources, the effects of exploitation on the abundance of these resources and to determine appropriate fishing levels to obtain the best possible catches at present and during future years. Fisheries research is most often concerned with the estimation of population mean and totals. e.g. the total catch by different types of gears in a particular landing centre.

Sampling error and Non-sampling error

Sampling error arises from the fact that samples differ from their populations in that they are usually small sub-sets of the total population. Non-sampling errors can be defined as errors arising during the course of survey activities rather than resulting from the sampling procedure. Non-sampling errors can occur because of problems in coverage, response, non-response, data processing, estimation and analysis.

Methods of Sampling

Sampling methods can be broadly classified into two categories:

Probability Sampling

Non-Probability Sampling

Probability Sampling: Probability sampling is based on random selection of units from a population. In other words, the sampling process is not based on the discretion of the researcher but is carried out in such a way that the probability of every unit in the population is pre-defined.

Some of the characteristics of a probability sample are:

- each unit in the sample has some probability of entering the sample,
- weights appropriate to the probabilities are used in the analysis of the sample and
- the process of sampling is automatic in one or more steps of the selection of units in the sample.

Non-probability Sampling: Non-probability sampling is based on the judgement of the researcher. The guiding factors in non-probability sampling include the availability of the units, the personal experience of the researcher and hidher convenience in carrying out a survey. Since these samples are not prepared through random sampling techniques, they are known as non-probability samples. Depending on the technique used, non-probability sampling includes purposive, incidental, snowball and quota sampling.

Sampling Scheme

In survey sampling a fixed finite population is under consideration, where the population elements are labeled so that each element can be identified. Probability sampling helps to select a random sample from a fixed population and assigns a probability of selection to each element. The probability of selection need not be equal for all the elements.

A specific sampling scheme is used in drawing the sample. The collection of techniques for the selection of the sample is termed as sampling scheme.

A sample scheme can be described by two factors.

- **Sampling method**. Sampling method refers to the rules and procedures by which some elements of the population are included in the sample. The basic sampling techniques which are commonly employed are simple random sampling, stratified sampling, cluster sampling, systematic sampling and sampling with unequal probabilities of selection of units particularly with probability proportional to size.
- **Estimator**. The estimation process for calculating sample statistics is called the estimator. Different sampling methods may use different estimators. For example, the formula for computing a mean score with a simple random sample is different from the formula for computing a mean score with a stratified sample. Similarly, the formula for the standard error may vary from one sampling method to the next.

The principal steps in a sample survey are given below.

i. Statement of the objectives of the survey: The objectives of the survey have to be defined clearly prior to the actual survey.

ii. Definition of the population to be sampled: The population will be used to denote the aggregate from which the sample is chosen. Whenever possible, the population to be sampled should obviously coincide with the population about which information is wanted.

iii. Determination of the data to be collected: It is well to verify that all the data are relevant to the purpose of the survey, and that no essential data are omitted.

iv. Methods of measurement: When the kinds of data that are needed have been decided, there may be a choice as to the methods of measurement to be employed.

v. Choice of sampling unit: As a preliminary to the selection of a sample, the population must be subdivided in some way into parts which will be called sampling units, or units. The sampling units must together comprise the whole of the population, and they must be non-overlapping, in the sense that every element in the population belongs to one and only one unit. The construction of a complete list of sampling units, sometimes called a frame, maybe one of the major practical problems. Sometimes the frame is impossible to construct, as with the population of fish in the sea.

vi. Selection of the sample: There are varieties of procedures by which the sample may be selected. The selection also involves a decision about the size of the sample, which in turn requires a provisional estimate of the cost of the survey, to ensure that the sample will fall within the allowable budget.

vii. Organization of the field work: In extensive surveys, many problems of administration are involved. The personnel must receive training in the purpose of the survey and in the methods of measurement to be employed and must be adequately supervised in their work. A procedure for early checking of the quality of the data may be invaluable. Plans must be made for handling non-response, that is, the failure of the enumerator to obtain information from certain of the units in the sample.

viii. Summary and analysis of the data: The first step is to edit the completed questionnaires, in the hope of amending recording errors, or at least of deleting data that are obviously erroneous. Decisions about tabulating procedure are needed in the case where answers to certain questions were omitted by some respondents or had to be deleted in the editing process. Thereafter, the tabulations which lead to the estimates are performed. Different methods of estimation may be available for the same data.

ix. Information gained for future surveys: The more information we have initially about a population, the easier it is to devise a sample which will give accurate estimates. Any completed sample is potentially a guide to improved future sampling, through the data which it supplies about the means, standard deviations, and nature of the variability of the principal measurements, and about the costs involved in getting the data.

References

Cochran, W. G. 1977. *Sampling Techniques*, third edition. John Wiley & Sons, Inc., New York, 428pp.

Hansen, M. H., W. N. Hurwitz, and W.G. Madow. 1953. *Sample Survey Methods and Theory*. John Wiley and Sons, Inc., New York.

Lohr, S. L. 1999. *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA, 494pp.

Sukhatme, P. V. and B. V. Sukhatme. 1970. *Sampling Theory of Surveys with Applications*. Iowa State University Press, Ames, IA. 452pp.

Thompson, M. 1997. *Theory of Sample Surveys.* Chapman & Hall, 312pp.

06

Simple Random and Stratified Sampling

Mini K. G.

Simple Random Sampling

Simple random sampling is a method of selecting a sample from a finite population in such a way that every unit of the population is given an equal chance of being selected. In practice, you can draw a simple random sample unit by unit through the following steps:

- Define the population
- Make a list of all the units in the population and number them from 1 to N.
- Decide the size of the sample, or the number of units to be included in the sample.
- Use either the 'lottery method' or 'random number tables' to pick the units to be included in the sample.

For example, you may use the lottery method to draw a random sample by using a set of 'N' tickets, with numbers ' 1 to N' if there are 'N' units in the population. After shuffling the tickets thoroughly, the sample of a required size, say n, is selected by picking the required n number of tickets. The units which have the serial numbers occurring on these tickets will be considered selected. The assumption underlying this method is that the tickets are shuffled so that the population can be regarded as arranged randomly. When the size of population is large, this procedure of numbering units on tickets and selecting one after reshuffling becomes cumbersome. Human bias and prejudice may creep in this method. Therefore, this method is generally discouraged.

The best method of drawing a simple random sample is to use a table of random numbers. These random number tables have been prepared by Fisher and Yates (1967). After assigning consecutive numbers to the units of population, the researcher starts at any point on the table of random numbers and reads the consecutive numbers in any direction horizontally, vertically or diagonally. If the read out number corresponds with the one written on a unit card, then that unit is chosen for the sample.

Simple random sampling can be regarded as the basic form of probability sampling applicable to situations where there is no previous information available on the population structure. There are two methods of simple random sampling namely simple random sampling with replacement (SRSWR) and simple random sampling without replacement

(SRSWOR). Sampling with replacement means that each unit selected in the sample is returned to the population before the next is drawn.

In SRSWR, one unit of element is randomly selected from population is the first sampled unit. Then the sampled unit is replaced in the population. The second sample is drawn with equal probability. The procedure is repeated until the requisite sample units *n* are drawn. The probability of selection of an element remains unchanged after each draw. The same units could be selected more than once. Let N denote the population size and n is sample size. As the population size remains the same after each draw, not only the probability of each unit being selected in the sample is $\frac{1}{N}$ at each draw, it remains same even when included in the sample more than once. In case of SRSWR, the number of all possible samples is N^n. The probability of drawing any of these N^n is $\frac{1}{N^n}$. For example, if there is three landing centres in a Tehsil, denoted by A, B, C (The population size N= 3). Then 3^2 = 9 possible samples of size 2 can be drawn through SRSWR. They are (A, A) (A B) (A C) (B A) (B B) (B C) (C A) (C B) (C C). Each sample can be selected with equal probability of $\frac{1}{9}$.

In SRSWOR, unlike SRSWR, once an element is selected as a sample unit, will not be replaced in the population pool. The selected sample units are distinct. SRSWOR is a method of selecting n units out of N such that every one of the ${}_N C_n$ distinct samples has an equal change of being drawn. The simple random samples are drawn unit by unit. The units in the population are numbered from 1 to N. A series of random numbers between 1 and N is then drawn, either by means of a table of random numbers or by means of a computer program that produces such a table. At any draw, the process must give an equal chance of selection to any number in the population not already drawn. The units that bear these numbers constitute the sample. With reference to the same example above in SRSWR, the possible samples with SRSWOR without giving any importance to the ordering of the units are (A B) (A C) (B C). The total number of samples is ${}_3 C_2 = 3$. Any of these samples have equal probability of selection and the probability of selection of each of the samples is $\frac{1}{3}$. So, in SRSWOR, the probability of selection of each unit at any draw is $\frac{1}{N}$ and the probability of inclusion of any unit in the sample is $\frac{n}{N}$.

Estimating the Population Mean

Let Y be the character of interest and $Y_1, Y_2, \dots, Y_i, \dots, Y_N$ be the values of the character on N units of the population. Let, $y_1, y_2, \dots, y_i, \dots, y_n$ be the sample of size n selected by SRSWOR.

The estimator for the population mean is given by

$$\hat{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

An unbiased estimator of variance of the population mean is given by

$$\hat{V}(\hat{\bar{Y}}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}$$, where $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad , \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Bound on the error of estimation (B)

$$B_y = t\,SE(\hat{\bar{Y}}) = t\sqrt{\hat{V}(\hat{\bar{Y}})}$$

Where *t* is the Student's t value for *n-1* degrees of freedom at the $1 - \alpha/2$ level of significance. For $\alpha = 0.05$, t = 1.96., Confidence Interval is given by C.I = $\hat{\bar{Y}} \pm B_y$

Estimating the Population Total

The estimator for the population total is given by

$$\hat{Y} = N\hat{\bar{Y}} = \frac{N}{n}\sum_{i=1}^{n} y_i$$

An unbiased estimator of the variance of the population total is given by

$$V(\hat{Y}) = N^2 V\left(\hat{\bar{Y}}\right) = N^2\left(1 - \frac{n}{N}\right)\frac{s^2}{n}$$

Where s^2 is an unbiased estimator of the population mean square S^2,

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 .$$

Advantages of Simple Random Sampling

One of the best things about simple random sampling is the ease of assembling the sample. It is also considered as a fair way of selecting a sample from a given population since every member is given equal opportunities of being selected. Another key feature of simple random sampling is that the sample will be a representative of the population. If the sample is not representative of the population, the random variation called sampling error will be large. An unbiased random selection and a representative sample are important in drawing conclusions from the results of a study.

Disadvantages of Simple Random Sampling

One of the most obvious limitations of simple random sampling method is its need of a complete list of all the members of the population. However, from a practical point of view, a list of all the units of a population is not possible to obtain for large populations. Even if it is possible, it may involve a very high cost which a researcher or an organisation may not be able to afford. Therefore, simple random sampling is difficult to realize. Also, in case of a highly heterogeneous population, a simple random sample may not necessarily represent the characteristics of the total population, even though all selected units participate in the investigation. In those cases, it is wiser to use other sampling techniques.

Stratified Random Sampling

When the population is heterogeneous, the whole population can be divided into sub-populations, called strata, to increase the precision of the estimates. In stratified sampling the population of N units is first divided into disjoint groups of N_1, N_2,.., N_h,.., N_L units, respectively. These subgroups, called strata, together they compromise the whole population, so that $N_1 + N_2 + ... + N_h + .. + N_L = N$. The strata should not overlap and each stratum should be sampled following some sampling design. The strata are sampled separately and the estimates from each stratum combined into one estimate for the whole population. If a simple random sample selection scheme is used in each stratum then the corresponding sample is called a stratified random sample.

Reasons for stratification

- To obtain estimates of known precision for certain subdivisions of the population by treating each subdivision as a stratum. Since sampling is done independently in each stratum, separate stratum estimates and their precision can be obtained by treating each stratum as a ``population'' in its own right. For example, in fishery surveys estimates may be required by state, district, month, landing centre, craft, species etc.
- For administrative convenience; for example stratification can provide survey organization to control the distribution of fieldwork among its regional offices.

- Sometimes different parts of the population may call for different sampling procedures.
- Stratification may often produce a gain in precision of the estimates of characteristics of the whole population. The amount in the gain depends on the type of stratification. If the population is heterogeneous and if it can be divided, using prior information about the population, into subpopulations (strata), each of which is internally homogeneous. If each stratum is homogeneous, that is characteristic under consideration vary little from one unit to another, a precise estimate (an estimate with smaller variance) of any stratum parameter can be obtained from a small sample in that stratum. These estimates can then be combined to obtain a precise estimate for the whole population.

Notations

The suffix h (h =1,2,...,L) denotes the stratum and i the unit within the stratum.

N_h - Total number of population units in stratum h.

n_h - Total number of sample units in stratum h.

$w_h = N_h / N$ - The h[th] stratum weight.

Y_{hi} - Value of the characteristic for the i^{th} unit in stratum h.

$Y_h = \sum_{i=1}^{N_h} Y_{hi}$ - Population total of Y – values for units belonging to stratum

$\overline{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi} = \frac{Y_h}{N_h}$ - Population mean of Y-values for units belonging to stratum h.

$\sigma_h^2 = \frac{1}{N_h}\sum_{i=1}^{N_h}\left(Y_{hi} - \overline{Y}_h\right)^2$ -Population variance of Y-values for units belonging to stratum h.

$$\overline{Y} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{N_h} Y_{hi}}{\sum_{h=1}^{L} N_h} = \sum_{h=1}^{L} W_h \overline{Y}_h$$ - Population mean of Y-values.

$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi} = \frac{y_h}{n_h}$ - Sample mean of Y-values for units belonging to stratum h.

$s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(y_{hi} - \bar{y}_h\right)^2$ - Sample variance of Y-values for units belonging to stratum h.

$\bar{y} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h} Y_{hi}}{\sum_{h=1}^{L} n_h}$ - Sample mean of Y-values.

Estimation of Population Mean and Variance

An unbiased estimator of $\bar{Y}$ is given by

$$\hat{\bar{Y}} = \sum_{h=1}^{L} W_h \bar{y}_h$$

Also, since sampling is done independently within each stratum

$$V\left(\hat{\bar{Y}}\right) = \sum_{h=1}^{L} W_h^2 \frac{N_h - n_h}{N_h - 1}\sigma_h^2$$

Note that $V\left(\hat{\bar{Y}}\right)$ can not be computed since it involves Y-values for all the units in the population. However, based on Y-values for the sampled units we can estimate $V\left(\hat{\bar{Y}}\right)$ by using the following formula.

$\hat{V}\left(\hat{\bar{Y}}\right) = \sum_{h=1}^{L} W_h^2 \frac{N_h - n_h}{N_h n_h} s_h^2$ which estimates $V\left(\hat{\bar{Y}}\right)$ unbiasedly.

Allocation in Stratified Random Sampling

In planning a study, requiring stratification of the population, an important consideration is how to allocate a total sample size n among the L identified strata. There are three types of allocation.

Allocation of a Sample to Strata

1. Equal: If the strata are presumed to be of roughly equal size, and there is no additional information regarding the variability or distribution of the response in the strata, equal allocation to the strata is probably the best choice. $n_h = n/L$

2. Proportional: If the strata differ in size, allocation of sample sizes to strata might be performed proportional to these stratum sizes. $n_h = \left(\frac{N_h}{N}\right)n$

3. Optimum (Neyman): The allocation which minimizes the variance of the estimator of the mean (and total) is given by $n_h = \frac{nN_h\sigma_h}{\sum_{h=1}^{L} N_h\sigma_h}$

Whenever the strata are heterogeneous among themselves and the variance of each stratum is small, the sampling variance of the mean or total value estimators obtained by stratified sampling will always be smaller than the simple random sampling. The relative sizes of strata must be known to obtain the full benefits of the stratification technique. Each stratum should be internally homogeneous. If information about heterogeneity is not available then consider all strata equally variable. A short stratified pilot survey can sometimes provide useful information about internal dispersion within strata. A small sized sample could be taken from a stratum if the variability among their units is small. A larger sample from a stratum should be taken if the stratum is larger, the stratum is more heterogeneous and the cost of sampling the stratum is low.

References

Cochran, W. G. 1977. *Sampling Techniques*, third edition. John Wiley & Sons, Inc., New York, 428pp.

Lohr, S. L. 1999. *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA, 494pp.

Sukhatme, P.V. and B.V. Sukhatme. 1970. *Sampling Theory of Surveys with Applications*. Iowa State University Press, Ames, IA. 452pp.

Thompson, M. 1997. *Theory of Sample Surveys.* Chapman & Hall, 312pp.

07

Systematic Sampling and Cluster Sampling

J. Jayasankar

Systematic Sampling

The sampling methods which are more popular and usually precede a discussion on Systematic sampling basically involve the methodology of selection of successive units with the help of random numbers. But there are situations where only the first unit is selected at random and the rest all the other units get selected automatically thereafter. This method is called systematic sampling.

As the entire sample is selected based on one unit, the mode of selection is based on a pattern. Such pattern involves regular spacing of units either spatially or temporally whose first occurrence is amenable to selection randomly. Suppose a population contains N units serially numbered from 1 to N and suppose further that N is expressible as a product of two integers k and n so that N=kn. Draw a random number less than k, say i, and select the unit with corresponding serial number and every kth unit in the population thereafter. Obviously the sample will contain the n units i,i+k, i+2k, ..., i+(n-1)k and is known as a systematic sample.

Some examples

(a) The selection of a maize (corn) field, every k^{th} mile apart, for the observation on pest incidence.

(b) Observing every k^{th} strip of forest to estimate timber.

(c) Selection of every k^{th} interval of time (say two hourly) for observing the number of fishing craft landing on the coast.

(d) Selection of every k^{th} village from a list of villages.

Systematic sampling which obtains the basic property of a probability sample just because of the selection of the first unit has many practical advantages. This method is simple to adopt and economical. The field level workers can easily succeed in performing this type of sampling as it has very little technical maneuvering. Further this ensures better control over the field work. By its very design, the systematic sampling ensures even spread of the sample through the length and breadth of the population.

On the disadvantage side, the relative position in the population of the different units included in the sample is fixed. Hence if the population exhibits some cyclicity then this method is liable to pitfalls. For example if every 7^{th} day is a holiday in a locality if the task is to estimate daily vehicular traffic, then a systematic sample selected with a typical holiday would be highly skewed. Another significant drawback of this method is that as only one cluster or group of preset units are selected, estimation of variability of the estimated mean, technically called as sampling variance is not obtainable.

Similarities with other methods

Systematic sampling resembles Stratified Random Sampling in that one sampling unit is selected from each stratum of k consecutive units. This similarity is strictly casual as the procedure followed while adopting stratified sampling is quite different. Systematic sampling resembles cluster sampling, a systematic sample being equivalent to a sample of one cluster selected out of k clusters of n units each.

The following schematic diagram shows the possible representation of a population considered for systematic sampling in clusters. (Sukhatme (1954)).

Row No.

Cluster Number	1	2	...	i ...	...	k
1	1	2	.	i	.	k
			.		.	
2	1 + k	2 + k	.	i + k	.	2k
.	.	.	.	.	.	.
j	1+ (j-1)k	2+(j-1)k	.	i+(j-1)k	.	jk
.	.	.	.	.	.	.
n	1+(n-1)k	2+(n-1)k	.	i+(n-1)k	.	nk

Since the first number less than or equal to k is to be chosen at random, every one of the k columns gets an equal chance of being chosen as the systematic sample.

Estimates of parameter

Under a systematic sampling design, the estimate of the commonest parameter, mean, is given by

Sample mean = sum of all the values of selected units comprising the sample/ the number of units selected or using notations, Sample mean $\bar{y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij}$

And this is used as the estimate of the population mean given by

$$\bar{y}_{..} = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}$$

The variance of the estimate is given by

$$\frac{\text{Sum of squared deviations of the means of each cluster from the population mean}}{\text{number of clusters (k)}}$$

or in notations

$$V(\bar{y}_{i.}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2$$

It is this parameter, the sampling variance of mean, which cannot be estimated as only one $\bar{y}_{i.}$ Is known at the end of the sampling operation and that itself is the estimate of population mean.

An example:

Consider a population with N=9, consisting of 1, 2, 3, 4, 5, 6, 7, 8, and 9. The arrangement for systematic sample plan is as follows:

k = 1	k = 2	k=3
1	2	3
4	5	6
7	8	9

If 1-in-3 systematic sample is taken assuming that n=3 and k=3, the following three samples are possible:

Sample 1: 1, 4, 7: mean=4

Sample 2: 2, 5, 8: mean=5

Sample 3: 3, 6, 9: mean=6

That means had we selected 1 as the initial random number the sample mean would have been $\bar{y}_{1.} = 4$. By definition, Var($\bar{y}_{1.}$) = (1 + 0 + 1)/3 = 2/3

An approximate estimate of variance could be the following.

$\hat{V}(\bar{y}_{1.}) = (\frac{1}{n} - \frac{1}{nk})(\frac{1}{n-1}\sum_{j=1}^{n}(y_{ij}^{2} - n\bar{y}_{i.}{}^{2}))$ which of course is not an unbiased estimate of the variance. For the said example the estimated variance is $(\frac{1}{3} - \frac{1}{3x3})(\frac{1}{3-1} * 18)$= 2.

Cluster Sampling

A sampling design pre supposes the division of the population into a finite number of distinct and identifiable units called as sampling units. Thus a population of fields under paddy in a given region might be regarded as composed of fields or groups of fields, on the same holdings, villages or other suitable settlements. A human population might similarly be regarded as composed of individual persons, families, or groups of persons residing in houses in villages. The smallest units into which the population can be divided are called the elements of the population and groups of elements are called as clusters. When the sampling unit is a cluster, the procedure of sampling is called cluster sampling. When the entire area containing the population under study is sub divided into smaller areas and each element in the population is associated with one and only one such small area, the procedure is alternatively called area sampling.

For many types of population a list of elements is not available and the use of an element as the sampling unit is therefore not feasible. The method of cluster or area sampling is available in such cases. Thus in a city a list of all the houses is readily available, but that of persons is rarely available. Again lists of fields are not available but those of villages are. Cluster sampling is therefore widely practiced in sample surveys.

The unique situations on the field could leave the clusters to be selected as equal or unequal. The procedures for estimation will slightly vary depending upon the cluster sizes and their equality. Here let us focus on the case of equal sized clusters.
We shall first consider the case of equal clusters and suppose that the population is composed of N clusters of M elements each, and that a sample of n clusters is drawn from it by the method of Simple Random Sampling.

For notational simplicity, let

y_{ij} denotes the value of the character for the jth element (j=1,2,..., M) in the ith cluster (i=1,2,..,N)

$\bar{y}_{i.} = \frac{1}{M}\sum_{j=1}^{M} y_{ij}$ be the mean for the elements of the ith cluster

$\bar{y}_{N.} = \frac{1}{N}\sum_{i=1}^{N} \bar{y}_{i.}$ be the mean of cluster means in the population

$\bar{y}_{..}= \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}$ be the mean per element in the population and

$\bar{y}_{n.}= \frac{1}{n}\sum_{i=1}^{n} y_{i.}$ be the mean of cluster means in a simple random sample of n clusters and this is the unbiased estimator for the population mean.

The variance of this mean is given by

$$V(\bar{y}_{n.}) = \frac{N-n}{Nn}\left(\frac{1}{N-1}\sum_{i=1}^{N}(\bar{y}_{i.} - \bar{y}_{N.})^2\right)$$

And its estimate being

$$\hat{V}(\bar{y}_{n.}) = \frac{N-n}{Nn}\left(\frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_{i.} - \bar{y}_{n.})^2\right)$$

An example

Let the catch from points in a harbour (11 points) wherein designated boats each are licensed to land for a month is given below.

Date/ Landing point	1	2	3	4	5	6	7	8	9	10	11	Means of selected clusters
01-Jan	61	72	40	87	36	36	68	75	98	69	26	
02-Jan	22	48	31	95	66	20	34	94	27	49	51	48.81818
03-Jan	1	76	34	77	31	15	86	8	95	27	81	
04-Jan	72	83	38	82	87	28	70	46	48	92	82	
05-Jan	40	22	24	76	97	82	8	83	82	54	75	58.45455
06-Jan	83	50	82	31	37	17	6	70	2	37	37	
07-Jan	59	26	81	92	0	32	83	45	31	60	84	
08-Jan	9	12	92	78	100	19	62	70	9	82	55	
09-Jan	97	31	59	57	71	31	50	79	42	79	83	
10-Jan	46	35	1	89	21	75	66	77	95	46	97	58.90909
11-Jan	76	76	80	68	45	20	27	86	56	59	17	
12-Jan	58	64	11	28	45	83	10	24	32	17	68	40
13-Jan	43	75	94	49	84	35	20	57	10	63	43	
14-Jan	77	4	3	97	12	93	8	52	15	98	87	
15-Jan	81	6	96	79	88	63	80	76	47	66	77	
16-Jan	12	52	73	54	8	23	53	31	57	71	13	

17-Jan	57	59	95	34	9	85	5	99	55	44	60	54.72727
18-Jan	36	27	42	45	84	76	67	19	28	99	80	
19-Jan	11	93	6	3	71	62	73	2	36	85	71	
20-Jan	33	20	99	63	70	30	80	51	40	4	22	
21-Jan	38	66	95	67	52	31	83	62	54	28	62	
22-Jan	59	84	94	11	13	87	57	76	48	92	86	
23-Jan	87	81	25	16	93	85	44	14	61	60	20	
24-Jan	8	38	18	88	50	17	75	70	81	59	50	
25-Jan	27	87	58	97	77	37	9	89	93	2	3	
26-Jan	1	58	63	6	25	49	57	84	76	40	43	
27-Jan	57	42	84	39	5	86	89	54	68	83	86	63
28-Jan	97	17	19	53	75	41	91	72	9	51	56	
29-Jan	66	12	43	15	19	74	60	70	37	73	19	
30-Jan	1	75	18	4	36	20	29	98	80	53	15	
31-Jan	55	76	94	68	36	77	91	65	65	68	60	
sample mean												53.98485
population mean												**53.32551**
estimated variance												9.363466

Suppose the month is sampled on the dates (2,5,10,12,17 & 27) that is n=6 against N=31 and M=10. Then the estimate of the resource landed based on the sample is 53.985.

Its estimated variance is as follows

$\hat{V}(\bar{y}_{n.}) = \frac{31-6}{(31*6)}(\frac{1}{6-1}((48.81818 - 53.985)^2 + \cdots + (63 - 53.985)^2)))$=9.363.

However as the entire population is known we can compare the estimated mean with the population mean. For this data set the estimated mean is 53.985 and the population mean is **53.325**.

Thus systematic and cluster sampling procedures form a very important pair of sampling designs which are quite practical and economically viable in estimating parameters of interest.

References

Sukhatme, P. V. 1954. Sampling Theory of Surveys with Applications: The Indian Society of Agricultural Statistics: New Delhi. p.491

08

Multistage Sampling

Mini K. G.

In the cluster sampling method, all the elements of the selected clusters are enumerated. Though the cluster sampling is convenient and economical, the method restricts the spread of the sample over the population and it results in the reduction in the efficiency of the estimator. There are many ways to modify cluster sampling for more complex sampling situations. One common modification is to take a sample of secondary units from within sampled clusters instead of inspecting every secondary unit within each sampled cluster. The process of drawing samples from the selected clusters is called sub-sampling.

The sampling method which consists of selecting the clusters and then randomly choosing a specified number of units from each selected cluster is known as two-stage sampling. The clusters which form the units of sampling at the first stage are called first stage units (fsu's) or primary stage units (psu's) and the elements within the clusters are called second stage units (ssu's). For example, to obtain a sample of fish caught in a commercial fishery, it may be necessary to first take a sample of boats and then take a sample of fish from each selected boat. The simplest type of multistage sampling is two-stage sampling. The estimation mean and variance is given below.

Let N be the number of primary units in the population and n be the number of primary units selected in the first stage. If each primary unit contains M subunits, of which m are chosen, the

$\bar{y}_{ij}$ = value obtained for the j^{th} element in the i^{th} primary unit

$\bar{y}_i = \sum_{j=1}^{m} \frac{y_{ij}}{m}$ = sample mean per element in the i^{th} primary unit

$\bar{\bar{y}} = \sum_{i=1}^{n} \frac{\bar{y}_i}{n}$ = over-all sample mean per element

$S_1^{\ 2} = \sum_{i=1}^{N} \frac{(\bar{Y}_i - \bar{\bar{Y}})^2}{N-1}$ = variance among primary unit means

$S_2^{\ 2} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \left(y_{ij} - \bar{Y}_i\right)^2}{N(M-1)}$ = variance among elements within primary units

The simplest estimates of the population total and mean per subunit are, respectively,

$$\hat{Y} = \frac{NM}{n}(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_n),$$

$$\hat{\bar{\bar{Y}}} = \frac{1}{n}(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_n)$$

Where $\bar{y}_i$ is the sample mean per subunit in the i^{th} primary unit

If the *n* units and the *m* subunits from each chosen unit are selected by simple random sampling, $\bar{\bar{y}}$ is an unbiased estimate of $\bar{\bar{Y}}$ with the variance

$$V(\bar{\bar{y}}) = \left(\frac{N-n}{N}\right)\frac{{S_1}^2}{n} + \left(\frac{M-m}{M}\right)\frac{{S_2}^2}{mn}$$

An unbiased estimate of $V(\bar{\bar{y}})$ is

$$v(\bar{\bar{y}}) = \frac{1-f_1}{n}{s_1}^2 + \frac{f1(1-f_2)}{mn}{s_2}^2$$

where $${s_1}^2 = \frac{\sum_{i=1}^{n}(\bar{y}_i - \bar{\bar{y}})^2}{n-1}, \quad {s_2}^2 = \sum_i^n \sum_j^m \frac{\left(y_{ij} - \bar{y}_i\right)^2}{n(m-1)}$$

The two stage sampling method can be generalized to three or more stages and is termed as multistage sampling. In one-stage cluster sampling, the estimate varies because due to one source: different samples of primary units yield different estimates. In two-stage cluster sampling, the estimate varies due of two sources: different samples of primary units and then different samples of secondary units within primary units. In general, if there a k stages of sub-sampling, there will be k sources of variability. Thus, variances and variance estimators for multistage sampling with k-stages will contain the sum of k components of variability.

Multistage sampling designs are used for a variety of practical reasons and this procedure is being commonly used in large scale surveys. This sampling procedure is more flexible as it permits the use of different selection procedures at different stages. The multistage sampling may be the only choice in a number of practical situations where a satisfactory sampling frame of ultimate stage units is not readily available and the cost of obtaining such a frame is large and time consuming.

Reference

Cochran, W. G. 1977. *Sampling Techniques*, third edition. John Wiley & Sons, Inc., New York, 428pp.

09

Sampling Design Followed by CMFRI for Deriving Estimates of India's Exploited Marine Fishery Resources

Mini K. G.

Sampling Design

India has a coast line of about 8219 km. Marine fish landings take place almost all along in the coast line throughout the day and sometimes during night. According to marine fisheries census 2010, there are 3288 fishing villages scattered along the coast line from where fishermen go for fishing and return to a landing centre which may be distinct from the fishing village. There are 1511 landing centres scattered along the coastline of the main land. Under these conditions collection of statistics by complete enumeration would involve a very large number of enumerators and a huge sum of money apart from the time involved in collection of data. In this situation a feasible solution for obtaining marine fish landings is the adoption of a suitable sampling technique for the collection of fish landing data. The sampling design adopted by the Central Marine Fisheries Research Institute (CMFRI) to estimate resource-wise/region-wise marine fish landings is based on stratified multi-stage random sampling design, in which, the stratification is over space and time. Over space, each maritime state is divided into suitable, non-overlapping fishing zones on the basis of fishing intensity and geographical considerations (Fig. 1).

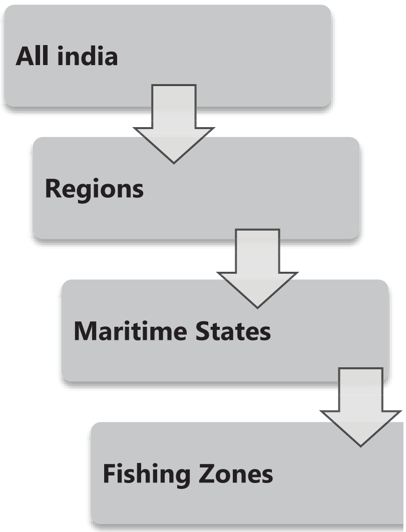


Figure 1. Stratification over Space

Selection of Landing Centre Days

The number of landing centres may vary from zone to zone. These zones have been further stratified into substrata, on the basis of intensity of fishing. Each zone is regarded as a stratum in space. The stratification over time is by calendar month. A zone and a calendar month constitute a space-time stratum. If in a zone, there are 25 landing centres and there are 30 fishing days in the month; we get 25 x 30 = 750 landing centre days which constitute the primary stage units (PSU). The fishing boats that land on a landing centre day forms the second stage units (SSU). The introduction of space–time stratification in the sampling methodology becomes necessary as the fish population is supposed to vary with respect to both space and time. The stratification is intended to reduce the variance in the sample estimates. The fish landings are found to vary considerably among the landing centres in a multi-centre zone, especially in different seasons and hence a zone is further stratified as major, minor and very minor centres etc. The centres in which either mechanised boats or 100 or more non-mechanised/ motorised boats are operating are considered as major centres. Similarly other strata are defined based on the number and type of fishing boats operating.

Further a month is divided into 3 groups, each of 10 days. From the first five days of a month, a day is selected at random, and the next 5 consecutive days are automatically selected. From this, three clusters of two consecutive days are formed. For example, for a given zone, in a given month, from the five days if the date (day) selected at random is 4, then the clusters formed from the first 10 day group are (4, 5), (6, 7) and (8, 9). In the remaining ten day groups, the clusters are systematically selected with an interval of 10 days. For example, in the above case, the cluster of days for observation in the remaining groups are (14, 15), (16, 17), (18, 19); (24, 25), (26, 27) and (28, 29). Normally, in a month 9 clusters of two days each can be obtained. From among the total number of landing centres in a zone, 9 centres are selected with replacement and allotted to the 9 cluster days selected as described earlier. These 9 days are evenly distributed among the strata in case of fishing zones with more than 1 landing centers. A landing centre day which is the PSU is the 24 hour duration from noon of the first day to the noon of the following day.

A landing centre day has been divided into 3 periods as given in the infographic below.

- Period 1: 1200 to 1800 hours on 1st day
- Period 2: 0600 to 1200 hours on 2nd day
- Period 3: 1800 hours to next morning 0600 hours

One field staff is usually provided to each zone. A field staff starts data collection from period 1 on each selected landing centre day. He will be present throughout the periods 1 and 2 at the centres. The data on landings during period 3 (night landings) is usually collected from the landing centre by enquiry on the following day morning. The sum of the observations on the 3 periods contribute the data for one landing centre day (24hrs). Thus in a 10 day period, data from 3 centre-days are sampled and consequently in a month 9 landing centre-days are sampled

Selection of fishing units and Recording of Fish Landings

It may not be practicable to record the catches of all boats landed during an observation period, if the number of boats/craft is large. A sampling of the boats/ craft become essential. When the total number of boats landed is 15 or less, all the boats are enumerated for catch composition and other particulars. When the total number of boats exceeds 15, the following procedure is followed to sample the number of boats.

Less than or equal to 15	• 100%
16 -19	• First 10 and the balance 50 %
20 - 29	• 1 in 2
30 - 39	• 1 in 3
40 - 49	• 1 in 4
50 - 59	• 1 in 5 and so on

From the boats, the catches are normally removed in baskets of standard volume. The weight of fish contained in these baskets being known, the total weight of the fish in each boat under observation has been obtained. The procedures of selection of the landing centre days and the boats landed on the selected day for single centre zones are the same as in the case of a stratum in a multi-centre zone.

Estimation Procedure

From the landings of the observed fishing units, the landings for all the units landed during the observation period are estimated. By adding the quantities landed during the two 6-hour's periods and during the night (12-hours) the quantity landed for a day (24-hours) at a centre that is the landings for each centre day included in the sample is estimated. From these, the monthly zonal landings are obtained. From the zonal estimates, district-wise, state-wise and all India landings are arrived. The corresponding sampling errors are also estimated. The estimation procedure is detailed in Srinath *et.al.* (2005).

Administration of the Survey

Plan of operation

The survey staff is given 10-12 weeks training course immediately after recruitment and is posted to the survey centres. They are permanent employees. Each survey centre is housed in 1-2 room accommodation and each centre is provided with literature connected with the identification of fish, a reference collection of local fish species, crustaceans and molluscs, field notebooks and registers. The programme of work for the following month is carefully designed by the staff of Fishery Resources Assessment Division at the CMFRI headquarters. Generally one field staff is allotted to each zone to collect the fish landings data. At the end of every month, the survey staff receives the programme of work for the next month by post, that includes the names of landing centres to be observed and details such as dates and time for observations at each landing centre. The field staff are instructed to send the data collected during every month to reach the Institute's headquarters at least by the end of first week of the subsequent month.

Supervision of data collection

Surprise inspections are carried out by the supervisory staff of the Institute and the enumerators are inspected while at work in the field and their field notebooks and diaries are scrutinised. The estimated zonal landings are always compared with the previous year's survey figures, and if any variation which cannot be explained is observed, the technique of interpenetrating sub-samples is adopted to detect observational errors. Zonal workshops are held periodically to review the progress of work and update the sampling frame and to impart refresher courses to the field staff.

Errors due to non-response, their magnitude and control

Non-response occurs when the regular field staff is not available to observe the centre-day included in the sample. Usually, arrangements are made at the Headquarters/Research/Regional Centre to minimise the non-response.

Analysis of Data

In the existing sampling methodology, the interest is to estimate gear-wise, species-wise landings for the state in a month, fishing effort according to different types of fishing boats and also in terms of man hours. The analysis is carried out at CMFRI headquarters. Before the data is processed for analysis it will be ensured that the data collection is made as per the approved schedule, by checking the appropriate proforma. The responsibilities and functions of staff at the headquarters are data coding, estimation and database management. The data analysis is computerised and estimates are made using the software developed by the Fishery Resources Assessment Division of the Institute. The processed data are again counter- checked for errors. When discrepancies are detected, the estimation procedure is scrutinised in detail.

Reference

M. Srinath, Somy Kuriakose and K. G. Mini, 2005. Methodology for the Estimation of Marine Fish Landings in India, CMFRI Special Publication No. 86, p.57

10

Sampling for Collection of Biological Data

K. K. Joshi

Introduction

In Fisheries Science the word stock is defined as a sub-set of one species having the same growth and mortality parameters, and inhabiting a particular geographical area. Most of the parameters in the stock assessment cannot be obtained or measured throughout the whole population. In fish population studies, it is impossible to measure the whole population and also the fish caught. Hence a sample of the population is collected assuming that a reasonable estimate can be obtained of the true value of the sample population. Random sampling can be defined as a sampling from a population where each entry has an equal chance of being drawn. In practical terms this means that any fish from the stock under investigation have the same probability of being sampled. Care should be taken to sample from a mixed or unsorted catch of commercial landing.

Field sampling

Fish landing centres provide a landing ground for a variety of crafts. The gear combination like mechanised trawl nets from multiday and single day operations, mechanised purse seines, mechanised drift gillnets, mechanised hook and lines, outboard hook and lines, outboard ring seines and non-mechanised sectors. Sampling was done at weekly intervals from as far as possible and if not and at fortnightly intervals. On each sampling day the units to be sampled were selected following Alagaraja (1984) and from each of the vessels sampled, information on depth, mesh size, catch etc. were collected.

Area of the trawling operation	• by enquiry
Depth of trawling	• by enquiry
Mesh size of the cod end of trawl net used	• by direct observation
The total fish catch	• by observation

The sampling strategy for species composition and body is to be designed to take into account the differences arising out of sorting the catch on board, depth and area of fishing. Catch sampling is carried out at the landing centre and care should be taken to avoid 'bias' affecting the sampling. Regarding the size of the 'sub-sample' it should be 1-2 kg from smaller fishes, 4-5 kg for medium sized fishes and 6-9 kg for bigger fishes. From each unit collect a random sample and place in a plastic bag, note down the total catch of fish in the boat. Using the raising factor the abundance of the different species in the catch can be calculated. Total length (TL) has to be taken at the landing centre using the measuring board and individual total weight (w) is collected using portable balance. Also record all the additional information i.e. total catch, mesh size, number of hauls, hooks and any other significant observation on weather and social issues.

Sampling form the landing centre can be performed directly and the routine length frequency and weight of the data can be collected for population analysis studies. The sampling for the inshore gears will allow to gather information on the temporal patterns of the population structure of same species in shallow coastal waters and an impact of such gear on them.

Laboratory analysis

Fish samples collected from the lading centre are kept in the ice box in separate bags have to be analysed as soon as possible in the same day. Individual total length (TL) measurements from tip of the snout to tips of the largest caudal fin rays are made with the fish placed on its right side, snout to the left on the measuring board. Note down the weight of the fish. Cut open the fish and examine sex and stage of fish. If difficulty arises to do the sex, examine the gonads using lens or dissection microscope. Then the different stages of sexual maturity have to be identified. It is possible to collect more data from the individual fish such as otolith, gonad weight, individual body weight and egg numbers and ova diameter frequency.

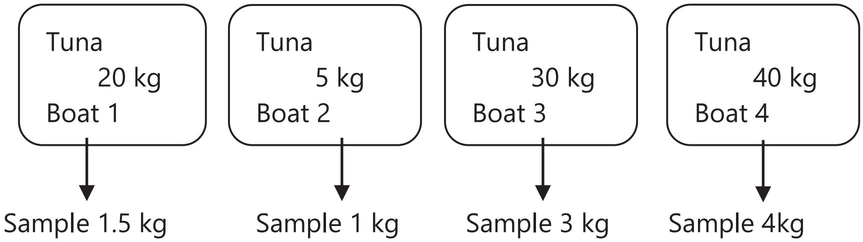
Analysis of fish samples: In the laboratory thaw and sort the samples by species. Take weight of each species in each plastic bag and raise to the estimated catch of species in the boat sampled. Take data on length on each species in each bag and weigh to the total catch of the above species in the boat. Then estimate weights of each species in the different boats sampled and pool and raise to the total catch of species for the day. Similarly the estimate length composition of the catch of each species from different boats and pool and raise to the total catch of the species estimated on the day. The data on species composition and length composition collected on each observation day were weighted respectively to the estimated total catch of the group and species obtained on that day and pooled and then raised to the estimated catch of the month.

Total catch weight: The estimated weight of the total catch obtained during that day by the sampled fishing unit.

Sample weight: Is the weight of the unsorted sample, all species mixed.

Sample species weight: Is the weight in the sample made by that species alone. It means the summation of the length groups of total weight.

Scheme of sampling of species illustrated



For studies on species composition of tuna and length composition, each species in each sample were first weighted to the total catch of tuna in the boat. Thus from the above example, from the 1.5 kg of species collected from Boat I containing 20 kg of tuna, the weight of each species in the sample is first raised proportionately to 20 kg. when each sample is thus raised to the catch of the boat the total tuna catch of the 4 boats are added up (i.e. 20+5+30+40= 95 kg) and the species composition and length composition of each boat is also added up to get the consolidated estimate in the 4 boats (95 kg) sampled. The consolidated tuna catch of the boat is raised to the day's catch and the consolidated species and length composition of the four boats are also raised to the day's catch to get the estimate for the day.

Maturation studies

After measuring the length and weight of each specimen, the belly was cut open to note the sex, colour and general appearance of the gonads, which were then carefully removed and preserve in 5% formalin, in labelled bottles. If the spawning is synchronous in the population, the studies on the maturation and spawning were carried out mainly on the basis of the ovaries otherwise the development of male gonads was also to be carried out separately.

i. Quantification of ovaries into different stages of maturation: Examine the appearance of the ovaries in fresh condition and note down the proportion of the area occupied by them in the body cavity. The structure and diameter range of the intraovarian ova were considered for quantification of ovaries into different stages of maturation.

ii. Ova diameter frequency distribution: For measurement of ova diameters, take transverse sections from the anterior, middle and posterior regions of the ovary, then tease ova out on micro slides taking utmost care to separate out all the ova in the samples. Observe ova diameter under the microscope, using an ocular micrometer, at a set magnification where, one micrometer division equals XXX mm. Measure the ova from the different regions of the ovary to see whether differences in the distribution of ova in different regions of the ovary occur.

In the immature ovaries in which the ova were minute and it is not possible to arrange them in rows; they were spread evenly on the slide and the diameters of the ova lying parallel to two horizontal guidelines on the slide were measured. In mature ovaries, samples for diameter measurements were taken after noting the total weight of the ovary and the sample weight. Arrange the ova diameters into groups by micrometer division (md) class intervals (i.e., 1-3, 4-6, 7-9 etc.) to determine the frequency distribution of ova in the ovary.

iii. Determination of length at first maturity: For determining the length at first maturity (L50), specimens with ovaries in stages IV and V of maturation have to be considered as mature and the proportion of such mature fish in each length group determined. The length at which about 50% of the fish are mature, has been taken as the L50. As almost all the adult fish during the spawning season (or at least peak spawning season) are expected to be in mature stage, it is desirable to consider representative samples collected during this period for this purpose to eliminate the possibility of growth in length influencing the estimate of the length at first maturity. erimentation. The following table tries to sum up the plan to be adopted during the course of scientific investigations and the type of data to be collected and the checklist to be prepared before venturing into such data collection.

iv. Spawning: The periodicity of spawning has been determined using the ova diameter frequency distribution in mature ovaries following Hickling and Rutenberg (1936) and De Jong (1940). The spawning period was determined using the data on maturation stages in different months and the months of occurrence of gravid fishes has been taken as the spawning period. For this purpose only fishes of and above the length at first maturity (L50) were considered as this would help in determining the peak spawning period more satisfactorily. The gonado-somatic index has also been studied for the purpose.

v. Fecundity: The estimated number of mature ova in ovaries represents the fecundity. (Total weight of ovary/weight of the sample) X Number of mature ova in the sample

Data records and files

Forms are the paper sheets used to the data either in the field or in the laboratory, whereas the tables are computer spreadsheets used for data input. First data collect in the forms and later input into the computer. The main purpose of the form is to make data collection work quicker and simple. This can be modified on the basis of experience gained during the initial weeks of sampling and modifications and adjustments aim to make available form more user friendly and to avoid confusion (Form No.1 and Form No.2).

Data input tables

There are two main types of spreadsheets: table for L/F and weight data and tables for maintaining maturation stages. All the tables are containing mainly weekly data; it is likely that the analysis work will be carried out on monthly basis. Monthly summary tables are prepared. Fishery data includes landing data, fishing effort, in the form of number of boat, number of hauls and number of hours. Catch, effort and the ratio of catch per unit effort give the index of abundance of stock. In biological point of view effort as a measure of mortality caused by fishing effort as a measure of the abundance of diversity of fish stock. Sampling for a single rapidly determined measure (total length) for large number of species from the field.

Raising factor

The use of raising factors is very vital step in combining and analysing sample data. Unless the sampling system used is very simple, the length distribution of the fish measured will be difficult from the estimated distribution of the population being sampled. When several samples have been taken from a population, each giving an estimate depending on the raising factor used on the actual distribution of the sample. In the case of fish are sorted into fairly precise size categories, the most important information about the size distribution of fish in the landing as a whole is the record of the quantities landed of each category. Here more than one set of raising factors may be used.

It is often better to sample from the landing centre where the entire sampling procedure can be controlled directly. All the other types of sampling may knowingly or unknowingly will create bias data. Assuming that the length sampling is unbiased, then the bias can only be introduced it, for fish of a particular size, fish of certain ages are more likely to appear in age sample.

L/F data

The length data have to be group into 5/10 mm class intervals. Raise the data on length-frequency distribution of a sample to the estimated total catch of the species on the date of observation from each of the centres separately. Pool the estimated length-frequency distribution of all the sampling days and then raise to the estimated total catch of the species from each of the centres. Thus the length-frequency distribution in the estimated catch of a species from each of the two selected landing centres was obtained. It is this data that formed the basic input for analysis of von Bertalanffy growth parameters and to extract population parameters using the FiSAT package.

File names

Care must be taken for naming data files as a huge amount of data will be collected. Naming of L/F, weight and sexual maturity data is illustrated below.

File name start with 2 letters of species name. It is followed by 1 number indicating the year (2000=1, 2001=2, 2003=3) by two numbers referring to the month (January=01, December=12) by two letters, the first letter to mean the kind of data (Length and weight data=L, sexual maturity stage data= S) and last letter is the state/place code (Kochi=K, Tuticorin=T, Mumbai=M).

Sardinella longiceps	SL
Thunnus albacore	TA
Liza parsia	LZ
Epinephelus tauvina	ET

Ex: L/F and w data of the *Sardinella longiceps* collected by the month of January 2000 from Kochi is SL101LK.

References

Alagraja, K. 1984. Simple method for estimation of parameters for assessing exploited fish stocks. Indian. J.Fish 31 (2): 177-208.

De Jong, J. K., 1940. A preliminary investigation of the spawning habits of some fishes of the Java Sea. Treubia, 17: 307-327.

Hickling, C. F. and Rutenberg, E., 1936. The ovary as an indicator of the spawning period in fishes. J. mar. biol. Ass. UK., 21: 311-317.

Sparre, P. and S. C.Venama. Introduction to tropical fish stock assessment .Part 1. Manual FAO Fisheries Technical Paper No.306.1, Rev.1 Rome, FAO, 376p.

Form 1- Collection of length data

Sample number		Date	
Species		Station	
Fishing vessel		Gear and mesh size	
Total catch (kg)		Sample (kg)	

Length (mm)

									Total									Total
20										50								
21										51								
22										52								
.										.								
.																		
.										.								
48										78								
49										79								

Length group	No	Weight (g)
20-24		
25-29		
30-34		
.		
.		
75-79		

Form 2 - Collection of maturity data

Maturity sample

Length class (mm)	Immature (unidentified)		Female					Male				
	1	Total	2	3	4	5	Total	2	3	4	5	Total
20-24												
25-29												
.												
.												
95-99												
100-104												
Total												
Observations:												

11 Trophodynamics and Review of Methods for Stomach Content Analysis of Fishes

P. U. Zacharia

Investigation of food and feeding of fishes has traditionally been an important field of activity in fisheries biology, but it is one in which there are great difficulties in correlating the results with the research made in the other fields (FAO, 1974). Investigations of the food of the fish cannot be considered in isolation but have to be discussed in relation to the whole marine environment, of which the fish constitute single elements.

Food chains and trophic levels

The production of organic substances (food) by photosynthesis is a process involving transformation of light energy into potential chemical energy. The transfer of this food energy from the producers through a series of consumers is called a food chain, each organism through which it is passed being a link in the chain.

Three different food chains may be recognized.

1	2	3
• The carnivore chain, where the energy is passed from smaller to larger organisms	• The parasite chain, where the energy is passed from larger to smaller organisms	• The saprophyte chain, where the energy is passed from dead organic matter to micro-organism in most cases

In reality food may be passed through parts of all three chains before it is finally decomposed into inorganic nutrients by the bacteria and fungi found at the end of every food chain. In other words, the species population within a community or ecosystem form many food chains which interconnect, anastomose or cross each other in a complex pattern, which is usually referred to as the **food web.**

Organisms which belong to the same link of the food chain as counted from the producer level are said to belong to the same trophic level. Thus the plants constitute the first trophic level, the herbivores the second, and the carnivores feeding on herbivores the third trophic level. Secondary carnivores feeding on third level carnivores belong to the fourth trophic level and so forth. However, there is a very definite limit to the number of possible links in a food chain, and consequently also to the number of trophic levels in any ecosystem. The reason for this is that only about 10 percent of the available energy is assimilated in passing from one trophic level to the next. At the top of the food chain there are usually only one or two major predators. The number of species in each trophic layer increases with approach to the first layer, giving rise to what is called a pyramid of numbers. For the major predators introduction of small amounts of pollutants into the first trophic layer can have fatal consequences because it is eventually concentrated in them.

Gross production and net production

Only a very small portion of the light energy absorbed by green plants that is transformed into food energy (gross production) because most of it is dispersed as heat. Furthermore, some of the synthesized gross production is used by the plants in their own respiratory processes, leaving a still smaller amount of potential energy (the net production) available for transfer to the next trophic level.

The loss of energy

True production of organic matter takes place only in the chlorophyll-possessing plants and certain synthetic bacteria, and this has been referred to as the primary production. Copepods and euphausids, convert plant material into protein that can be assimilated by the animals which eat them but which themselves could not exist on plant material. In reality, of course, they only assimilate and store energy derived from the primary producers. They are called secondary producers, a term which of course fits animals at higher trophic levels just as well because they too - although indirectly - utilize the primary production of the plants. The loss of energy is generally referred to as the respiratory loss because the organisms utilize the food energy by oxidizing it. Because of the respiratory losses the food chains cannot be very long and the number of trophic levels in natural communities is therefore seldom more than four or five and often only three. It also means that the total amount of food available decreases with increasing trophic level. For this reason, the largest animals are found feeding on either plants or other animals which are in a low trophic level as, for example, whales on krill and elephants on plants.

Studying food and feeding of fishes

The study of the feeding habits of fish and other animals based upon analysis of stomach content has become a standard practice (Hyslop 1980). Stomach content analysis provides important insight into fish feeding patterns and quantitative assessment of food habits is an important aspect of fisheries management. Lagler (1949) pointed out that the gut contents only indicate what the fish would feed on. Accurate description of fish diets and feeding habits also provides the basis for understanding trophic interactions in aquatic food webs. Diets of fishes represent an integration of many important ecological components that included behavior, condition, habitat use, energy intake and inter/intra specific interactions. A food habit study might be conducted to determine the most frequently consumed prey or to determine the relative importance of different food types to fish nutrition and to quantify the consumption rate of individual prey types. Each of these questions requires information on fish diets and necessitates different approaches in how one collects and analyzes data. Here, we outline qualitative and quantitative techniques used to describe food habits and feeding patterns of fishes. For a better understanding of diet data and for accurate interpretation of fish feeding patterns, time of day, sampling location, prey availability and even the type of collecting gear used need to be considered before initiating a diet study or analyzing existing diet data.

Stomach contents can be collected either from the live or fresh died fish. Regardless of the method, investigators should ensure that the removal technique effectively samples all items in the gut. Other wise data will be skewed toward items that are more easily displaced from the stomach. Alternatively, live fish can be sacrificed and stomach contents removed for analysis. If fish are to be sacrificed, they should be preserved immediately either by freezing or by fixing in formalin. Stomach contents will continue to digest, rendering rapid preservation of the fish or removed contents necessary to prevent loss of resolution. As in most fish groups feeding behavior of juveniles and adults vary distinctly attention should be taken to encounter more samples which will include all size groups of the particular fish. The specimens either from live or preserved should be measured to its total length to the nearest 1mm and weight to the nearest 0.1 g. Cut open the fish and record the sex and maturity stage of the fish. Remove the stomach and preserve them in 5% neutralized formalin for further analysis. For the analysis, a longitudinal cut must be made across the stomach and the contents are transferred into a petri dish. The contents then keep for five minutes to remove excess formalin and then examine under binocular microscope. Identify the gut content up to the genus and if possible up to species level depending up on the state of digestion. Various taxa digest at different rates. As such, all recently consumed taxa may be present in the foregut but only resistant items remain in the hindgut. To avoid bias when both easily digested prey and resistant prey are present, only the immediate foregut (e.g., stomach) should be sampled.

Prey items in fish stomachs are often not intact. Hard parts such as otoliths, scales, cleithra or backbones have diagnostic, species specific characteristics useful for identifying prey. Alternatively, partially digested prey may be identified using unique biochemical methods such as allozyme electrophoresis, or immunoassays. An important fact assessed by the examination of the stomach is the state or the intensity of feeding. This is judged by the degree of distension of the stomach or by the quantity of food that is contained in it. The distension of the stomach is judged and classified as ‘gorged or distended’, ‘full’, ‘3/4full’, ‘1/2full’ etc by eye estimation.

Fish diets can be measured in a variety of ways. Methods of gut contents analysis are broadly divisible into two, viz., qualitative and quantitative. The qualitative analysis consists of a complete identification of the organisms in the gut contents. Only with extensive experience and with the aid of good references it is possible to identify them from digested, broken and finely comminuted materials. Quantitative methods of analysis are three types, viz., numerical, gravimetric and volumetric. All these types of analysis are widely employed by different workers. The following outline of methods is based mainly on the reviews by Hynes (1950), Pillay (1952), Windell(1968), Hyslop (1980) and Chipps et al (2002).

1. Numerical methods

The numerical methods are based on the counts of constituent items in the gut contents. The numerical methods have been adapted in different ways to assess the relative importance of food items and these can be classified under four distinct heads, viz., a) Occurrence, b) Dominance, c) Number and d) Point (Numerical) methods.

a) Frequency of Occurrence: Stomach contents are examined and the individual food organisms sorted and identified. The number of stomachs in which each item occurs is recorded and expressed as a percentage of the total number of stomachs examined.

Frequency of Occurrence, = $O_i = \frac{J_i}{P}$, where, J_i is number of fish containing prey i and P is the number of fish with food in their stomach.

This method demonstrates what organisms are being fed upon, but it gives no information on quantities or numbers and doest not take in to consideration the accumulation of food organisms resistant to digestion. For instance, three organisms in a stomach, say, prawn, rotifers and diatoms, present in the ratio of 1:200:2000 would all be treated by this method as 1:1:1 with reference to the stomach in question. This method holds good even when there is differential distribution of various food organisms in the water for the same reason that it is not biased by size or numbers of organism comprising the food. Many have used this method as an indicator of inter-specific competition while some utilized this method to illustrate the seasonal changes in diet composition.

b) Number method: The number of individual of each food type in each stomach is counted and expressed as a percentage of the total number of food items in the sample studied, or as a percentage of the gut contents of each specimen examined, from which the total percentage composition is estimated.

Percent by number, $N_i = \frac{N_i}{\sum_{i1}^{Q} N_i}$, where, N_i is the number of food category i

This method has been employed successfully by several workers in studies on the food of plankton feeding fishes where the items can be counted with ease. In the basic number method, no allowance is made for the differences in size of food items. So in the studies on the food of fishes other than plankton feeders, the number method has very limited use. The counting of comminuted plant matter in the stomach of fish is impracticable and will not yield correct evaluations. So also in the analysis of the gut contents of a carnivore which may consist of only one large sized fish and a couple of small larvae, the counting are of little value computations. These are summed to give totals for each kind of food item in the whole sample, and then a grand total of all items. The quotient of these gives the percentage representation, by number, of each type of food item.

c) **Dominance method**: Essentially the dominance method is a partial improvement of the occurrence method, viz., the lack of consideration of the quantities of the food items present in the stomach, sought to be remedied. The stomach contents comprising the main bulk of the food materials present, is determined and the number of fish in which each such dominant food material is present is expressed as a percentage of the total number of fishes examined. The percentage composition of the dominant food materials can also be expressed by this method as in the occurrence method.

Though in an analysis of dominance the bulk of the food material is taken in to account, it can yield only a very rough picture of the dietary of a fish. More over, items which are less dominant due to environmental reasons may escape notice. Though this defect can also be remedied to a certain extent by the examination of large samples spread over a long period of time, a system of assay that takes in to account the relative importance of food constituents will obviously be more suitable in gut content analysis.

d) **Points (Numerical) Method:** The points method is an improvement on the numerical method where consideration is given to the bulk of the food items. The simple form of points method is the one in which the counts are computed falling a certain organisms as the unit. In a more modified form, the food items are classified as 'very common', 'common', 'frequent', 'rare', etc., based on rough counts and judgments by the eye. In this arbitrary classification the size of the individual organisms is also given due consideration. The contents of all stomachs are then tabulated and as a further approximation, different categories are allotted a certain number of points and the summations of the points for each food item are reduced to percentages to show the percentage composition of the diet. This method is essentially a numerical one; the volume being only a secondary consideration and it is only in the counts that a certain amount of accuracy can be claimed.

2. Volumetric methods

Many workers consider the volume as a more satisfactory method for quantitative analysis of gut contents. As Hynes (1950) pointed out, volume forms a very suitable means of assessment, this is especially so in the case of herbivorous and mud feeding fishes where the numerical methods "become meaningless as well as inaccurate". Even in cases where the numerical methods are suitable, volume has been considered as an essential factor to be reckoned with, and in all improved numerical methods the volume of the food items is taken in to consideration in some way or other. The chief methods that are employed in assessing the volume of food items in the gut contents of fishes are:

a) Eye estimation method: This is probably the simples and easiest means of determining the volume of food constituents. In this method the contents of each sample is considered as unity, the various items being expressed in terms of percentage by volume as estimated by inspection. This method of analysis is subjective in nature and the investigators personal bias is likely to influence the results very greatly. This defect can be minimized to a great extent by the examination of large samples conducted over a long period.

b) Points (Volumetric) method: This method is a variation of the eye estimation method. Here instead of directly assessing the volume by sight as in the previous method, each food item in the stomach is allotted a certain number of points based on its volume. Certain workers have taken into account both the size of the fish and the fullness of the stomach in the allotment of points. The diet component with highest volume was given 16 points. Every other component was awarded 16, 8, 4, 2, 1 and 0 points depending on the volume relative to the component with the highest volume. Percentage volumes within each subsample were calculated as:

$$\alpha = \frac{\text{Number of points allocated to component } \alpha}{\text{Total points allocated to subsample}} \text{ X } 100$$

where, α is the percentage volume of the prey component α

This method is quite useful for analyzing omnivorous and herbivores where measuring volumes of microscopic organisms such as diatoms and filamentous algae are very difficult.

c) Displacement method: The displacement method is probably the most accurate one for assessing the volume. The volume of each food item is measured by displacement in a graduated container such as a cylinder with the smallest possible diameter for accuracy. This method is eminently suited in the estimation of the food of carnivorous fishes. But the differential rate of digestion of the food items may sometimes affect he accuracy of the observations. However, if the collections are made when the fish are on feed, this defect can be easily overcome. A knowledge of the volumes of the different size groups of the food items ay be of great help in estimating the volume of the whole item form the semi digested fragments

3. Gravimetric method

The gravimetric method consists of the estimation of the weight of each of the food items, which is usually expressed as percentages of the weight of the total gut contents as in other quantitative methods.

$$\text{Percent by weight, } W_i = \frac{W_i}{\sum_{i=1}^{Q} W_i},$$

Where, W_i is the weight of the prey i

Generally the wet weigh of the food after removing superfluous water buy pressing it dry between filter papers is taken for this purpose. Dry weight estimation is more time consuming and is usually employed where accurate determinations of calorific intake is required. The limitation of weight as a criterion of analysis has already been referred in the consideration of the method of assessing the condition of feed. Besides these, the accurate weighing of small quantities of food matter is extremely difficult and impracticable in studies of large collections. This method is, therefore generally employed only in conjunction with other methods to demonstrate seasonal variations in the intensity of feeding.

Table: Example of results obtained using different methods of estimation of stomach contents for two *Lactarius lactarius*

L. lactarius 1 (LL1). 1. *Stolephorus bataviensis*, 9 cm long, weight 5 g, volume 7 ml, 6 *Acetes* each 3.0cm long, weight 300mg vol. 2ml, 1 *Bregmaceros* ,4cm, 1 g, vol. 1 ml.

L. lactarius 2 (LL2). 1. *Stolephours bataviensis*, 7 cm long, weight 3 g, volume 4 ml, 4 *Acetes* 2.5 cm long, weight 250 mg, vol.1 ml.

Food	Method	Fish LL1	LL2		%	Total of which % expressed
S. bataviensis	Occurrence	1	1	2	40	All food occurrences
Acetes		1	1	2	40	
Bregmaceros		1	0	1	20	
S. bataviensis	Numerical	1	1	2	15.4	All food organisms
Acetes		6	4	10	76.9	
Bregmaceros		1	0	1	7.7	
S. bataviensis	Dominance	1	1	2	100	All fish
Acetes		1	1	2	100	
Bregmaceros		1	0	1	50	

Food	Method	Fish LL1	LL2		%	Total of which % expressed
S. bataviensis	Total Volume	7	4	11	73.3	Total food volume
Acetes		2	1	3	20	
Bregmaceros		1	0	1	6.7	
S. bataviensis	% volume	70	80	75	75	Food volume
Acetes		20	20	20	20	
Bregmaceros		10	0	5	5	
S. bataviensis	Gravimetric	5	3	8	67.8	Total weight of food
Acetes		1.8	1	2.8	23.7	
Bregmaceros		1	0	1	8.5	

Food analysis indices

A. Simple indices

1) **Index of fullness**: This is measured as the ratio of food weight to body weight as an index of fullness, which is very widely employed. (The ratio of corresponding volume can also be used.) This index can be applied to the food in the stomach, or to that in the whole digestive tract. It is usually expressed as parts per 10,000 (%00, or parts per decimile); that is:

$$\text{Fullness index} = \frac{\text{weight of the stomch contents x 10,000}}{\text{weight of fish}}$$

2) **Index of selection or forage ratio**: Most fishes have a scale of preference for the organisms in their environment, so that some are consumed in large numbers, others moderately, some not al all. A quantitative index of such differences called as the forage ratio. A study of the quantities of different organisms available to the fish is made, and also of the various items in their stomachs; then;

$$\text{Selection index} = \text{forage ratio} = \frac{s}{b}$$

where, s = percentage representation by weight, of a food organism in the stomach and b = percentage representation of the same organism in the environment. The lower limit for this index is 0; its upper limit is indefinitely large.

3) **Index of electivity:** Ivlev (1961) proposed a somewhat different quantitative measure of selection which has been widely used as mean of comparing the feeding habits of fishes and other aquatic organisms with the availability of potential food resources in natural habitats. The relationship is defined as

$$\text{Electivity index} = E = \frac{s - b}{s + b}$$

The index has a possible range of -1 to +1, with negative values indicating avoidance or inaccessibility of the prey item, zero indicating random selection form the environment, and positive values indicating active selection.

B. Compound indices

In an attempt to consolidate the desirable properties of individual diet measures (e.g., Ni, Wi. Foi), compound indices were developed that combine two or more measures into a single index. The belief is that compound indices capture more information than do single component measures (Chipps *et al* 2002).

1) Index of Preponderance: (Natarajan and Jhingran, 1961)

This index gives a summary picture of frequency of occurrence as well as bulk of various food items. It provides a definite and measurable basis of grading the various food elements. The bulk of food items can be evaluated by 1) Numerical 2) volumetric and 3) Gravimetric methods. As the numerical method is not suited to the index with the frequency of occurrence it magnifies the importance of smaller organisms which may appear in enormous numbers. Therefore either volumetric or gravimetric are best to assess the food items quantitatively. If we V_i and O_i are the volume and occurrence index of food item i. then,

$$\text{Index of preponderance } I_i = \frac{V_i O_i}{\sum V_i O_i} X\ 100$$

Example: The 'Index of Preponderance' of food items of *Catla catla* (Ham.) is given in the table 2 with rankings in brackets.

Table 2 : Index of Preponderance (Natarajan and Jhingran, 1961) of adult Catla

Food items	Percentage of Occurrence (O_i)	Percentage of volume (V_i)	$V_i O_i$	$\frac{V_i O_i}{\sum V_i O_i}$ X 100
Crustaceans	24.5	57.1	1398.95	64.50 (1)
Algae	27.3	24.0	655.20	30.06 (2)
Plants	6.4	8.2	52.48	2.41 (3)
Rotifers	10.8	2.4	25.92	1.19 (4)
Insects	3.6	6.0	21.60	0.99 (5)
Protozoa	0.6	0.3	0.18	0.01 (8)
Molluscs	….	….	……	……
Polyzoa	….	….	……	. ….
Detritus	10.0	1.3	13.00	0.60 (6)
Sand and mud	16.8	0.7	11.76	0.54 (7)
Total	100	100	2179.09	100

According to the index crustacea and algae constitute 1 and 2 ranks in Catla catla. While third, fourth and fifth places are held by plants, rotifers and insects. In grading the food elements accidental and incidental inclusions like sand, mud, etc., may be left out of consideration.

2) Index of Relative Importance (IRI):- Leo Pinkas et al (1971)

This index is an integration of measurement of number, volume and frequency of occurrence to assist in evaluating the relationship of the various food items found in the stomach. It is calculated by summing the numerical and volumetric percentages values and multiplying with frequency of occurrence percentage value.;

$$\text{Index of relative importance, } IRI_i = (\%N_i + \%V_i)\,\%O_i,$$

where, N_i, V_i and O_i represent percentages of number, volume and frequency of occurrence prey **i** respectively.

Example:. Index of Relative Importance of pelagic preflexion summer flounder, *Paralichthys dentatus* larvae (Grover, 1998).

Prey	$\%N_i$	$\%V_i$	$\%O_i$	$(\%N_i + \%V_i)\,\%O_i$	% IRI
Tintinnids	28.7	3.3	37.6	1203.20	19.3
Copepod nauplii	20.0	10.2	41.2	1244.24	20.0
Copepodites	16.0	61.4	30.0	2322.00	37.3
Calanoids	0.6	4.9	2.0	11.00	0.2
Cyclopoids	0.6	2.0	2.4	6.24	0.1
Copepod eggs	16.0	1.2	34.8	598.56	9.6
Bivalve larvae	12.1	14.8	28.0	753.20	12.1
Invertebrate eggs	3.7	0.9	11.6	53.36	0.9
Other	2.3	1.3	9.2	33.12	0.5

In pelagic preflexion summer (*Paralichthy dentatus)* larvae, copepodites composed the bulk of the diet (61.4% Vol, 37.3 % IRI) and formed the most important prey. Copepod nauplii, the second most important prey, composed 20.0% (N and IRI). Tintinnids, despite being the most abundantly ingested prey (28.7% N); ranked third in importance at 19.3% (IRI). Bivalve larvae and copepod eggs were the only other prey that accounted for >1% of the diet, and together they composed 21.7% (IRI).

References

Chipps S.R and E.J. Garvey 2002. *Assessment of Food Habits and Feeding Patterns,* USGS South Dakota Cooperative Fish and Wildlife Research Unit, Department of Wildlife and Fisheries Sciences, South Dakota State University, Brookings, SD 57007.

FAO, 1974. Manual of Fisheries Science Part 2 - Methods of Resource Investigation and their Application. Fish.Tech paper 115.p.255.

Grover, J.J. 1998. Feeding habits of pelagic summer flounder, *Paralichthys dentatus*, larvae in oceanic and estuarine habitats. Fish.Bull, 90 (2): 248-257.

Hynes, H.B. N. 1950. The food of the freshwater sticklebacks *(Gastrosteus aculeatus)* and *Pygosteus pungitius*) with a review of methods used in studies of the food of fishes. J. Anim. Ecol., 19: 36-58.

Hyslop, E. J. 1980. Stomach contents analysis: a review of methods and their application. *J.Fish. Biol,* 17:411-429.

Ivlev, V.S. 1961. Experimental ecology of the feeding of fishes. Yale University Press, New Haven, Conn.

Lagler, K.F. 1949. *Studies in freshwater biology*, Ann Arbor, Michigan.

Natarajan, A.V and A. C. Jhingran. 1961. 'Index of preponderance'-a method of grading the food elements in the stomach analysis of fishes. *Indian J. Fish*, 8: 54-59.

Pillay, T.V.R. 1952. A critique of the methods of study of food of fishes. *J. zool. Soc. India.,* 4: 1885-200.

Pinkas, L., M.S.Olipahnt, and I.L.K.Iverson 1971. Food habits of albacore, bluefin tuna, and bonito in Californian waters. Calif. Dep.Fish Game, *Fish. Bull,* 152: 1-105.

Seaburg, K.G. 1957. A stomach sampler for live fish. *Progre. Fish. Cult*. 19: 137- 144

Shchoener, T. W. 1970. Non synchronous spatial over lapof lizards in patchy habitats. *Ecol,* 51: 408-418.

Strauss, R. E. 1979. Reliability estimates for Ivlev's electivity index, the forage ratio, and a proposed linear index of food selection. *Transactions of the American Fisheries Society* 108:344-352.

Windell, J.T. 1968. Food analysis and rate of digestion. In. W.E. Ricker (editor), methods of assessment of fish production in fresh waters, 2nd ed., P. 215-226. IBP (Int. Biol. Programme). Handb.3.

12

Introduction to Fish Stock Assessment and Management

T. V. Sathianandan

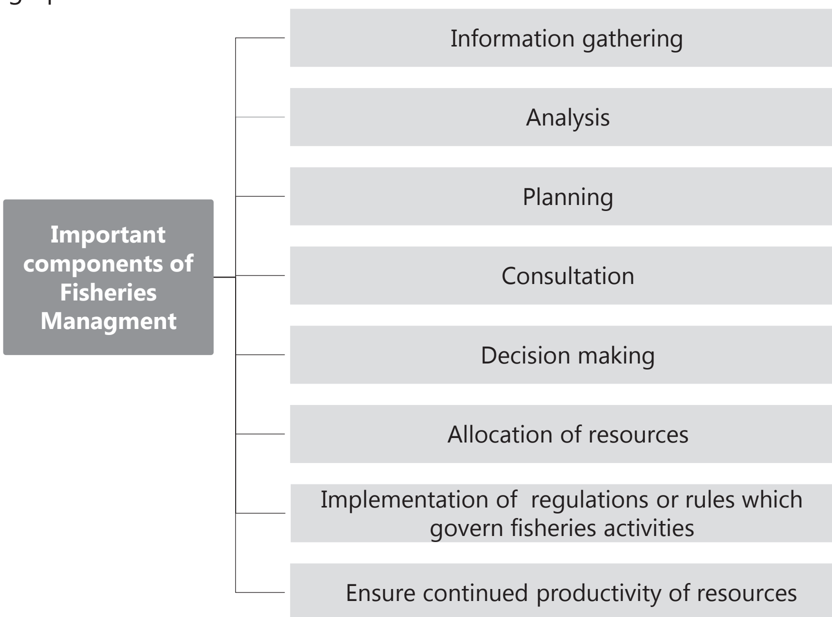
It is a known fact that when resources are owned by a large group of people, no one takes responsibility for maintaining the resource. In such situation the tendency of each person is to use the resource to the maximum extent without bothering to conserve the resources or invest in the resources because those who do not contribute also would get benefit. This is generally known as "the tragedy of the commons". To prevent "the tragedy of the commons" most common property resources are held in trust and managed for the people by the government. Fish living in public waters are a common property resource. The government has the responsibility of managing this resource for the benefit of all citizens, even those who do not fish.

Fish is a renewable resource which has to be managed using some basic principles for the benefit of the citizens generation after generation. Renewable resources like finfish and shellfish are living things that replenish themselves naturally and can be harvested, within limits, on a continuing basis without being eliminated. The scientific principles behind this renewability are well known and provide the basis for its management. As a kind of biological insurance against natural calamities all living organisms produce more offspring than that is necessary should survive to become adult. Most individual fish and shellfish produce tens of thousands to millions of eggs, most of which do not survive to become juveniles and even fewer live to become adults.

In a habitat of unfished population, the population biomass increases over time due to reproduction and growth of individual fish and it will approach a maximum beyond which the biomass can not grow which is known as the carrying capacity of the habitat. Compared to a fished population, the unfished population will have lot of older, larger fish that dominate the habitat limiting the survival of the young fish produced each year to become old fish. When fishing begins, many large older fish are removed which brings down the biomass below the carrying capacity and increases the chances of survival for smaller, younger fish. This extra production of the younger fish together with the effects of harvesting fish can result in sustainable production. The unfished population is relatively stable population with moderate production and the fished population is dynamic with a higher turnover of individual fish as the older fish are replaced by younger, faster growing fish. Some of this new production must be allowed to survive and reproduce to maintain the population. The remaining is the surplus production which is available for harvest.

Fisheries management

The basic goal of fishery management is to find out the quantity of fish that can be harvested (optimum yield) in a sustainable manner which may be subjected to changes based on political, economic, and social considerations. Highly conservative management can result in loss of fish production due to under-harvesting, while too liberal or no management can result in population depletion due to over-harvesting. Unlike mineral resources, if the fishery resources are well managed, their duration is unlimited. The fundamental basis for conservation and management of fisheries resources is its biological characteristics. Fisheries management draws on fisheries science in order to find ways to protect fishery resources so that sustainable exploitation is possible. The important components of fisheries management are depicted in the infographic.



The reality is that fisheries management is not about managing fish but about managing the associated people and industry through regulations. The success of fisheries management is associated with human factors like reactions of fishermen and the implications of the management regulations for the stakeholders. FAO of the United Nations, advises that the precautionary approach should be applied when "ecosystem resilience and human impact are difficult to forecast and hard to distinguish from natural changes. It suggests that when an action risks harm, it should not be proceeded until it can be scientifically proven to be safe.

Fisheries management history

Even 700 years ago there were restrictions on fishing imposed by Māori tribes of New Zealand, residents for about 700 years against taking more than what could be eaten. Also, as an offering to their sea god 'Tangaroa', they release back to the sea the first fish caught. Attempts started in the 18th century itself to regulate the North Norwegian fishery which resulted in forming a law in 1816 on the Lofoten fishery, known as territorial use rights. The first resource protection based fisheries management by a government was for the North European fisheries after the first over fishing conference held in London in 1936. Ray Beverton and Sidney Hold, two British fisheries researchers published a seminal work on North Sea commercial fisheries dynamics In 1957 and this became the theoretical platform for North European fisheries management schemes in 1960. In south west coast of India, the oil sardine fishery had totally collapsed during the early 1940s and the oil sardine fishery was officially closed for 5 years (1943-47) by the erstwhile British Government. Many countries have set up government departments for controlling aspects of fisheries within their Exclusive Economic Zone.

Traditional management practices aim to reduce the number of old, slow-growing fish, leaving more room and resources for younger, faster-growing fish. The assumption was that younger spawners would produce plenty of viable larvae. Research revealed that large, elderly females are far more important than younger fish in maintaining productive fisheries. The larvae produced by these older maternal fish grow faster, survive starvation better, and are much more likely to survive than the offspring of younger fish. Failure to account for the role of older fish may help explain recent collapses of some major US West Coast fisheries. One way to prevent such collapses is to establish marine reserves, where fishing is not allowed and fish populations age naturally. Some of the fisheries management mechanisms are listed in the infographic.

Fisheries Management Mechanisms

- Prohibiting devices such as bows and arrows, spears, or firearms
- Prohibiting nets
- Setting minimum mesh sizes
- Limiting the average potential catch of each vessel in the fleet
- Prohibiting bait
- Snagging
- Limits on fish traps
- Limiting the number of poles or lines per fisherman
- Restricting the number of simultaneous fishing vessels
- Limiting a vessel's average operational intensity per unit time at sea.
- Limiting average time at sea

In most countries fisheries management rules are based on the Code of Conduct for Responsible Fisheries (CCRF), agreed by different countries in an international meeting of the United Nations FAO session in 1995. In 2005 the Fisheries Centre at the University of British Columbia reviewed the performance of the world's major fishing nations against CCRF. International agreements are required to regulate fisheries in international waters. The desire for such agreement led to three conferences on the 'Law of the Sea' and to the treaty known as United Nations Convention on the Law of the Sea (UNCLOS) and derived the concept of exclusive economic zone (EEZ), extending up to 200 nautical miles (370 km) from a nation's coast and allocated certain sovereign rights and responsibilities for resource management to individual countries.

Fish Stock Assessment

Living resources are limited but renewable. The term stock in fisheries refers to genetically distinct populations within a species that are unique biological identities. Fish stock assessment can be described as the search for the exploitation level that in the long run will give maximum yield from the fishery. The aim of fish stock assessment is for a fishing strategy that gives the highest steady yield year after year. The primary objective of fish stock assessment is to provide advice on the optimum exploitation of aquatic living resources such as fish and shrimp. Social well being of the fishermen and the economic success of the fishing industry are based on the status of fish stock. Different regulatory measures imposed on the fishery are closed seasons, marine protected areas, protected species, ban on destructive fishing gears and methods, mesh size regulation, use of excluder devices etc.

In India though the fishery is largely open access type fishing is governed by certain laws and acts shown in the infographic. In all the maritime states along the west coast of India 45 days during June – August of every year is closed season except for Goa where it is for 60 days. Along the east coast 45 days during April – May of every year is closed season for fishery. Also, there is restriction on fishing by mechanized vessels near the coast in almost all the maritime states ranging from 5 – 10 km. Currently there are 31 Marine Protected Areas (MAP), mainly in Andaman & Nicobar and area under MPA is 6.16% of the area in the coastal biogeographic. Oil wells in Bombay High and Godavari Basin also function as MPAs. Under the Indian wildlife protection act, 24 species of molluscs, 10 species of elasmobranches, 5 species of turtles, 1 species

- Indian fisheries act 1897
- The wild life protection act 1972
- MFR regulation bill 1978
- MFRA of maritime states enacted from 1980
- Maritime zones of India act 1981
- Environment protection act 1986

of grouper and all species of sea horse, sea cucumber, sponges, seafans, whales, dolphins and sea cows are protected.

Catch Quotas

Systems that use individual transferable quotas (ITQ), also called individual fishing quota limit the total catch and allocate shares of that quota among the fishers who work that fishery. Fishers can buy/sell/trade shares as they choose. A large scale study in 2008 provided strong evidence that ITQ's can help to prevent fishery collapse and even restore fisheries that appear to be in decline.

Ecosystem based Fisheries

Tony Pitcher and Daniel Pauly proposed that rebuilding ecosystems, and not sustainability per se, should be the goal of fishery management. Sustainability is a deceptive goal because human harvesting of fish leads to a progressive simplification of ecosystems in favour of smaller, high turnover, lower trophic level fish species that are adapted to withstand disturbance and habitat degradation

Ecopath with Ecosim

Ecopath with Ecosim is an ecosystem modeling software suite which is an initiative by NOAA led by Jeffrey Polovina and later developed at University of British Columbia. In 2007, it was named as one of the ten biggest scientific breakthroughs in NOAA’s 200years history. This software suit revolutionized scientists ability to understand complex marine ecosystems world wide and is used widely in fisheries management as a tool for modeling and visualizing complex relationships that exist in real world marine ecosystems.

13 Analysis of Length Weight Relationship

Somy Kuriakose

Length-weight relationship (LWR) is of great importance in fishery assessments. Length and weight measurements can give information on the stock composition, life span, mortality, growth and production. Length-weight (L-W) relationships are very useful for fisheries and ecological research because they are used:

- to convert growth-in-length equations to growth-in-weight, for stock assessment models;
- for the estimation of the biomass of a species based on length frequency distributions from both onboard surveys and underwater observations;
- as an estimate of the condition of fish; and
- for between-region comparisons of life histories of a certain species.

The relationship between weight (W) and length (L) in fishes has the form:

$$W=aL^b$$

In this equation, the parameters ***a*** and ***b***, usually termed as length weight parameters are to be estimated with the available length-weight data. Each species of fish will have a specific length-weight relationship or specific length- weight parameters. It may also differ between sexes and between stocks or those belonging to different geographical regions. The parameter ***a*** is a scaling coefficient for the weight at length of the fish species. The parameter ***b*** is a shape parameter for the body form of the fish species.

The length-weight relationships were originally used for estimating the weight corresponding to a given length and to provide information on the condition of fish.

In theory, one might expect that the exponent ***b*** would have a value of roughly ***b*** = 3 because the volume of a 3-dimensional object is roughly proportional to the cube of length for a regularly shaped solid. Length is one dimensional whereas weight which depends on volume is three dimensional. Hence, there is thinking that weight of a fish is proportional to cube of the length of the fish. That is, there exists cubic relationship between weight and length of a fish.

In practice, fish that have thin elongated bodies will tend to have values of b that are less than 3 while fish that have thicker bodies will tend to have values of b that are greater than 3. Thus this also help to determine whether somatic growth is isometric (b=3) or allometric. Values of b smaller, equal and larger than 3 indicate isometry, negative allometry and positive allometry respectively. When b>3, large specimens increase in height or width faster than in length, either as the result of a change in body shape with size, or because the large specimens in the sample are in better condition than the small ones. Conversely, when b<3, either the large specimens have changed body shape, i.e., become more elongated, or the small specimens were in better nutritional condition at the time of sampling.

Thus the growth of fish length and weight is not proportionate or the relationship between length and weight is not linear. This means that when the length is increased the increase in weight is not proportionate to it. It is rather non-linear type of relationship. The estimation procedure for length – weight relationship is through linear regression. Since the above model of length-weight relationship is not linear it has to be transformed into linear type by applying logarithmic transformation. If we take logarithm *(natural logarithm with base e)* the above model will become linear as

$$\ln(W) = \ln(a) + b\ln(L) \quad or \quad Y = A + b\,X$$

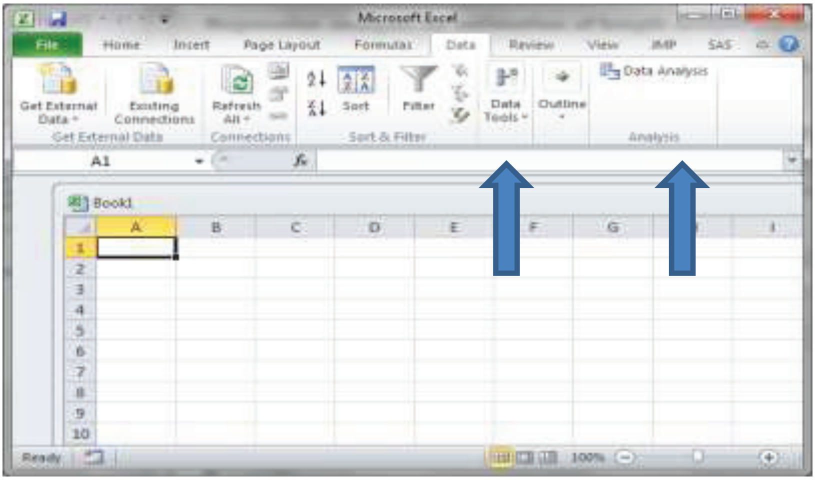
where ln*(a)* is the intercept and *(b)* the slope or regression coefficient. The above relationship is now linear and we can use the ordinary linear regression method for estimating the parameters of the relationship.

Data for fitting the length-weight relationship is collected randomly from the commercial catches and should represent fishes of all sizes, smallest to the biggest, and there should be enough samples for the analysis and estimation through regression. If our aim is examine difference in length weight relationship between different sexes then data should be collected separately for males and females.

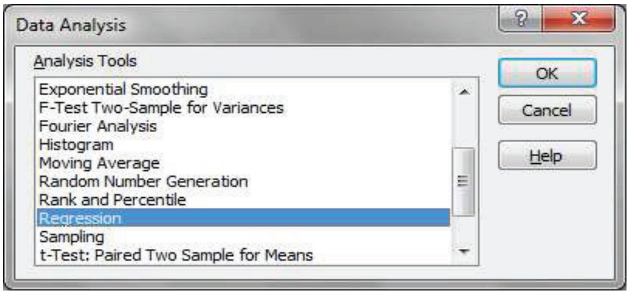
Regression analysis for estimation of length weight parameters

In Microsoft Excel we can do the analysis using the regression analysis tool. Enter the data on length and weight of samples in two columns as shown in Fig. 5. Generate two columns as the logarithmic values of the length and weight by using the natural logarithm function 'ln'. The transformed data will be used for estimation of parameters.

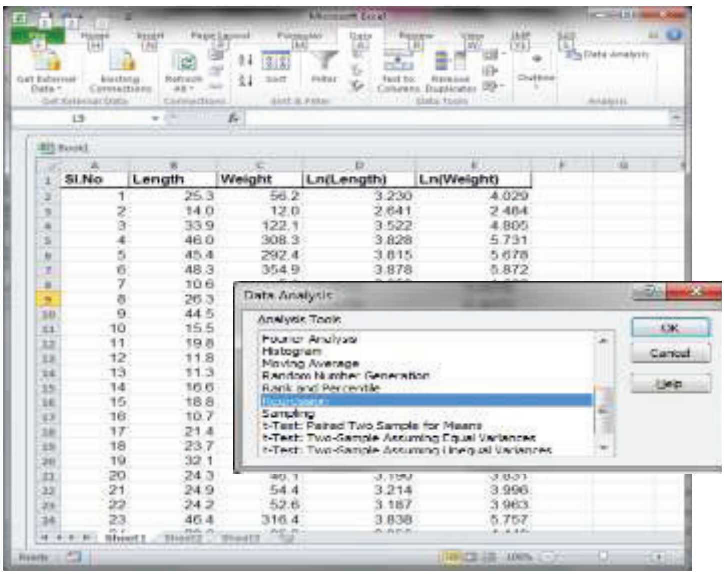
Select Data from the Main Menu, and Select Data Analysis



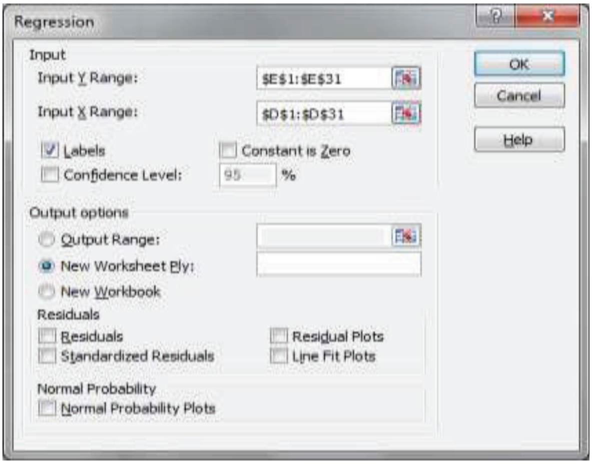
Select 'Regression' from the 'Data Analysis' dialog box and click OK.



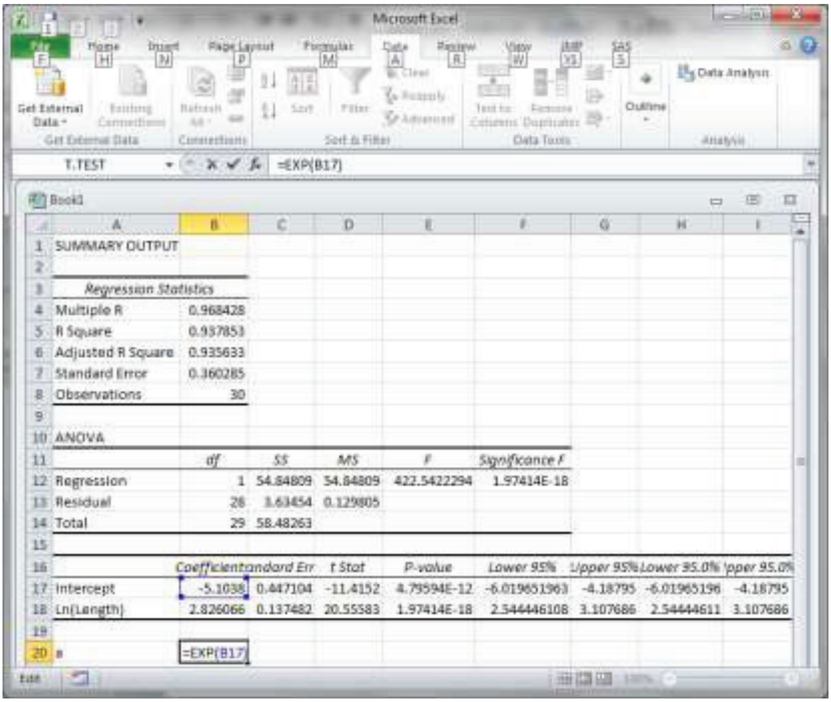
The following example demonstrates the use of this tool for estimation of length-weight parameters.



You will be presented with the dialog box. Specify the Log transformed weight data and label for "Input Y Range:" (D1:D31) and the log transformed length data and label for "Input X Range:" (E1:E31). Check the "Labels" box (since you included data labels in your input ranges), provide a new worksheet name under "Output options" and click OK.



The output will be obtained in a new sheet as given below.



The output includes regression statistics, ANOVA and the estimates of coefficients. The estimate of parameter 'a' is calculated from the value given against intercept and the estimate of parameter 'b' is that given against Ln(length) coefficient (here it is the value against 'ln(Len)' which is 2.826). The estimate of 'a' is calculated as the exponent of the intercept value which can be obtained by using the 'exp' function. For example here the intercept value is in cell B17 and to obtain the estimate of 'a' in a blank cell use the function '=exp(B17)' and we get the value of a as 0.00607.

The goodness of fit of the regression model is indicated by the 'R square' value in the output. It should be high for the relationship fitted to be good. In the example it is 0.96 indicating a good fit. The maximum value of 'R square' is 1.0 and the minimum is zero.
Using the estimated values of the parameters and the original data we can calculate the expected values of weight for the lengths in the sample data. This is done by substituting the estimated values in the relationship and calculating the weights corresponding to each length in the sample.

Statistical Test for b=3 (Isometric Relationship)

In statistical test of hypothesis this is testing for the null hypothesis $H_0 : b = 3$ against the alternative hypothesis $H_1 : b \neq 3$. The test criterion for this statistical test is a Student's t statistic with (n-2) degrees of freedom where n is the total number of observations.

Since this test criterion is for a linear regression, for the length-weight relationship situation we should use the log transformed values for the X and Y variables. Therefore, X values are the log transformed values of length and Y values are the log transformed values of the weights.

The test statistics for this is

$$t_{n-2} = \frac{(b-3)\sqrt{(n-2)\sum_{i=1}^{n}(x_i-\bar{x})^2}}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2 - b^2\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

This value has to be compared with the table value of student's t for n-2 d.f for making inferences about the null hypothesis.

If the table value of Student's t is higher than the calculated value, we accept the null hypothesis that b=3. In that case we infer that the length weight relationship is said to be isometric or there is cubic relationship between length and weight.

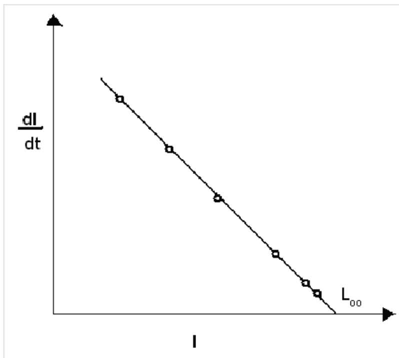
The length-weight relationship in fishes can be affected by a number of factors including season, habitat, gonad maturity, sex, diet, and stomach fullness, health and preservation techniques, and differences in the length ranges of the specimen caught. The exact relationship between length and weight differs among species of fish according to their inherited body shape, and within a species according to the condition (robustness) of individual fish. Condition sometimes reflects food availability and growth within the weeks prior to sampling. But, condition is variable and dynamic. Individual fish within the same sample vary considerably, and the average condition of each population varies seasonally and yearly. Sex and gonad development are other important variables in some species.

14

Von Bertalanffy Growth Model – Growth Parameters

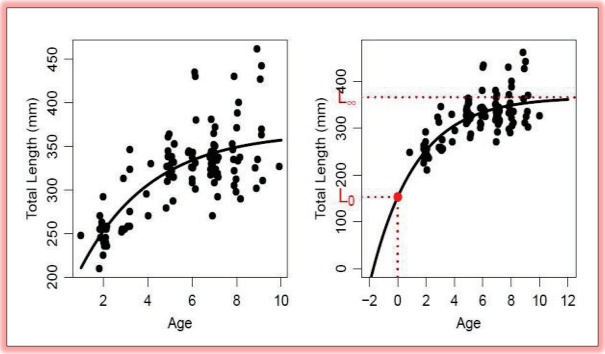
T. V. Sathianandan

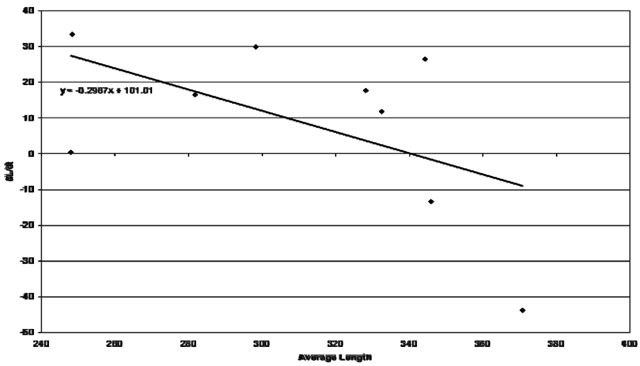
A growth curve is an empirical model of the evolution of a quantity over time. Some growth curves for certain biological systems display periods of ***exponential growth***. **A Gompertz curve,** named after Benjamin Gompertz, is a sigmoid function. It is a type of mathematical model, where growth is slowest at the start and end of a time period. In biology, a growth model is a depiction of length or weight of animals as a function of age. In the case of fish populations, the study of growth is to determine the body size as a function of its age. The growth model developed by **von Bertalanffy** (1934) has been found to be suitable for the observed growth of most of the fish species. This model expresses length as a function of age of the animal. Fish increases in length as they grow older but their growth rate which is the increment in length per unit time decreases as they grow old. When the rate of growth is plotted against the length, in most cases it will look almost like a straight line with descending limb (negative slope). This line will cut the x-axis at a point where the rate of growth is zero. This is the point beyond which the fish will not grow further and the length of the fish at this point is known as the asymptotic length denoted by L_∞.



Example: Length-at-age for a portion of a sample of male Atlantic croakers (left) and average length-at-age are given in the following table. The figures show plots of the growth curve and growth rates.

Sample	Age	Length	Avg.Age	Avg Length
1	1	248	1	248.0
2	2	210	2	248.4
3	2	225	3	281.8
4	2	236	4	208.3
5	2	240	5	328.2
6	2	245	6	345.9
7	2	255	7	332.5
8	2	258	8	344.3
9	2	263	9	370.8
10	2	270	10	327.0
11	2	292		
.	.			
.	.			





To develop the growth model the above phenomenon can be represented by means of a differential equation

$$\frac{dl}{dt} = K(L_\infty - l)$$

This can be rewritten as $\frac{dl}{L_\infty - l} = K\,dt$

The required growth model is then obtained by integrating the above differential equation to yield,

$$-\log(L_\infty - l) = K\,t + C$$

where *C* is a constant to be determined.

Expressing this equation for the length *l* we get,

$$l = L_\infty - Ce^{-K\,t}$$

When $t = 0$ the length *l* also will be zero so that we get, $0 = L_\infty - C$.

Hence $C = L_\infty$ and we get the equation as $l = L_\infty(1 - e^{-K\,t})$

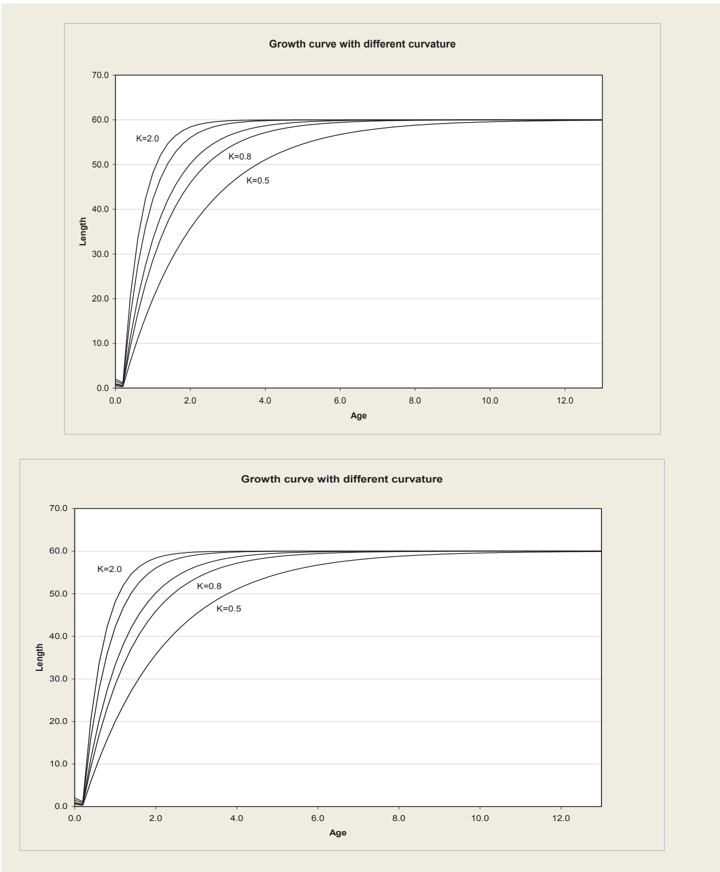
But usually the length will be zero at a different point $t = t_0$ so that we get the solution for the constant as

$$C = L_\infty e^{K t_0}$$

Hence the general growth equation is obtained by substituting the above value for C as

$$l = L_\infty (1 - e^{-K(t - t_0)})$$

Parameters of the model are *K*, L_∞ and t_0. Here *K* is termed as the curvature, L_∞ is known as the asymptotic length and t_0 is the age at birth.



15

Estimation of Growth Parameters

T. V. Sathianandan

The data commonly used for fish stock assessment is the length frequency data collected periodically by sampling from commercial catches. The data so obtained will consist of animals of different age group. The animals born on same day (single spawning) is termed as a cohort and the animals of the same age will not have same length rather it will vary with a mean and variance. If we make a histogram of their length most of the animals will fall at the middle and it will have the well known bell shape. The sample collected at a time will be a mixture of such bell shaped distributions corresponding to different age groups. If we are able to trace the length distributions of each cohort separately from its initial age up to its life span then we would be able to work out its growth and growth model parameters. As the sample collected by us from commercial catch will be a mixture of cohorts of different age groups the problem reduces to resolution of individual components (known as normal distributions or Gaussian components) from the mixture.

Resolution of Gaussian components from polymodal distributions

The frequency distribution of length obtained from a sample of fish is usually skew and polymodal. The modes corresponding to individual age groups are very useful in separating the different Gaussian components of which it is assumed to be composed off. Here the problem is to resolve a distribution into Gaussian components. Different procedures are available for resolution of a mixture into Gaussian components. These are probability paper method, parabola method and Bhattacharya's method. Among this the last method is most popular.

Probability paper method: Decomposition of polymodal frequency distributions using probability paper method was introduced by Harding (1949) and later modified by Cassie in 1950. This involves dissection of the distribution at points of inflexion of the probit plot, followed by correction for over lap of components. In this method, the cumulative percentages of the frequency distribution are first plotted against the mid points of the classes on a probability graph paper and the point of inflexion are marked. Cumulative percentages of these points are the keys for separation of the components and each segment between them are due to separate distributions. Each of these components is then extracted by adjusting the original cumulative percentages within in segments so that the total is 100. These adjusted values if plotted on the same probability paper will be linear. The means of each separated component are estimated from the actual frequencies falling in the corresponding region.

Parabola method: If the frequency distribution of random variable distributed as normal has *y* as the frequency for a class with mid value *x* then we can express y as

$$y = N \int_{x-c/2}^{x+c/2} f(x)\,dx$$

where *f(x)* is the probability distribution of a normal random variable with mean μ, standard deviation σ, *c* the class interval and *N* the total frequency. An approximation for the relation is

$$y \approx \frac{(Nce^{-(x-\mu)^2})/2\sigma^2}{\sqrt{2\pi\sigma}}$$

$$\ln(y) \approx \ln(\frac{Nc}{\sqrt{2\pi\sigma}}) - \frac{(x-\mu)^2}{2\sigma^2}$$

The above equation is of the form which is a quadratic equation representing a parabola. The axis of symmetry of the above parabola will be at $x = \mu$. Hence, if we plot the natural logarithm of the class frequencies against the mid values of the classes we can represent the different peaks with different parabolas each corresponding to a normal distribution whose mean is the point where the axis of symmetry intersects the *x*-axis.

Bhattacharya's method: If $y(x)$ denote the observed frequency of the class with *x* as its mid value and *h* the class width, then

$$y = \int_{x-h/2}^{x+h/2} \sum_{i=1}^{k} N_i f(x, \mu_i, \sigma_i)\,dx$$

$$\approx \int_{x-h/2}^{x+h/2} N_r f(x, \mu_r, \sigma_r)\,dx$$

$$\approx \int_{x-h/2}^{x+h/2} N_r \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu_r)^2}\,dx$$

$$\ln(y) = \ln(\frac{hN_r}{\sigma_r\sqrt{2\pi}}) - \frac{h^2}{24\sigma^2} - \frac{\sigma_r^2 - (h^2/12)}{2\sigma_r^2} t^2$$

by ignoring terms with higher orders in *h* and

$$t = \frac{(x-\mu_r)}{\sigma}$$

$$\Delta t^2 = 2h(x - \mu_r + \frac{h}{2}\sigma_r^2 \text{ and}$$

$$\Delta \ln(y) = -h(\sigma_r^2 - h^2/12) - (x - \mu_r + h/2)/\sigma_r^4$$

That is, the graph of $\Delta \ln(y)$ against the mid value of the class will be linear. If denote the x intercept and the angle the line makes with the negative direction of the x-axis then the mean and standard deviation of the Gaussian component corresponding to this region are estimated as a plot of $\Delta \log y(x) = \log(y(x+h)) - \log(y(x))$ against *x* is to be made first. Then the number of regions where the graph look like straight lines with negative slope, indicate the number of components (under certain conditions). By connecting the points in the regions fit straight lines for these regions. If θ_r is the angle it makes with the *x* axis and λ_r is the *x* intercept for the r^{th} region for $r = 1, \cdots, k$ then the mean and variance of the r^{th} component is estimated as

$$\mu_r = \hat{\lambda}_r + h/2$$

$$\hat{\sigma}_r^2 = (dh\, Cot\hat{\theta}_r / b) - h^2/12$$

where *b* and *d* denote the relative scales for *x* and $\Delta \log y(x)$ respectively. The proportions of the mixture can be estimated as

$$\hat{p}_i = \hat{N}_i \Big/ \sum_{i=1}^{k} \hat{N}_i$$

where $\hat{N}_i$ is the total frequency of the i^{th} class and it is estimated by

$$\hat{N}_r = \sum y(x) \big/ \sum \hat{p}_r$$

Here, the summation being restricted to the region under consideration and

$$\hat{p}_{ri} = P(\frac{x + h/2 - \hat{\mu}_i}{\hat{\sigma}_i}) - P(\frac{x - h/2 - \hat{\mu}_i}{\hat{\sigma}_i})$$

where *P(x)* is the distribution function of standard normal variate.

Estimation of Growth parameters

Once we have data on age and corresponding length obtained from the above procedure we may use any one of the following methods as per the situation to estimate the growth parameters.

Gulland and Holt Plot: For small values of Δt (need not be kept constant), the required expression is

$$\frac{\Delta L}{\Delta t} = K L_\infty - K \bar{L}_t \text{ where } \Delta L = L_{t+\Delta t} - L_t \text{ and } \bar{L}_t = \frac{(L_{t+\Delta t} + L_t)}{2}$$

By regressing $\frac{\Delta L}{\Delta t}$ on $\overline{L}_t$ (of the type $y = a + bx$) we can get estimates of the growth parameters as

$$\hat{K} = -\hat{b} \text{ and } \hat{L}_{\infty} = -\frac{\hat{a}}{\hat{b}}$$

Example: The first two columns of the following table pertain to the age and corresponding average length of animals of a cohort. The growth parameters can be estimated by calculations in the remaining columns and a followed regression. The steps followed are

1. Generate column dL as the increment in length (difference of consecutive values of L(t))
2. Generate column dt as the increment in age (difference of consecutive values of Age(t))
3. Compute values in column dL/dt as the ratio of values in dL and dt)
4. Compute the mean length Lbar(t) as the average of consecutive values of L(t)

Age (t)	L(t)	dL	dt	dL/dt	Lbar(t)
1	25.7	10.3	1	10.3	30.85
2	36.0	6.9	1	6.9	39.45
3	42.9	4.6	1	4.6	45.20
4	47.5	3.2	1	3.2	49.10
5	50.7	2.1	1	2.1	51.75
6	52.8	1.4	1	1.4	53.50
7	54.2				

Now regress the values in column dL/dt with the values in Lbar(t). That is carryout regression analysis with values in column dL/dt as Y values and values in Lbar(t) as X values and obtain the regression coefficients *a* and *b*.

Regression Statistics	
Multiple R	0.999922
R Square	0.999844
Adjusted R Square	0.999804
Standard Error	0.046844
Observations	6

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	22.36353	0.11182	199.9952	3.75E-09
Lbar(t)	-0.39163	0.00245	-159.872	9.18E-09

The estimates of coefficients in the regression model obtained through the regression analysis are a = 22.36353 and b = -0.39163 and the estimates of growth parameters are

$$\hat{K} = -\hat{b} = 0.39163 \text{ and } \hat{L}_{\infty} = -\frac{\hat{a}}{\hat{b}} = -\frac{22.36356}{-0.39163} = 57.1$$

Ford-Walford Plot: The growth equation can be brought into the form

$$L_{t+\Delta t} = a + b\,L_t \text{ where } a = L_\infty(1-b) \text{ and } b = e^{-K\,\Delta t}$$

When Δt is constant we can get estimates of a and b by regressing $L_{t+\Delta t}$ on L_t and the estimates of growth parameters can be obtained as

$$\hat{K} = -\frac{\ln(\hat{b})}{\Delta t} \text{ and } \hat{L}_\infty = \frac{\hat{a}}{(1-\hat{b})}$$

Example: For the same set of data the column L(t+1) is made with the next value of L(t). As per the Ford-Walford plot we regress the values in L(t+1) with values in L(t) and find the regression coefficients a and b.

Age (t)	L(t)	L(t+1)
1	25.7	36.0
2	36.0	42.9
3	42.9	47.5
4	47.5	50.7
5	50.7	52.8
6	52.8	54.2
7	54.2	

Regression Statistics	
Multiple R	0.999987
R Square	0.999974
Adjusted R Square	0.999968
Standard Error	0.039173
Observations	6

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	18.7018	0.074708	250.3308	1.53E-09
L(t)	0.672493	0.001713	392.5672	2.53E-10

The estimates of coefficients in the regression model obtained through the regression analysis are a = 18.7018 and b = 0.672493. Thus the estimates of growth parameters are

$$\hat{K} = -\frac{\ln(\hat{b})}{\Delta t} = -\frac{\ln(0.672493)}{1} = 0.3968 \text{ and } \hat{L}_\infty = \frac{\hat{a}}{(1-\hat{b})} = \frac{18.7018}{(1-0.672493)} = 57.1$$

Method of Chapman and Gulland: When Δt is constant, using the growth equation we can make the relation

$$L_{t+\Delta t} - L_t = c\,L_\infty - c\,L_t \text{ where } c = 1 - e^{-K\,\Delta t}$$

Through a regression of $(L_{t+\Delta t} - L_t)$ on L_t we can arrive at a regression relation of the form $y = a + b\,x$ and using the estimates of coefficients of this regression equation we can estimate the growth parameters as

$$\hat{L}_\infty = -\frac{\hat{a}}{\hat{b}} \text{ and } \hat{K} = -\frac{\ln(1+b)}{\Delta t}$$

Example: For the given data first we generate a column with values L(t+1)-L(t) and regress these values on L(t) to obtain the constants a and b in the linear regression equation.

Regression Statistics	
Multiple R	0.999945
R Square	0.999891
Adjusted R Square	0.999863
Standard Error	0.039173
Observations	6

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	18.7018	0.074708	250.33083	1.528E-09
L(t)	-0.32751	0.001713	-191.182	4.49E-09

The estimates of a and b from the regression analysis are a = 18.7018 and b = -0.32751. The corresponding estimates of growth parameters are

$$\hat{L}_{\infty} = -\frac{\hat{a}}{\hat{b}} = -\frac{18.7018}{-0.32751} = 57.1 \text{ and } \hat{K} = -\frac{\ln(1+b)}{\Delta t} = \frac{-\ln(1-0.32751)}{1} = 0.3968$$

ELEFAN – Electronic Length Frequency Analysis

The first component ELEFAN-I in the system of ELEFAN is the program for estimation of growth parameters from length frequency data. It was first developed in 1978 and it consisted of (i) component for separation of samples into normally distributed components (ii) estimation of growth parameters by generating the growth curve and minimizing the sum of squared deviations from the means of the component distributions. Later versions incorporated an algorithm which by passes the sample separation step and fits the growth curve to peaks defined independently of any assumed underlying distribution.

- Data pre-processing: ELEFAN-I uses a simple high-pass filter to identify peaks and troughs in length frequency data. The high pass filter used is a running average over 5 classes which leads to the definition of peaks as those parts of the length frequency distribution that are above the corresponding moving average and those below the corresponding running average are the thoughs separating peaks.
- Steps involved in fitting of the growth curve in ELEFAN-I are
  i. Calculate the maximum sum of points available in a set of length frequency samples. These are points which can be accumulated by one singe growth curve. It is termed as available sum of peaks (ASP).
  ii. Trace through the set of length frequency tables sequentially arranged in time for any arbitrary input of growth parameters L_{∞} and *K*. A series of growth curves

starting from the base of each of the peaks are then projected forward and backward in time to meet all other samples or the same sample again and again.

iii. Accumulate points obtained by each growth curve when passing through the troughs separating peaks.

iv. Select the curve which pass through most peaks and avoid most troughs and accumulate the largest number of points called Explained Sum of Peaks (ESP).

v. Decrement or increment the values of L_∞ and K until the ratio ESP/ASP reaches a maximum.

The growth model used in ELEFAN-I is the seasonally oscillating version of the generalized von Bertalanffy Growth Function (VBGF) of the form

$$L_t = L_\infty[1 - \exp(-KD(t - t_0) + \frac{CKD}{2\pi} Sin(2\pi(t - t_s))]^{\frac{1}{D}}$$

where

L_t is the predicted length at age t.

L_∞ is the asymptotic length

K is the growth constant – stress factor by Pauly 1981.

D is another growth constant – termed as surface factor by Pauly 1981

C is a factor that express the amplitude of the growth oscillations.

t_0 is the age at which the fish would have had zero length

t_s sets the beginning of the sinusoidal growth oscillation with respect to t = 0

In ELEFAN-I the model is used with two of the original parameters replaced (i) t_s with winter point WP and (ii) t_0 is described as a factor used to adjust a growth curve to an absolute age scale. Here a parameter "T0" is internally used to fulfil the role of t_0. Winter point WP designates the period of the year, expressed as a function of a year when growth is slowest. In northern hemisphere WP is often found to be near 0.2 (February) while for the southern hemisphere WP is often a value close to zero. The relation between WP and t_s is

$$t_s + 0.5 = WP$$

When D = 1 and C = 0 the model will take the form of the normal VBGF used for fisheries research. When 0 < C < 1 growth oscillates seasonally and when C > 1 growth oscillates strongly.

16

Age Determination in Fishes using Hard Parts

E. M. Abdussamad

Fisheries management relies on understanding the fish population dynamics while determining the biological parameters, including size at maturity, duration of spawning season, mortality estimates, age and growth, Accurate information on age of fish is an important pre-requisite for extracting precise information on growth, mortality, recruitment and other fundamental population parameters of fishes for stock assessment. The outcome of conventional age estimates using length frequency data depends upon the sample quality, selectivity of the fishing gear etc. The stock assessment results may therefore be affected and sometimes give results which is having no bearing on reality. The hard parts of the fishes also grow with the fish and the growth process may left some inscription on such parts and if that can be interpreted properly, will get precise idea on growth. These inscriptions may result from either changes in the environment which the fish inhabits, or food availability, or physiological states of the fish. However, free swimming fishes always lives in ideal conditions and do not leave any environment related markings in their skeletal structures. So interpretation of hard part inscriptions need utmost care.

Ageing techniques

There are four approaches to age the fish.

i. Direct observation of fish in confinement or marking/tagging recapture technique

This is the oldest technique described initially by the fish culturists. Tagging and marking experiments are conducted as the data collected are useful in estimating the population size, mortality rates and migration. Tagging does not enable individual fish to be aged unless the age of the fish at tagging is known. The method is very useful for fish living in areas where the growth is continuous throughout the year. It is useful when large numbers of fish recaptured at annual intervals are available. However, cultivated or tagged fish seldom have the same growth rate as that of the wild or untagged fish. Tagging or marking of fish usually involves considerable time and recapturing is not assured.

ii. Injection (chemical marker) technique

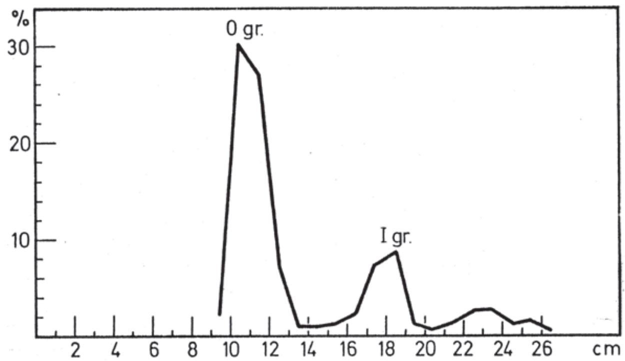
Artificial time markers can be introduced into skeletal structures by injecting chemicals into fish. The initial works were based on the use of lead acetate but this is toxic and tetracycline is now commonly used. It has the advantage of being an antibiotic drug, stable in solid form Tetracycline is readily absorbed by vertebrate animals and deposited in bony structures where calcification is taking place. In teleost fish, the tetracycline is laid down as a narrow ring timing the point of injection. The areas in which tetracycline is deposited in skeletal tissue appears fluoresce yellow under ultraviolet light, enabling them to be detected easily. However this is not a popular technique

iii. Identification of cohorts based on length frequency data of fish

Length frequency data are used in various analytical, graphical and software assisted techniques to estimate the age, growth and other population parameters. The common methods employed are:

1. Petersen method

This is a single sample method and is very simple, fastest but most inaccurate method of ageing fishes. This method can be used only with species which have a restricted spawning season so that the fish bred in a single season can be identified as a single mode in a polymodal length distribution. The mode with the lowest value is identified as O-year group fish. Subsequent modes will be 1-year group, 2-year group fish and so on. The method can be very good for young fish but becomes increasingly less useful for older fish as the growth rate slows down and the modes merge. In practice length-frequency distributions of fish caught over the shortest time period possible are plotted; the shorter the time period the more precisely the modes will be defined. A regular sequence of such length frequency distributions enables the progression of the modes to be followed.



2. Monthly modal progression analysis

Length frequency data collected at random from the commercial and experimental fishing are used to estimate the age of the age and growth of the fish.

3. Scatter diagram technique of monthly modal length

By plotting the monthly modal values of the length frequency data of fish as a scatter diagram, growth as well as the number of broods recruiting per year can be estimated.

4. Bhattacharya method

This is a graphical method of splitting a composite distribution into separate normal distributions, i.e. when several age groups or cohorts of fish are represented in the same sample. (For details consult FAO Fisheries Technical Paper No. 306.1, Rev.)

5. Probability paper/plot method

This method aims to resolve the normally distributed components of a length frequency distribution.

6. Age determination using hard parts of fish

Fishes grow continuously, but growth rate varies over time and season. These changes in growth rates may be reflected as zones or bands in the hard parts. By tracking down these inscriptions age of the fishes can be determined. Hard parts on which growth lines may found include scales, otoliths, opercular bones, spines, vertebrae etc. During the slow growth phase rings/bands will be laid close together, whereas during fast growth phase they will be laid far apart.

Among skeletal structures, otoliths and scales are most widely used as they are easy to collect and store. The opercular bones of the head, pectoral and pelvic girdles dorsal spine etc. were also widely used.

Otoliths

There are three pairs of otoliths in teleost fishes. These are three-dimensional structures but do not necessarily grow at the same rate equally in all dimensions. If there is a pattern in the otolith it will be composed of a number of concentric shells with different radii. Depending on the amount of organic material in each shell or zone, its appearance will vary from extremely opaque to completely hyaline (transparent). Fir reading otoliths it is usually preferable to identify and count the opaque zones, as characteristic growth patterns if any will usually appear and also more visible in the opaque zones. Among the three, Sagittal otoliths are generally used for age determination as they are the largest and easy to collect and process. They are located in the sacculus of the inner ear.

Scales

Scales vary in shape depending on the fish and body shape. Scales at the shoulder of the fish between the head and the dorsal fin is best suited for age determination. Scales are almost two-dimensional structures. The anterior part is formed of a series of sclerites which should extend in a regular pattern from the centre of the scale. The structural discontinuities used for age determination result from irregularities in the pattern of the sclerites; they may be slightly distorted or they may be slightly closely spaced than the majority of the sclerites; usually the discontinuities are narrow and they are usually called ‘rings’.

Scales are thin structures they need no preparation before viewing; the scales should be cleaned before they are stored. For reading, the slide with mounted scales is placed on the stage of a low-power microscope. The magnification used depends upon the size of the scale; in general, the lowest possible magnification is the best because it enables the whole scale pattern to be seen.

Validation

In this section it is assumed that some pattern of structural discontinuities (for ease of reference termed 'rings' for both otoliths and scales, except when referring specifically to one of the structures) exists in the structure which is being used for age determination and that it has been made visible by some technique. The next step is to determine whether any time-scale can be allotted to the pattern of rings. This time-scale need not be annual (Different methods are given in infographic).

a	b	c
• By observing the timing of ring formation	• By following a strong year-class through the fishery	• By using the Petersen method

Finally the age and growth estimates obtained under conventional methods can be cross checked and validated for correctness using the hard part inscription. However, detailed information on the biology and population dynamics of the fishes under question in an essential pre-requisite for any ageing work.

17

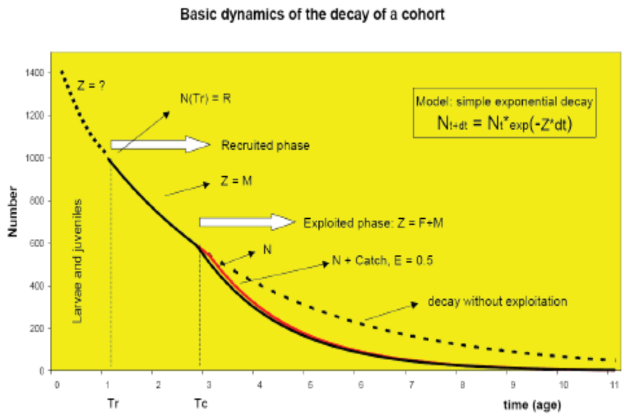
Estimation of Mortality

J. Jayasankar

Fish as a natural resource follows most of the established behavior expected of ay similar animal. The birth and ensuing recruitment, growth, reproduction and death, technically referred to as mortality are well defined phases of any animal's life time and fishes are no exceptions to this. As is evident from its logic, fish populations increase in their abundance, popularly termed as biomass, by birth of animals or by growth apart from occasional immigrations. The loss of animals is mostly through death (mortality) which could occur due to ageing, natural mortality, or due to fishing, fishing mortality apart from the predation inflicted by larger animals in the sea. Hence mortality phenomenon happens to be the single most important cause of change in abundance of fish in any defined population or technically referred to as stock. By its sheer importance as the leveling force in face of animals with varying degrees of reproduction, mortality assumes an important position in the study of dynamics or fluctuations in the biomass of a given resource. Like its growth counterparts mortality too has well laid conceptualization coupled with clearly defined procedures of measurement or technically termed as estimation. Thus the phenomenon of rate of loss of animals in a particular population is a parameter to be estimated with the sampled animals. The measurements taken from the sampled individuals help an assessor to find out the composition of fish available at various ages and such information collected over a period of time will enable the observer to find out the rates at which fish of a particular age die due to natural and unnatural causes.

The inevitability of the mortality phenomenon can be understood by the fact that for a group of contemporarily hatched fish the number can only dwindle over time. The contemporaries or those individuals who were hatched almost at the same epoch are technically termed as cohorts. The phenomenon of mortality applies to each such group of cohorts and how they decline in number through time. To clearly delineate this process of decline in numbers it is essential to follow the fate of the cohort. As mentioned earlier cohort is a batch of fish of all of approximately the same age and belonging to the same stock. (Sparre and Venema 1998). All fish of a cohort are assumed to have the same age at given time so that they all attain the recruitment age at the same time. In the context of mortality one is interested in the number of survivors from a cohort as a function of age. As mortality is split into natural and fishing induced ones, estimating the mortality entails the determination of total mortality (natural mortality + fishing mortality) first and then splitting this into natural and fishing mortalities as appropriate.

The progress of a cohort over time is displayed below in figure. In a cohort model it is assumed that R individuals are recruited into the fishery at the age t_r (denoting age at recruitment). From this age fish are exposed to some degree of natural mortality ,M. After certain time these fish are exposed to fishing at age t_c (age at first capture) denoted by F for fishing mortality. At some point tmax the older fish are not vulnerable to fishing. This setup assumes an all or none type of selection popularly referred to as knife edge selection, whereby at tc either none or all fish in an age class are either recruited or not or vulnerable or not, and once vulnerable all age classes are vulnerable. (Sparre and Venema 1992).



Dynamics of a cohort

The dynamics of similarly aged fish of a stock are assumed to follow the model of natural decay, whereby the reduction in numbers due to total mortality is an exponential function of the number of cohorts at the beginning of the period. Notationally the rate of change in numbers or number of losses or number of animals died in a small epoch is given by the following equation

$$\frac{\Delta N(t)}{\Delta t} = -Z * N(t)$$

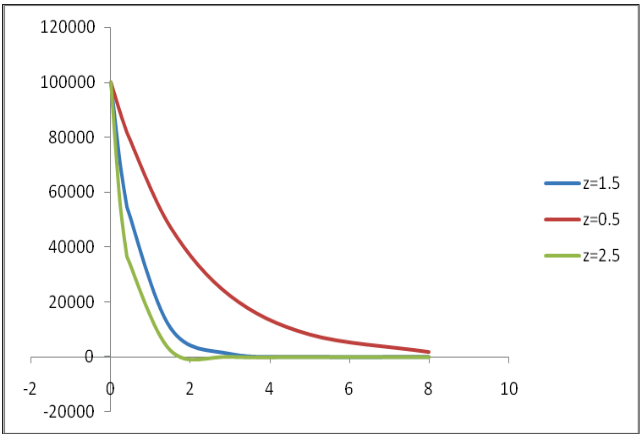
where the deltas indicate the change in numbers and a small interval of time, say one day or week etc. Z is the coefficient of reduction or popularly known as rate of annual instantaneous mortality usually scaled to account for one year. N(t) indicates the number of individuals alive at time t, preferably converted to years. This total mortality is supposed to be the arithmetic sum of natural mortality M and fishing mortality F. Notational depiction is as follows.

$$Z=M + F$$

A gentle mathematical juggling would yield the number of individuals alive at time t which follows the time of recruitment of the cohorts into the fishery at Tr could lead to an equation N(t)=N(Tr)*exp(-Z(t-Tr)). That is the number of individuals available at the present time in years is a function of the difference between the time at recruitment and the present time and also the number of live individuals at the time of recruitment.

Of course this whole relationship assumes that the rate of instantaneous mortality Z remains constant throughout. As a corollary it can be said that higher the value of Z the faster would be the decline in numbers and the lower would be the maximum age.

As an example let us look at a case where the recruitment number N(Tr) of a cohort is 1,00,000 and total instantaneous mortality Z = 1.5 per year. Assuming Δt as one day that is 1/365th of a year then, the number of survivors at different time intervals is given in the table.



As it can be seen that the loss is very severe in the initial phases as compared to the last stage like 8 years whereby the cohorts effectively vanish. The steepness depends upon the value of Z and it is better depicted in the following chart.

As is evident the decline is the steepest in the z=2.5 case and it is the slowest in z=0.5 case. The X axis indicates the time gap after Tr in years and the Y axis entries indicate the number of surviving animals.

Age of cohorts (t years)	Time interval in years	Number of survivors (N(Tr)*exp(-Z*(t-Tr)))
tr	0	100000
tr+ one day	0.00274	99590
tr+0.2	0.2	74082
tr+0.3	0.3	63763
tr+0.4	0.4	54881
tr+0.4+one day	0.40274	54656
tr+1.5	1.5	10540
tr+3.0	3	1111
tr+5.0	5	55
tr+8.0	8	1

After the animals obtain age of first capture they are most vulnerable to fishing mortality, whereas upto the age of recruitment the decline in numbers is mostly due to predation or disease ie natural mortality. Assuming that Z=F+M, the number of cohorts caught in a period from t1 to t2 in years is expressed as the function of total and fishing mortality as follows.

$$C(t1,t2)=F/Z*[(N(t1)-N(t2)]$$

wherein N(t1) and N(t2) are the animals available at time periods t1 and t2. This equation is of extreme importance in fish stock assessment and is famously referred as Baranov's equation or Catch Equation. The fraction F/Z is also very important from assessment point of view and is popularly referred to as "Exploitation Rate".

At the same period the number of animals dying due to natural causes is

D(t1,t2) = M/Z*[N(t1)-N(t2)]

The catch equation can be rewritten by involving the number of individuals at the beginning ie t1 as follows:

C(t1,t2) = N(t1)*F/Z*(1-exp(-Z(t2-t1)))

One major assumption which is the soul of this entire conceptualization is the fact that during the time interval (t1,t2) the situation at the ground is not fluctuating enough to influence the mortality rates, F and M. But criticisms are always possible on the count that natural mortality rates tend to differ with aging and younger fishes which are possibly smaller in size are less prone to fishing mortality as compared to their older counterparts.

Another conceptualization based on catch equation is the "Average number of survivors during the time period (t1,t2)" which is given by

$$\bar{N}(t1, t2) = N(t1) * \frac{1 - e^{-Z*(t2-t1))}}{Z * (t2 - t1)}$$

Estimation of total instantaneous mortality (Z)

a) From Catch Rates

There are very many ways of estimating Z from the data collected from research fishery. One such method is the method based on catch rates or Catch Per Unit Effort (CPUE) which is the ratio of total quantity of fish caught to the total number of units of gear utilized to catch the same. When the fish are caught with the same gear whose catchability coefficient (q) with respect to a particular resource is constant, the proportion of surviving members of the cohorts at two time periods (t1,t2) is equal to the ratio of the catch rates at the two time periods recorded by exploratory survey. That is

$$\frac{N(t2)}{N(t1)} = \frac{CPUE(t1)}{CPUE(t2)}$$

A slight modification of the catch equation would lead to the following relationship when the number of cohorts available at the time limits viz N(t1) and N(t2) are known.

$$Z = \frac{1}{t2 - t1} * \log(\frac{N(t1)}{N(t2)})$$

Using the previous two relationships it can be derived

$$Z = \frac{1}{t2 - t1} * \log(\frac{CPUE(t1)}{CPUE(t2)})$$

When the data available pertains to commercial fisheries, where the time series is on quarterly or annual basis, the equation used could be similar to the one described previously and the CPUE is calculated as the catch of cohort during the period (t1,t2) divided by the effort during that period. The catch rate can then be expressed as the product of average number of survivors in the period (t1,t2) and the catchability coefficient of the gear.

b) Heincke's method

Assuming that mortality rate (Z) is constant throughout the life of an individual, the following equation holds based on certain algebraic norms.

$$Z = -\ln\left(\sum_{t=1}^{\infty} N(t) / \sum_{t=0}^{\infty} N(t)\right.$$

which is called the Heincke's equation. In plain words the mortality rate is the negative value of the ratio between the number of surviving individuals from age 1 to those surviving from age 0. Substituting CPUE's at each year in the place of N(t)'s this equation assuming that they are proportional the same reads as

$$Z = -ln\frac{CPUE(1) + CPUE(2) + CPUE(3\ and\ above)}{CPUE(0) + CPUE(1) + CPUE(2) + CPUE(3\ and\ above)}$$

c) Robson- Chapman Method

Another estimate of Z is proposed by Robson and Chapman (Sparre & Venema, 1992) and the formula is

$$Z = -ln\frac{N(1) + 2 * N(2) + 3 * N(3) + \cdots}{N(0) + 2N(2) + 3N(2) + 4(N3) + \cdots - 1}$$

d) Linearised Catch Curve Method

Ideally for estimating most of the parameters including the mortality rate, the type of data required is the number of sampled and raised animals belonging to a cohort at various age categories. However in fishery sampling age determination is a time and manpower consuming exercise and invariably aging is done by using the length of the animals sampled and their categories thereof. Here length is used as an alibi for age. Further it is worth recalling that age and length are functionally linked through the Von Bertalanffy Growth Function (VBGF). Using the inversion of the VBGF length can be converted into age. The specific relationship is as follows:

$$t(L)Z = t_0 - \frac{1}{K} * \ln(1 - \frac{L}{L_\infty})$$

where t(L) is the age at length L units (cm or mm) and t_0, L_∞ and K are the classical VBGF parameters. Using this in the equation relating the logarithm of catch rate over a small time interval and the mid –time interval which is as follows:

$$\ln\left(\frac{C(t,t+\Delta t)}{\Delta t}\right) = c - Z * (t + \frac{\Delta t}{2})$$

which in turn can be rewritten using the catch and length information as

$$\ln\frac{C(L_1, L_2)}{\Delta t} = c - Z * t\left(\frac{L_1 + L_2}{2}\right).$$

Here the change in time

$$\Delta t \text{ is given by } \frac{1}{K} * \ln(\frac{L_\infty - L_1}{L_\infty - L_2})$$

Thus from this linear function, the total instantaneous rate of mortality can be estimated as the negative slope. It can be noted that in the previous equations c is a term which is made of constant terms or in other words by terms which are not involving either time or length at different classes.

To put this linearised catch curve method into action, a plot of

$$\ln\frac{C(L_1,L_2)}{\Delta t} \quad \text{against} \quad t\left(\frac{L_1+L_2}{2}\right)$$

has to be made. Only the stable range of t values which are in the fully exploited range of the animal's life and which also is not close to t_∞ (age at maximum length of the animal) must be included for the computation of the coefficients of regression. This procedure is partially subjective which must be given due care.

Example

A worked out example of estimating total instantaneous mortality rate Z from length frequency data is given below.

The case is that of Upeneus vittatus from Manila Bay, Philippines (quoted in Sparre and Venema 1992) and the length intervals and catch numbers of the pseudo cohorts is given below. The VBGF parameters are K=0.59 per year; L_∞=23.1 cm; and t_0=0

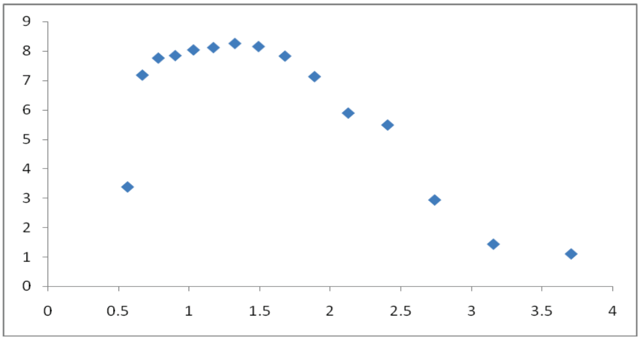
L1-L2	C(L1,L2)	t(L1)	Δt	t(L1+L2)/2	ln(C(L1,L2)/Δt)
6-7	3	0.51	0.102	0.56	3.381395
7-8	143	0.612	0.109	0.665	7.179252
8-9	271	0.721	0.116	0.778	7.756284
9-10	318	0.837	0.125	0.898	7.841493
10-11	416	0.961	0.135	1.027	8.033166
11-12	488	1.096	0.146	1.168	8.114464
12-13	614	1.242	0.16	1.32	8.252576
13-14	613	1.402	0.177	1.488	8.14997
14-15	493	1.579	0.197	1.675	7.825061
15-16	278	1.776	0.223	1.884	7.128205
16-17	93	2	0.257	2.123	5.891279

17-18	73	2.257	0.303	2.402	5.484482
18-19	7	2.56	0.37	2.735	2.940162
19-20	2	2.93	0.474	3.151	1.439695
20-21	2	3.404	0.66	3.702	1.108663
21-22	0	4.064	1.096	4.525	-
22-23	1	5.16	4.064	6.188	-1.40217
23-24	1	-	-	-	-

To select the most appropriate portion of the length intervals a plot of

$$\ln\frac{C(L_1,L_2)}{\Delta t} \text{ against } t\left(\frac{L_1+L_2}{2}\right)$$

is made and in the above case it looks like this:



As is evident from the scatter the first 7 observations and the last two observations do not follow the steady fall pattern and hence can be avoided. So only the points in the mean time range 1.5 to 3.15 are considered for estimating the regression coefficients. In this present case the estimated slope is -4.19433 and hence the estimated Z rate is 4.19.

e) The cumulated catch curve method

Another approach to estimate Z from length frequency data is the Cumulated Catch Curve method propounded by Jones and Van Zalange. The main difference here is in the time range (t1,t2) t2 is assumed to be very large to be near ∞ and that would lead the linearised catch curve equation to become

LnC(t, ∞)=d- Z*t where C(t, ∞) is called cumulated catch curve equation. Then the Jones and Van Zalinge equation for length converted catch curve would be

$$\ln(C(L, L_\infty)) = a + \frac{Z}{K}\ln(L_\infty - L)$$

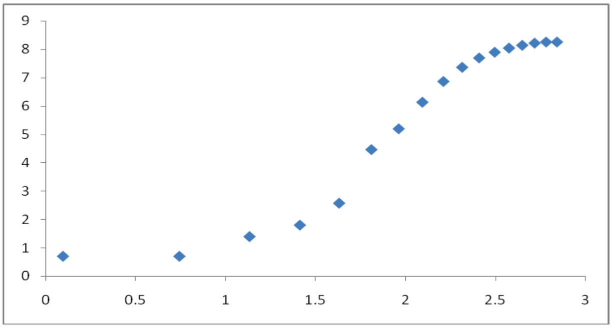
After selecting the appropriate portion of the scatter between

$$\ln\big(C(L, L_\infty)\big) \text{ and } \ln(L_\infty - L).$$

Then from the slope the Z is estimated as slope * K. For the previous example the cumulated catch curve approach is done as follows:

L1-L2	C(L1,L2)	C(L1,L∞)	ln(C(L1,L∞))	ln(L∞-L1)
6-7	3	3816	8.246958	2.839078
7-8	143	3813	8.246172	2.778819
8-9	271	3670	8.207947	2.714695
9-10	318	3399	8.131237	2.646175
10-11	416	3081	8.033009	2.572612
11-12	488	2665	7.887959	2.493205
12-13	614	2177	7.685703	2.406945
13-14	613	1563	7.354362	2.312535
14-15	493	950	6.856462	2.208274
15-16	278	457	6.124683	2.091864
16-17	93	179	5.187386	1.960095
17-18	73	86	4.454347	1.808289
18-19	7	13	2.564949	1.629241
19-20	2	6	1.791759	1.410987
20-21	2	4	1.386294	1.131402
21-22	0	2	0.693147	0.741937
22-23	1	2	0.693147	0.09531
23-24	1	1	0	-2.30259

The plot based on the last two columns is shown here. From the plot it can be seen that the most appropriate range to be used for estimating the slope is the x range from 1.41 to 2.31 and the corresponding slope value is 6.51. Hence the estimated Z rate = Slope*K= 6.51* 0.59= 3.84.



f) Beverton and Holt's Z-equation based on length data

Beverton and Holt (Sparre and Venema 1992) have shown that there exists a functional relationship between Z and the average length of fish $\bar{L}$ which is given by

$$Z = K * \frac{L_\infty - \bar{L}}{\bar{L} - L'}$$

where L′ is some length for which all fish of that length and longer are under full exploitation and it is the lower limit of the class interval of lengths from which point full exploitation is presumed.

For example if the VBGF parameters of a cohort are K=0.45 per year and $L\infty$ = 100 cm and if it is assumed that L'=45 cm then the Z estimates for the following data are as given below:

Length Group	Mid Length	N(1960)	N(1970)	N(1980)	N(1960)*Mid Length	N(1970)*Mid Length	N(1980)*Mid Length
45-50	47.5	256	268	212	12160	12730	10070
50-55	52.5	237	226	161	12442.5	11865	8452.5
55-60	57.5	211	180	116	12132.5	10350	6670
60-65	62.5	187	141	79	11687.5	8812.5	4937.5
65-70	67.5	161	105	52	10867.5	7087.5	3510
70-75	72.5	138	76	31	10005	5510	2247.5
75-80	77.5	113	50	17	8757.5	3875	1317.5
80-85	82.5	87	30	8	7177.5	2475	660
85-90	87.5	62	15	3	5425	1312.5	262.5
90-95	92.5	36	6	1	3330	555	92.5
95-100	97.5	12	1	0	1170	97.5	0
	Sum	1500	1098	680	95155	64670	38220
	$\bar{L}$				63.43667	58.898	56.20588

Where N(1960) indicates numbers caught in year 1960 and so on. The mean length here is a weighted average of the lengths detailed.

Based on these figures the Z values for various years are as follows:

$$Z(1960)=0.3 * \frac{100-63.44}{63.44-45} = 0.6 \text{ per year}$$

$$Z(1970)=0.3 * \frac{100-58.90}{58.90-45} = 0.9 \text{ per year}$$

$$Z(1980)=0.3 * \frac{100-56.21}{56.21-45} = 1.2 \text{ per year}$$

g) Power- Wetherall method

As a special application of the Beverton- Holt's Z- equation it can be expressed that

$$\bar{L}=a + b*L' \text{ where } Z/K=-(1+b)/b \text{ and } L_\infty=-a/b$$

or alternatively

$$b=-K/(Z+K) \text{ and } a=-b* L_\infty$$

This means that plotting $\bar{L}$ -L' against L' gives the estimates of a and b and from them the parameters L_∞ and Z can be estimated.

h) Pauly's empirical equation for Natural Mortality Estimation

Pauly (Sparre and Venema 1992) made regression analysis to functionally link natural mortality M with VBGF parameters and climatic parameters and the empirical formula arrived by him is given below:

Rate of Natural Mortality per Year (M)= $-0.0152-0.279*\ln L_\infty+0.6543*\ln K+0.463*\ln T$
where T is the average annual temperature at the surface in degrees centigrade.

The following table gives the estimates of natural mortality for various combinations of T and VBGF parameters.

T=5°C					T=25°C			
L∞	K=0.1	0.5	1.0	2.0	0.1	0.5	1.0	2.0
10	0.24	0.7	1.1	1.7	0.51	1.5	2.3	3.6
80	0.14	0.38	0.6	1.0	0.29	0.8	1.3	2.0
200	0.10	0.30	0.47	0.7	0.22	0.6	1.0	1.6

Method of computing

The above discussed methods of estimating rates of mortalities can be implemented practically either by manual means (highly exhausting) or by using computer based spread sheets or by software custom made for this purpose.

References

Sparre, P and Venema, S.C .1992. Introduction to Tropical Fish Stock Assessment. Part 1. Manual, FAO Fisheries Technical Paper No. 306.1, Rev. 1., Rome. 376 p.

18 Virtual Population Analysis

T .V . Sathianandan

Cohort analysis is a general terminology for analytical techniques used in the estimation of mortality and population size using catch-at-age data. Different methods are available under the class of cohort analysis namely Virtual Population Analysis by Fry, Gulland and Cohort Analysis by Pope and also multispecies VPA. In these models we trace the decline in abundance of cohorts as they age and pass through the fishery. The decline in cohort abundance is thus used to determine mortality rates. Originally the method was developed to trace the population size over the life space of a cohort and later devised to estimate fishing mortality rates, current abundance and recruitment size. The cohort analysis concept was first developed by Derzhavin (1922) with his work on sturgeon. In 1949, Fry termed Derzavin's model as Virtual Population Model and applied it to trace historical population size. In 1965, Gulland attempted calculations for estimating historical population sizes starting with the oldest age group and worked backwards with a guess of fishing mortality rate F termed as terminal F.

The basic concept behind cohort models is that the size of a cohort when it first enters the fishery, known as recruitment size, can be approximately calculated using catches of that cohort across all the period it is in the fishery. The historical population size is known as virtual population. The term 'virtual' is used to distinguish something that is merely conceptual from something that has physical reality. It is otherwise referred as cohort analysis because each cohort is analyzed and traced separately. These methods are now used to estimate cohort size, historic patterns of fishery mortality, age structure and recruitment size using data on numbers of fish harvested. The methods were initially developed for age-structured data and later extended for length based data. The basic theory behind age-structured analysis is that the abundance of fish of age (t+1) which are the survivors from the year before, equals the abundance of fish of age t reduced by fishing and natural mortalities.

Gulland's VPA model

It is based on two basic equations. The first one is the exponential decay model given by

$$N_{t+1} = N_t \, e^{-(F_t + M_t)} \qquad (1)$$

and the second equation is Baranov's catch equation given by

$$C_t = \frac{F_t}{F_t + M_t} N_t \left(1 - e^{-(F_t + M_t)}\right) \qquad (2)$$

Here N_t refers to the number of fish of age *t* in the year class, F_t and M_t are instantaneous rates of fishing and natural mortalities for fish of age *t* in the year class (constant within a year) and C_t is the catch in number of fish of age *t* in the year class. Here the assumptions are (i) fishing takes place continuously through out the year and (ii) fishing and natural mortality rates are constant within a year.

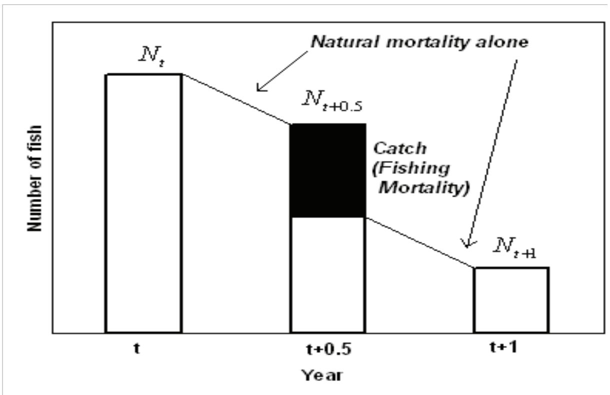
By dividing equation (1) with equation (2) we get the following expression

$$\frac{N_{t+1}}{C_t} = \frac{Z_t\, e^{-Z_t}}{F_t\left(1 - e^{-Z_t}\right)} \qquad (3)$$

Here $Z_t = F_t + M_t$ is the instantaneous rate of total mortality. Using the above three equations, starting from the terminal class, if N_{t+1}, C_t and M_t are known, we can solve for F_t from (3) and then compute N_t using (1). This procedure can be repeated to obtain F_{t-1} from (3) and then compute N_{t-1} using (1) knowing the values of N_t, C_{t-1} and M_{t-1}. Thus we can calculate these values recursively backward for all the age classes. To solve for F_t from (3) which is a non-linear equation in F_t iterative procedures have to be used and the calculations are not strait forward.

Pope's Cohort Analysis

Pope (1972) simplified Gulland's VPA model to avoid the iterative calculation procedure. In Pope's method, an year is broken into two parts with catch assumed to have occurred instantly in the middle of the year (say t+0.5). The natural mortality operates throughout the year so that by t+0.5 half of it (M/2) operated on the population size and the remaining half (M/2) is operated between t+0.5 and t+1.



Accordingly if N_t, $N_{t+0.5}$ and N_{t+1} are the population sizes at year t, t+0.5 and t+1 respectively then we have the following relation based on the exponential decay model (1).

$$N_{t+0.5} = N_t \, e^{-\left(\frac{M_t}{2}\right)} - C_t$$

$$N_{t+1} = N_{t+0.5} \, e^{-\left(\frac{M_t}{2}\right)}$$
$$= \left(N_t \, e^{-\left(\frac{M_t}{2}\right)} - C_t \right) e^{-\left(\frac{M_t}{2}\right)} \qquad (4)$$

The above equation can be rearranged to obtain an expression for backward calculation as

$$N_t = \left(N_{t+1} \, e^{\frac{M_t}{2}} + C_t \right) e^{\frac{M_t}{2}} \qquad (5)$$

Thus starting from the terminal year class if N_{t+1}, C_t and M_t are known using (5) population size N_t can be calculated backward. Once N_t is available corresponding fishing mortality rate can be calculated by rearranging (1) as

$$F_t = \ln\left(\frac{N_t}{N_{t+1}} \right) - M_t$$

By rearranging (2) we can get the expression for the population size as

$$N_t = \frac{C_t}{\left(\frac{F_t}{Z_t} \right)\left(1 - e^{-Z_t} \right)} \qquad (6)$$

If for the terminal class, the terminal $\left(\frac{F_t}{Z_t} \right)$ is known, say R (terminal F/Z), then the population size for the terminal class can be estimated as

$$N_t = \frac{C_t}{R\left(1 - e^{-\left(\frac{M}{1-R}\right)} \right)}$$

Length based VPA

In length based virtual population analysis, in the derivation of Pope's model, instead of age classes we have length classes and hence we have to consider the mid point of length classes. In place of t, t+0.5, t+1 we have L_t, $L_{t+0.5}$, L_{t+1} and when converted into age using inverse von Bertalanffy growth equation, if $dt = t_2 - t_1$ is the age difference corresponding to lengths L_t and L_{t+1} we will have the following equation corresponding to (5).

$$N_t = \left(N_{t+1} \, e^{\frac{M_t dt}{2}} + C_t \right) e^{\frac{M_t dt}{2}}$$

Different computational steps involved in length based Virtual Population Analysis is listed below with a numerical example

Step-1 For each length class with lower and upper class limits l_1 and l_2 compute corresponding ages t_1 and t_2 using the inverse VBGF given by

$$t = t_0 - \frac{\ln\left(1 - \frac{l_t}{L_\infty}\right)}{K}$$

Step-2 Corresponding to each length class calculate the age difference, $dt = t_2 - t_1$

Step-3 Assuming natural mortality rate to be constant through out the period, for each class calculate the quantity

$$H(l_1, l_2) = e^{\frac{Mdt}{2}}$$

Step-4 Starting from the terminal class we first calculate N_{last} using a reasonable guess of $\frac{F}{Z} = R$ for the terminal class (say 0.5) as

$$N_{last} = \frac{C_{last}}{R\left(1 - e^{-\left(\frac{M}{1-R}\right)}\right)}$$

Step-5 From the terminal class calculate population size back words using the recursive calculation formula

$$N_t = \left(N_{t+1}\ H(l_1,l_2) + C_t\right) H(l_1,l_2)$$

Step-6 Calculate $\frac{F_t}{Z_t}$ for the class using the equation

$$\frac{F_t}{Z_t} = \frac{C_t}{N_t - N_{t+1}}$$

Step-7 Calculate fishing mortality rate for the class as

$$F_t = \frac{M\frac{F_t}{Z_t}}{1-\frac{F_t}{Z_t}}$$

Step-8 For each class calculate total mortality rate as $Z_t = F_t + M$

Step-9 Calculate the average weight of each animal belonging to the class as

$$W_t = q\left(\frac{l_1^b + l_2^b}{2}\right)$$

Step-10 For each class calculate the mean number of animals belonging to the class as

$$N_{mean} = \frac{N_t - N_{t+1}}{Z_t}$$

Step-11 For each class calculate the Biomass as

$$B_t = W_t N_{mean}$$

Step-12 For each class calculate the Yield as

$$Y_t = W_t C_t$$

Example: Consider the following length frequency data set with necessary estimates of parameters as given below

Asymptotic length, $L_\infty = 130$

Growth rate, K = 0.1

Age at which length is zero, $t_0 = 0$

Natural Mortality coefficient, M = 0.28

Length weight parameters, q = 0.00001 & b = 3

Class		Freq.												
l_1	l_2	C_t	t_1	t_2	dt	$H(l_1,l_2)$	N_t	$\frac{F_t}{Z_t}$	F_t	Z_t	W_t	N_{mean}	B_t	Y_t
6	12	1823	0.4725	0.9685	0.4960	1.0719	98919.3	0.1255	0.0402	0.3202	0.0097	45370.0	441.0	17.7
12	18	14463	0.9685	1.4904	0.5219	1.0758	84392.7	0.5805	0.3874	0.6674	0.0378	37335.2	1411.3	546.7
18	24	25227	1.4904	2.0410	0.5506	1.0801	59475.8	0.7920	1.0661	1.3461	0.0983	23663.8	2325.7	2479.3
24	30	8134	2.0410	2.6236	0.5827	1.0850	27623.0	0.6979	0.6468	0.9268	0.2041	12575.5	2566.9	1660.3
30	36	3889	2.6236	3.2424	0.6188	1.0905	15967.8	0.6369	0.4911	0.7711	0.3683	7919.1	2916.5	1432.2
36	42	2959	3.2424	3.9020	0.6596	1.0967	9861.5	0.6785	0.5910	0.8710	0.6037	5007.0	3022.8	1786.4
42	48	1871	3.9020	4.6082	0.7062	1.1039	5500.5	0.6977	0.6462	0.9262	0.9234	2895.4	2673.6	1727.7
48	54	653	4.6082	5.3680	0.7599	1.1122	2818.8	0.5792	0.3855	0.6655	1.3403	1694.0	2270.5	875.2
54	60	322	5.3680	6.1904	0.8224	1.1220	1691.5	0.5072	0.2882	0.5682	1.8673	1117.4	2086.6	601.3
60	66	228	6.1904	7.0865	0.8961	1.1337	1056.6	0.5234	0.3075	0.5875	2.5175	741.4	1866.4	574.0
66	72	181	7.0865	8.0709	0.9844	1.1478	621.0	0.5890	0.4013	0.6813	3.3037	451.1	1490.3	598.0
72	78	96	8.0709	9.1629	1.0920	1.1652	313.7	0.5817	0.3894	0.6694	4.2390	246.5	1045.1	406.9
78	84	16	9.1629	10.3889	1.2260	1.1873	148.7	0.2823	0.1101	0.3901	5.3363	145.3	775.2	85.4
84	90	46	10.3889	11.7865	1.3976	1.2161	92	0.5000	0.2800	0.5600	6.6085	164.3	1085.7	304.0

19

Thompson and Bell Yield and Biomass Prediction Model

T .V . Sathianandan

Thompson and Bell (1934) developed a model to predict catch and stock size for a given fishing patter. The model developed by them is an age structured model. This model is used to predict the effects of changes in fishing effort on future yields. The length based version of this model usually takes necessary inputs, the number of recruits and fishing mortality rates for each length classes, from a length based cohort analysis. When recruitment details are not available, the model can still be used to get relative figures (in the form of per 1000 recruits). Other inputs necessary for prediction using Thompson and Bell model are, growth parameters, natural mortality rates for each length group, and length weight relationship parameters. It also requires an array of average weights per fish for different age/length groups to calculate yield. For economic analysis it requires a sequence of values (price) for different age groups.

The model makes predictions of catch in numbers, total number of deaths, yield, mean biomass and value for each age group. Changes in effort can be introduced by multiplying the F-array by a suitable factor and corresponding predictions can be obtained. Thompson and Bell model can be used to study the effect of certain management measures (such as closed seasons, increase in fishing effort etc.) on the yield, biomass and value. The outputs from Thompson and Bell analysis are for each length class the stock size, the catch in numbers, the yield in weight, the biomass and the time required to grow from the lower limit to the upper limit of the length group. Finally the totals of the catch, yield and mean biomass are also obtained.

The length converted Thompson and Bell analysis use the F - array (fishing mortality rates for each length class) estimated through cohort analysis as the reference F - array and assesses the effect of raising or reducing the F - array by a certain factor. The prediction made by length converted Thompson and Bell analysis is a prediction of the average long term catches assuming recruitment to remain constant.

Calculations involved in length based Thompson and Bell predictions

Inputs for the analysis consists of

- Fishing mortalities for different length groups
- Number of fish in the smallest length group (recruitment size)

- The growth parameters
- Natural mortality rate(s)
- Length-weight relationship parameters
- Average price per kg for different length groups
- The i^{th} length class is $(L_i - L_{i+1})$
- Calculate the total mortality sequence as (*x* is the multiplier used to raise or reduce the fishing mortality rates sequence, *x* = *1* for the current level of exploitation)

$$Z_i = M + x\, F_i$$

- Calculate population size for successive classes

$$H_i = \left[\frac{L_\infty - L_i}{L_\infty - L_{i+1}}\right]^{\frac{M}{2K}}$$

$$N_{i+1} = N_i \left(\frac{\frac{1}{H_i} - x\frac{F_i}{Z_i}}{H_i - x\frac{F_i}{Z_i}} \right)$$

- Calculate Catch for each class

$$C_i = [N_i - N_{i+1}]\, x \left(\frac{F_i}{Z_i}\right)$$

$$\overline{w_i} = a\left[\frac{L_i + L_{i+1}}{2}\right]^b$$

- Calculate average weight for each length class

$$Y_i = C_i\, \overline{w}_i$$

- Calculate yield for different classes

$$V_i = Y_i\, \overline{v}_i$$

- Calculate value for each class

$$\overline{N}_i = \frac{N_i - N_{i+1}}{Z_i\, \Delta t_i}$$

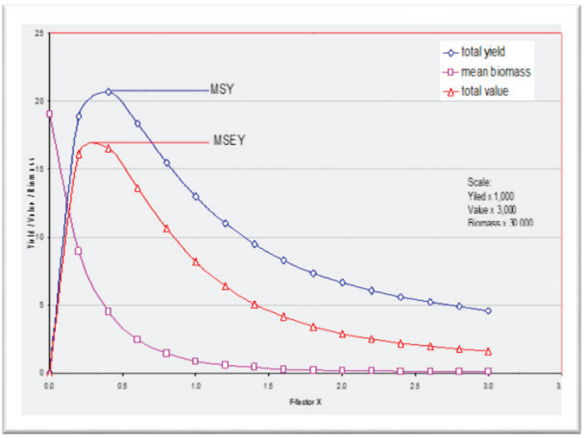
Length group	F_i	N_i	C_i	Yield	Biomass	Value
				Y_i	B_i	V_i
6-12	0.04	98919.3	1823	13.3	330.7	13.3
12-18	0.39	84392.7	14463	488.1	1260.1	488.1
18-24	1.07	59475.8	25277	2336.3	2191.5	2336.3
24-30	0.65	27623.0	8143	1601.0	2475.2	2401.5
30-36	0.49	15967.8	3889	1397.6	2845.9	2096.4
36-42	0.59	9861.5	2959	1755.2	2970.1	3510.5
42-48	0.65	5500.5	1871	1704.9	2638.4	3409.9
48-54	0.39	2818.8	653	866.2	2247.1	2165.5
54-60	0.29	1691.5	322	596.3	2069.4	1490.8
60-66	0.31	1056.6	228	570.1	1853.8	1710.3
66-72	0.40	621.0	181	594.6	1481.9	1783.8
72-78	0.39	313.7	96	405.0	1040.1	1215.0
78-84	0.11	148.7	16	85.0	772.0	255.1
84-inf	0.28	92.0	46	563.5	2012.6	1690.6
Total			59967	12977.1	26188.8	24567.1

- Length based Thompson and Bell analysis for estimation of MSY and MSEY (Table prepared with F-factor x = 2.0)

Length group	F_i	N_i	C_i	Yield	Biomass	Value
6-12	0.08	98919.3	3611.6	26.3	327.6	26.3
12-18	0.77	82724.1	26041.2	878.9	1134.4	878.9
18-24	2.13	47271.6	32863.1	3043.4	1427.4	3043.4
24-30	1.29	10092.9	5154.2	1014.5	784.2	1521.7
30-36	0.98	3823.1	1652.3	593.8	604.6	890.7
36-42	1.18	1669.7	881.6	523.0	442.5	1046.0
42-48	1.29	609.2	351.7	320.5	248.0	640.9
48-54	0.77	181.3	74.9	99.3	128.8	248.3
54-60	0.58	79.3	27.4	50.8	88.1	126.9
60-66	0.62	38.5	14.9	37.3	60.6	111.9
66-72	0.80	16.8	8.5	27.9	34.7	83.6
72-78	0.78	5.4	2.8	11.9	15.3	35.8
78-84	0.22	1.5	0.3	1.7	7.5	5.0
84-inf	0.56	0.8	0.5	6.7	12.0	20.2
Total			59967	12977.1	26188.8	24567.1

- Length based Thompson and Bell analysis for estimation of MSY and MSEY (Total yield and mean biomass worked out with different F-factor)

x	total yield	mean biomass	total value
0.0	0	571297	0
0.2	18903	268193	48329
0.4	20717	135343	49701
0.6	18360	73209	40925
0.8	15474	42376	31836
1.0	12977	26189	24567
1.2	10999	17216	19168
1.4	9470	11976	15236
1.6	8287	8761	12370
1.8	7365	6697	10259
2.0	6636	5316	8680
2.2	6053	4357	7480
2.4	5580	3670	6554
2.6	5191	3163	5829
2.8	4868	2780	5253
3.0	4596	2484	4790



<u>Final Results</u>

MSY = 20919
MSEY = 51544
Biomass at MSY = 163296
Biomass at MSEY = 188207
F-factor x at MSY = 0.343
F-factor x at MSEY = 0.301

20 Beverton and Holt's Yield Per Recruit Model

K. G. Mini

Introduction

Beverton and Holt's yield-per-recruit model is one of the commonly applied method for providing management advice for fish populations. The yield-per-recruit model is widely used mainly due to its relative simplicity and the fact that the model inputs like natural mortality, age/length at capture and weight-at-age are readily available for most species, or reasonable guesses can be made by inference with similar species.

The basic yield-per-recruit model divides the population into several growth groups to represent individual variability in growth. All animals in a growth group grow according to the same growth curve but the growth curve differs among growth groups. This approach to modelling individual variability in growth was chosen because it is simpler than attempting to develop a size-structured or an age- and size-structured population dynamics model. The extended yield-per-recruit model considers the dynamics of the population by age within each of n growth groups.

Assumptions

The yield per recruit model by Beverton and Holt (1957) is a "steady state model". A model that describes the state of the stock and yield when the fishing pattern has been the same over a long period so that all recruited fish alive are exposed to fishing is termed as a steady state model. The Beverton and Holts model makes the following assumptions.

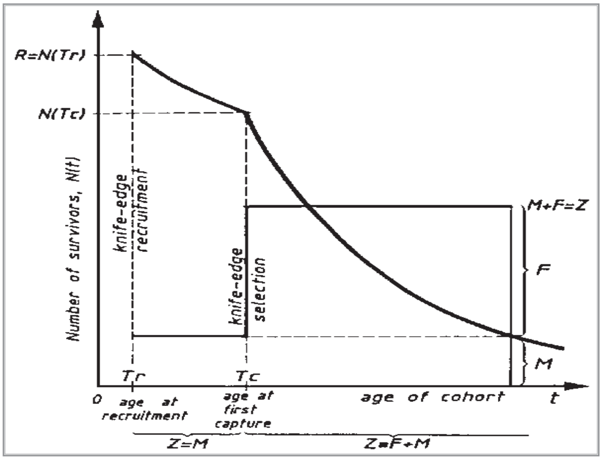
- Recruitment is constant
- All fish of a cohort are born on the same day
- Recruitment and selection are knife-edged
- The fishing and natural mortalities are constant through out the phase after recruitment
- There is a complete mixing within the stock.
- Growth in weight is isometric. That means b=3 in $W_t = a L_t^b$

Yield-per-recruit model

The assumed life history of a cohort in the Beverton and Holt model, as shown in Fig. 1 is as follows:

1) At age T_r all fish belonging to a given cohort recruit to the fishing grounds at the same time: "knife-edge recruitment".

2) From age T_r to age T_c the cohort is not exposed to any fishing mortality. In that period they suffer only from natural mortality, M, this is assumed to remain constant throughout the life span of the cohort.



3) At age T_c, the "age-at-first-capture", the cohort is assumed to be suddenly exposed to full fishing mortality, F, which is assumed to remain constant for the rest of the cohort's life. The catch from the cohort is therefore assumed to be zero until the cohort has attained the age T_c.

Fig.1 The life history of a cohort as assumed in the Beverton and Holt model

The number of survivors at age T_r is the recruitment to the fishery

$$R = N\,T_r \qquad \text{---------(1)}$$

The number of survivors at age T_c, the age at first capture is

$$N(T_c) = R\,e^{-M(T_c - T_r)} \qquad \text{---------(2)}$$

The number of survivors of the recruited cohort at age t is

$$N_t = N(T_c)\,e^{(-(M+F)(t-T_c))}$$

$$N_t = R\,e^{(-(M)(T_c-T_r)-(M+F)(t-T_c))} \qquad \text{--------(3)}$$

The fraction of the recruits surviving up to age t is given by

$$N_t/R = e^{(-(M)(T_c-T_r)-(M+F)(t-T_c))} \qquad \text{--------(4)}$$

This means that Eq. (4) gives the number of fish at time t per recruit. That means the fraction of each fish that is recruited to the fishery.

The numbers caught between a very small interval $(t, t+\Delta t)$ is given by

$$C(t, t+\Delta t) = \Delta t \, F \, N(t) \qquad \text{------- (5)}$$

The above equation gives the number of fish caught from a cohort, in the time period from t to $t+\Delta t$ when Δt is small. To obtain the corresponding yield in weight, this number should be multiplied by the individual weight of a fish. If Δt is small, then the body weight of a fish will remain approximately constant during the time period from t to $t+\Delta t$, and the yield becomes

$$Y(t, t+\Delta t) = \Delta t \, F \, N(t) w(t) \qquad \text{-------- (6)}$$

where w_t is the body weight of a t years old fish.

The yield per recruit for the time period from t to $t+\Delta t$ is divided by the number of recruits, R

$$\frac{Y(t, t+\Delta t)}{R} = \Delta t \, F \frac{N(t)}{R} w(t) \qquad \text{-------- (7)}$$

where N_t/R is defined by Eq. (4).

The above equation is the Beverton and Holt model for a short time period. To obtain the total yield per recruit for the entire life span of the cohort, Y/R all the small contributions defined by Eq.(7) has to be added up. The expression for the total yield per recruit is

$$\frac{Y}{R} = F e^{(-M(T_c - T_r))} W_\infty \left[\frac{1}{Z} - \frac{3S}{Z+K} + \frac{3S^2}{Z+2K} - \frac{S^3}{Z+3K} \right] \qquad \text{-------- (8)}$$

where:

$S = e^{-K(T_c - t_0)}$

K = von Bertalanffy growth parameter

t_0 = von Bertalanffy growth parameter

T_c = age at first capture

T_r = age at recruitment

W_∞ = asymptotic body weight

F = fishing mortality

M = natural mortality

Z = F + M, total mortality

The model allows us to calculate Y/R with varying inputs of the different parameters, such as F and T_c and then assess which effect the various input values have on the yield-per-recruit of the species under investigation. It is important to note that t_c and F are the two parameters over which the fishery managers have control. Fishing mortality F is proportional to effort and t_c is a function of gear selectivity which in turn is related to mesh size. Hence Y/R can be considered as a function of *F* and t_c, and often Y/R values are calculated for varying inputs of F and plotted for finding optimum value of F. This curve is known as yield per recruit curve and it often has a maximum that corresponds to the Maximum Sustainable Yield (MSY). This maximum changes as the value of t_c used is changed. By varying *F* and t_c simultaneously we can obtain a combination of F and t_c which gives the highest value for MSY.

However, yield-per-recruit analysis is not without its weaknesses. Prime amongst these are the assumptions that population is in steady state and that recruitment is independent of the size of the spawning stock. The last of these assumptions is only needed if the results of the analysis are used to make inferences about likely yields (rather than levels of yield-per recruit) Recent analysis of stock and recruitment data sets for many species reveal that the assumption that recruitment is independent of spawning stock size is probably invalid for many, if not most, stocks. Other weakness of the yield-per-recruit approach as it is usually applied is that it completely ignores the dynamics of movement and migration and assumes that length is related deterministically to age.

References

Sparre, P., and S. Venema. 1992. Introduction to tropical fish stock assessment. Part 1. Manual. FAO Fish. Tech. Paper 306/1 rev 1. FAO, Rome.

21

Trawl Surveys and Biomass Estimation

T. V. Sathianandan and Grinson George

Estimates of biomass and annual yield can be derived from bottom trawl surveys especially for monitoring demersal fish stocks. But the estimation of total biomass from this based on catch per unit effort estimates involves some crucial assumptions. The mean catch per unit area is an index of the stock abundance. This is on the assumption that it is proportional to the abundance. Using swept area method this index of stock abundance can be converted into an absolute measure of biomass.

Objectives of bottom trawl survey are

- Estimation of the total biomass and catch rates
- Estimation of biomass of selected species
- Collection of biological data such as length frequency data for estimation of growth and mortality parameters
- Collection of environmental data

The bottom trawl is a conical net bag with wide mouth fitted with weights on the ground rope and floats on the head rope. The net is kept open by tow otter boards which are wooden or iron structures which are towed by the warps attached forward of their centre so that they tend to diverge. These may be very long and sweep the sea bed over a wide area. They frighten the fish towards the advancing net and increases its effectiveness. The shape of the net varies depending on the kinds of fish targeted and the types of bottom. The ground rope is fitted with roller gear so that the trawl can be used on stony bottom without any damage. The tail end of the gear from which the captured fish are removed is called the codend where most of the size selection takes place. In order to obtain a representative sample of all the size ranges of the species the mesh size should be relatively small at the codend.

For estimation of stock sizes a completely randomized design or a stratified random sampling design is preferred and in most cases stratified sampling design is preferred. Strata are constructed in accordance with the density distribution of the fish so that areas with high/medium/low densities are separated. For stratification some prior information is required which is obtained in a first survey following simple random sampling design and the variability obtained is used for stratification. The distribution of hauls within strata should be random taking into account the practical difficulties. The number of hauls possible in a given period can be calculated as

Number of hauls per day = T / (t2+t3+t4)
where T is the number of hours available per day, t2 is the duration of one haul, t3 is the time used for shooting and hauling the trawl and t4 is the average time taken to cover distance between stations. It is important to standardize the length of the haul throughout the survey, since the catchability of species and sizes often depends on the duration of haul. Following are the important points to be remembered while recording data from a trawl survey.

- The objective of the survey determines the data items to be recorded. Eg. Biomass estimation, length frequency analysis, mortality estimation.
- Data items include specification of gear, haul duration, positions at start and end of haul, wire length, wing spread, bottom type, depth etc.
- Catch record should include total weight, species composition, length frequencies for selected species.
- Data should be well organized to facilitate processing.
- There should be a log summarizing the whole cruise
- There should be fishing log that provide information on vessel's position, time of start, end of haul gear rigging etc. Summery information on catch should also be recorded in the fishing log.
- Detailed information on catch in terms of length, weight, sex, maturity stage etc. for each specimen should be recorded along with length frequency distributions.

Swept Area Method

Trawl sweeps a well defined path, the area of which is the length of the path times the width of the trawl which is called the swept area. It is estimated as

$$a = D \times h \times X2 \text{, where } D = v \times t$$

Here, v is the velocity of the trawl over the ground when trawling, h is the length of the head-rope, t is the time spent for trawling and X2 is the fraction of the head-rope length, h, which is equal to the width of the path swept by the trawl and the wing spread is $h \times X2$. Different values of X2 in use are 0,4 to 0.6 for South east Asian bottom trawls, 0.5 as a compromise suggested by Pauly and 0.6 in the Caribbean suggested by Klima. Catch per unit area estimated by dividing the catch by the swept area is used for the estimation of biomass. When exact positions of the start and end of the haul are available the distance covered in nautical miles is estimated as

$$D = 60 \times \sqrt{(Lat1 - Lat2)^2 + (Lon1 - Lon2)^2 Cos^2(0.5(Lat1 + Lat2))}$$

where Lat1, Lat2 are the latitude at start and end of haul in degrees, Lon1,Lon2 are longitude at start and end of the haul in degrees. When the velocity of the vessel and its course together with direction and speed of the current are available, then the distance covered per hour is calculated as

$$D = \sqrt{VS^2 + CS^2 + 2VS \times CS \times Cos(dirV - dirC)}$$

where VS is the velocity of the vessel in knots (nautical miles per hour), CS is the velocity of current in knots, dirV is the course of vessel in degrees and dirC is the direction of current in degrees.

If cw is the catch in weight of a haul and t the time spent in hauling (in hours) the cw/t is the catch in weight per hour. If a is the swept area then a/t is the swept area per hour. Then the catch per unit of area is obtained as

$$CPUA = \frac{cw/t}{a/t} = \frac{cw}{a} \quad kg/nm^2$$

If X1 is the fraction of the biomass in the effective path swept by trawl which is actually retained in the gear and let $\overline{cw/a}$ is the mean catch per unit area of all hauls, then an estimate of the average biomass per unit area is

$$\bar{b} = \frac{\left(\overline{cw/a}\right)}{X1} \quad kg/nm^2$$

Let Anm^2 be the total area under investigation, then the estimate of total biomass for this area is obtained as

$$B = \frac{\left(\overline{cw/a}\right)A}{X1} \quad kg$$

22 Minimum Legal Size and its Estimation

T.V. Sathianandan and K. Sunilkumar Mohamed

Restrictions on size of fish that are caught are used as one of a number of measures considered for the sustainable management of fish stocks all over the world. The simple logic behind this conservation principle is to provide chance to the younger ones to grow, mature and reproduce at least once and contribute to the population before they are taken away in the catch. In many countries, there are legally implemented size (or length) limits for different species in the catch in the fishery including recreational fishing. Such size limits are arrived based on scientific research about the species especially its reproductive features. Though in most cases size restrictions are for the minimum size, there are restrictions on maximum size in some species were larger individuals contribute more to the population growth (example: Asian seabass younger ones are males and become females and spawn when they grow larger).

Advantages of implementing minimum legal size in a fishery

- Reduce juvenile fishing which results in economic loss
- Prevent growth over-fishing which is detrimental to the stock
- Maintain spawning stocks
- Ensures enough fish survive to grow and spawn
- Conservation of heavily exploited stocks
- Share catch more equally
- Reduce illegal marketing of fish
- Promote ethical and responsible fisheries
- Maximize marketing and economic benefits
- Promote aesthetic values of fish

In aquaculture the ultimate aim is to produce as many fish as possible in the shortest possible time which could be achieved through increased growth rate. an increased growth rate most probably will be accompanied by a subsequent decrease in age and size at sexual maturity. Since it is not economical to rear the species beyond sexual maturity, size at maturity is important for aquaculture also.

Information on size and reproductive behaviour of the species are necessary for a management regime to ensure that sufficient number of juveniles reach maturity and contribute to the growth of the population. An individual in a population is said to be fit when it survives to sexual maturity and contribute to the gene pool of the population and collectively, those surviving individuals determine the survival of the population. Thus it is very important to study about the reproductive biology of the fish for better understanding and management of an exploited ecosystem. As the reproductive behaviour vary highly from species to species

Some of the key measurements used for size regulation in fish include size at first maturity or size at which 50% of fish are mature (L50) and minimum size at maturity or size of the smallest mature fish. Proper estimation of these size measurements is very useful for fish stock management. Different methods have been proposed to estimate L50 and other measures of maturity size. According to a very useful study, each individual fish should be identified as reproductive or non reproductive. Although diverse methods are available for assessment of L50, most of the researchers apply some kind of logistic functions.

Thus restrictions on size of the animals that are caught is extensively used as one of the different means necessary for conservation of fish stocks. Accurate estimates of female age or length at maturity are thus critical for conservation of exploited fishery resources. Information on age and length at maturity based on histological evaluation of maturity status is therefore needed for different species. Fishery biologists prefer to conceive size at first maturity as the average size at which 50% of the individuals are mature. Size at 50% maturity (L50%) is commonly evaluated for wild populations as a biological reference point.

To estimate (L50%), a sample of organisms known to have just reached sexual maturity could be made available and their arithmetic mean size can be used as an estimator. One accepted method of estimating the size at first maturity is by sampling the mature animals from the population following a suitable sampling design. But the sample needed to obtain such a design based estimator (Sampling Design) for wild populations might be too expensive and would involve time-consuming histological procedures. With this conception, the estimator is usually not based on a sampling design but on a statistical model of the relation between body size and the number of individuals that are mature from a total number at each of many size intervals.

The most preferred model is the Logistic regression model to fit sigmoid curves to the proportion mature by length. The mathematical expression for a logistic regression model is

$$p(x) = \frac{e^{b_0+b_1x}}{1+e^{b_0+b_1x}}$$

Here p(x) is the probability that a fish is mature in a given length *x*. The parameters in the model b0 and b1 determine the shape and location of the sigmoid curve. Once estimates of the parameters of the model are available we can workout the length corresponding to any required proportion (size of the animal for which a given percentage of the animals will be mature) using the expression (except for 0 and 100%)

$$x = \frac{\ln\left(\frac{p}{1-p}\right) - \hat{b}_0}{\hat{b}_1}$$

where $\hat{b}_0$ and $\hat{b}_1$ are the estimates of the parameters in the logistic regression model.

Logistic regression model parameters can be estimated by adopting different statistical procedures. One method is through regression analysis after linearising the model by log transformation as shown below where *p* is the proportion mature having length *x* in the observed data.

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 x$$

The above method create estimation problems when the observed data have samples with proportions 0, 0.5 and 1.0 as the left hand side of the above equation become indeterminate or not defined for these cases. Some authors have suggested some adjustments in the observed data to handle this situation. A well accepted method is to use the statistically popular method of maximum likely hood which requires specific statistical softwares. Another alternative is to use Bayesian estimation for the logistic regression model which is explained here using the OpenBUGS computer software.

Bayesian methods are widely used in fisheries for stock assessment to obtain posterior probability densities of parameters of interest. Two important advantages of Bayesian inference are i) it provides estimates of posterior probability densities of unknown parameters of the model rather than the usual point estimates (ii) prior knowledge about the model parameters can be incorporated into the estimation process.

OpenBUGS is an open source version of WinBUGS, a statistical software for Bayesian analysis using Markov Chain Monte Carlo (MCMC), which is downloadable from www.openbugs.net. It is the windows version of the original DOS version BUGS (Bayesian inference Using Gibbs Sampling) software developed by MRC Biostatistics Unit, Cambridge, and Imperial College School of Medicine, London in 1989.

Practical Example

The source of data used for demonstration of Bayesian estimation using OpenBUGS is from the following publication accessed on line:

INFORMATION REPORTS NUMBER 2009-04, "Length and age at maturity of female yelloweye rockfish (*Sebastes rubberimus*) and cabezon (*Scorpaenichthys marmoratus*) from Oregon waters based on histological evaluation of maturity" by Robert W. Hannah, Matthew T. O. Blume and Josie E. Thompson, Oregon Department of Fish and Wildlife Marine Resources Program, 2040 Southeast Marine Science Drive, Newport, Oregon 97365, U.S.A

Number of female yelloweye rockfish sampled, number and proportion mature, by length (cm)

Length	Observed Number	Number Matured	Proportion	Length	Observed Number	Number Matured	Proportion
31	1	0	0.00	51	3	3	1.00
32	4	0	0.00	52	5	5	1.00
33	2	0	0.00	53	5	5	1.00
34	3	0	0.00	54	2	2	1.00
35	2	0	0.00	55	0	0	
36	4	2	0.50	56	1	1	1.00
37	5	2	0.40	57	4	4	1.00
38	4	1	0.25	58	2	2	1.00
39	4	2	0.50	59	0	0	
40	5	2	0.40	60	3	3	1.00
41	7	6	0.86	61	2	2	1.00
42	7	6	0.86	62	1	1	1.00
43	6	6	1.00	63	0	0	
44	8	7	0.88	64	1	1	1.00
45	5	5	1.00	65	2	2	1.00
46	19	19	1.00	66	0	0	
47	9	8	0.89	67	1	1	1.00
48	9	9	1.00	68	1	1	1.00
49	7	6	0.86	69	0	0	
50	3	3	1.00	70	1	1	1.00

OpenBUGS code for the logistic model

Download the OpenBUGS software (Version 3.0.3 or higher) from the website "http://www.mrc-bsu.cam.ac.uk/bugs" and install it on a computer system. Start the software and proceed with the following steps.

1. Open a new OpenBUGS page by choosing 'New' from the File Menu and copy the given code into the blank page (the portion from 'model' to the last line starting with 'list'.
2. Replace the input sample data portion (do not disturb the structure) with the original data where the x portion is for the lengths of samples, n portion is for the number of samples of each length observed and r is the numbers that are mature corresponding to each sample.
3. From the Model menu open the specification tool
4. Double click on the word "model" in the open page containing the code to select it and click on the check model button in the specification tool box. At the bottom left corner of the open page "model is syntactically correct" message should appear.

5. Double click on the word 'list' in the data portion of the open page to select it and click on the load data button in the specification tool box. At the bottom left corner of the open page "data loaded" message should appear.

6. Click on the compile button in the specification tool box. At the bottom left corner of the open page "model compiled" message should appear.

7. Double click on the word 'list' in the initialization portion of the open page (last line) and click on the load inits button in the specification tool box. At the bottom left corner of the open page "model is initialized" message should appear.

8. Now close the specification tool box.

9. Open the sample monitor tool box from the inference menu. Type the parameter names (b0, b1,b0.star,rhat) one at a time in the box against node and press the set button. Repeat it with other parameter names and close the sample monitor tool box once finished.

```
model
	{
	for( i in 1 : N ) {
		r[i] ~ dbin(p[i],n[i])
		logit(p[i]) <- b0.star + b1 * (x[i] - mean(x[]))
		rhat[i] <- n[i] * p[i]
		culmative.r[i] <- culmative(r[i], r[i])
				}
		b0 <- b0.star - b1 * mean(x[])
		b1 ~ dnorm(0.0,0.001)
		b0.star ~ dnorm(0.0,0.001)
	}
#
# Input sample data
#
list( x =
c(31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,4
8,49,50,51,52,53,54,56,57,58,60,61,62,64,65,67,68,70),
n = c(1,4,2,3,2,4,5,4,4,5,7,7,6,8,5,19,9,9,7,3,3,5,5,2,
1,4,2,3,2,1,1,2,1,1,1),
r = c(0,0,0,0,0,2,2,1,2,2,6,6,6,7,5,19,8,9,6,3,3,5,5,2,1,
4,2,3,2,1,1,2,1,1,1), N = 35)
#
# Initial values for parameters
#
list(b0.star=0, b1=0)
```

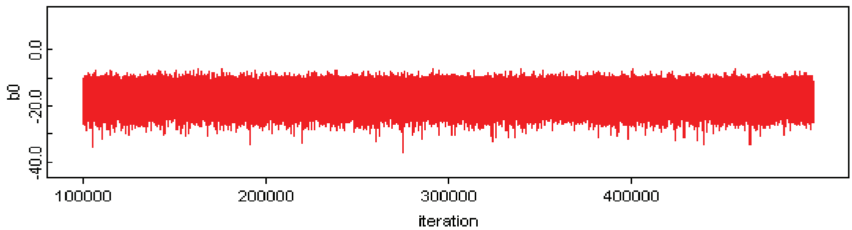
10. Open the update tool box from model menu. Replace the number in the update box with your choice number of updates (say, 100000 or more for good results) and click on the update button. The MCMC algorithm starts and the number of updates completed will be displayed in the iteration box. Close the update tool box one the iteration/updating is complete.

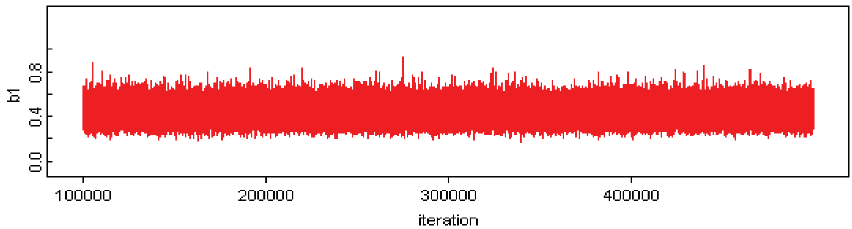
11. Open the sample monitor tool box from the inference menu again. Select the parameter name by clicking on the down arrow against the node (* for all set parameters) and click on the respective buttons to get information about the MCMC results.

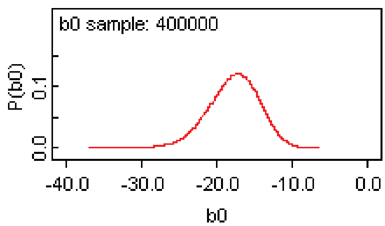
The important items are

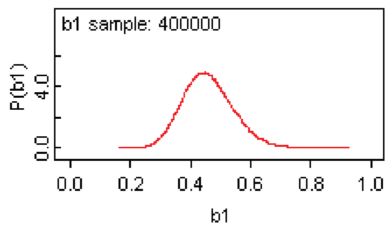
Name on the Button	Purpose
History	Graphical display of iteration history
Density	Graphical display of the probability density
Stat	Summary of statistics estimated

Here, for each parameter, the estimation history and posterior probability density plot should be examined before accepting the estimates displayed when ‘stat’ button is pressed. The history plot should be oscillating steadily in an acceptable range and the density plot should be smooth. For the sample data, the history plots, posterior probability density plots and summary statistics for the two parameters in the model obtained with 5,00,000 updations, omitting the initial 1,00,000 are given below.



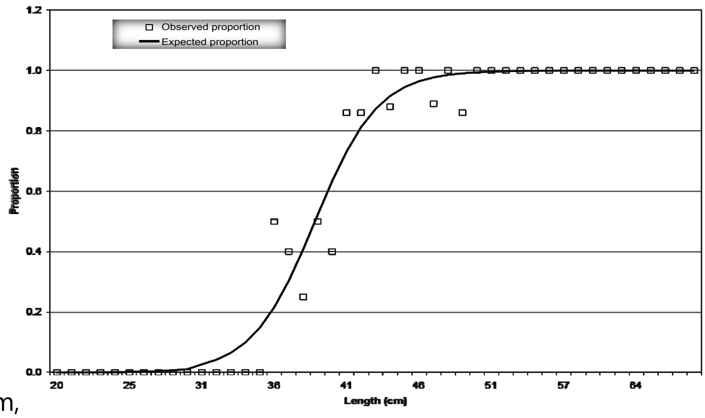






	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
b0	-17.83	3.395	0.01676	-25.06	-17.61	-11.81	100001	400000
b1	0.4595	0.08398	4.226E-4	0.3111	0.454	0.6386	100001	400000

The Bayesian estimates of the parameters of the logistic regression model for the sample data are $\hat{b}_0 = -17.83$ and $\hat{b}_1 = 0.4595$ and the plot of the observed proportions and the fitted sigmoid curve are given below. From the fitted model, the estimates of L25, L50 and L75 (lengths corresponding to 25%, 50% and 75% are mature) for the species are 36.4cm, 38.8cm and 41.2cm respectively.



23

Macro Analytical Models - Surplus Production Models

T .V . Sathianandan

Production models are classified into two major groups namely Macro/ Global/ Synthetic models and Micro Analytical Models. Macro models are based on quite simple equations where both the state of the population and fishing activity are each described by a single variable. These models take into account only the interrelationship between observable inputs such as fishing effort and observable output which is the yield obtained from the fishery. Surplus production models are Macro analytical models.

Surplus production models are an important approach to the study of harvested populations dynamics. In surplus production models the stock is considered as a single unit of biomass and modeling is not based on any age structure, length structure or dynamics of the population in terms of growth and mortalities. Instead, in these models the entire stock, the fishing effort and the total yield obtained from the stock are studied and a relationship between these are established without considering any micro level details such as growth, mortality, age at first capture, mesh size effect etc. The objective here is to obtain optimum levels of effort, which gives the maximum yield that can be sustained over a long period. These models do not demand much data for the analysis and for this reason these models more popular. When reasonable estimates are available for the yield and corresponding fishing efforts over a period of time these models can be used for obtaining optimum levels of effort and corresponding yield estimates.

Change in biomass depends on recruitment, growth and mortality. This can be represented by the following equation

$$B_{t+1} = B_t + R_t + G_t - Z_t$$

where B_t is the biomass at time *t*, R_t is the weight of the new recruits into the fishery, G_t is the total increase in the weight of the animals due to growth and Z_t is the weight of the animals died during the period. Then production is given by

$$P_t = B_{t+1} - B_t = R_t + G_t - Z_t$$

The population is in equilibrium when production is zero.
When

$P_t = 0$, population is in equilibrium

$P_t > 0$, population is in surplus

$P_t < 0$, population is in depletion

The population may collapse when P_t goes beyond some values. Here biomass is a point time concept and yield or production is a period concept.

At given time *t*, under fishing activity f_t and population state B_t, the change in B_t is assumed to depend on population state and fishing activity. Hence the equation used commonly to define surplus production models is

$$\frac{dB_t}{dt} = g(f_t, B_t)$$

Different versions of this model are given by different workers, such as

1. Pella and Tomlinson

$$\frac{dB_t}{dt} = r\, B_t \left[1 - \left(\frac{B_t}{B_0}\right)^{m-1}\right] - q\, f_t\, B_t$$

2. Graham Schaefer's model

$$\frac{dB_t}{dt} = r\, B_t \left[1 - \left(\frac{B_t}{B_0}\right)^{2}\right] - q\, f_t\, B_t$$

3. Exponential model

$$\frac{dB_t}{dt} = r\, B_t \left[1 - \ln\left(\frac{B_t}{B_0}\right)\right] - q\, f_t\, B_t$$

Here B_0, r, m and q are parameters of the model which have to be estimated using data on yield and fishing effort.

In surplus production model the rate of increase in biomass is taken as a function of biomass itself so that the relative change is given by the equation

$$\frac{1}{B_t}\frac{dB_t}{dt} = f(B_t) - F_t \quad \text{where } F_t = q\, f_t$$

and F_t is the reduction in biomass due to fishing. When the production is surplus the relative change in biomass will be positive and it will be zero when the population is in the state of equilibrium and hence $f(B_t) = F_t$ at equilibrium.

Graham-Schaefer Model: In this model the first order differential equation is used to describe the rate of change of stock biomass B_t due to production. In the absence of fishing the rate of change in the biomass is assumed to be a function of current population size only.

That is

$$\frac{dB_t}{dt} = rB_t - \frac{r}{K} B_t^2$$

where B_t is the biomass at time *t*, *K* is the carrying capacity beyond which the population can not grow and *r* is the intrinsic rate of increase in stock per unit time. When fishing mortality is added to this model it becomes,

$$\frac{dB_t}{dt} = (r - F_t)B_t - \frac{r}{K} B_t^2$$

$$= \alpha_t \ B_t - \beta \ B_t^2$$

where $\alpha_t = (r - F_t)$, $\beta = \frac{r}{K}$ and F_t is the instantaneous rate of fishing mortality.

For a short period $(t = h, t = h + \delta)$ during which the instantaneous rate of fishing mortality F_t is constant, the solution of the differential equation gives

$$B_{h+\delta} = \begin{cases} \dfrac{\alpha_h B_h e^{\alpha_h \delta}}{\alpha_h + \beta B_h e^{\alpha_h \delta - 1}} & when\, \alpha_h \neq 0 \\ \dfrac{B_h}{1 + \beta\, \delta\, B_h} & when\, \alpha_h = 0 \end{cases}$$

and yield during the same period denoted by Y_h is

$$Y_h = \int_{t=h}^{t=h+\delta} F_t \, B_t \, dt$$

and solution of this integral yields

$$Y_h = \begin{cases} \dfrac{F_h}{\beta} \ln[1 - \dfrac{\beta\, B_h (1 - e^{\alpha_h \delta})}{\alpha_h}] & when\, \alpha_h \neq 0 \\ \dfrac{F_h}{\beta} \ln[1 + \delta\, \beta\, B_h] & when\, \alpha_h = 0 \end{cases}$$

The estimated average biomass during this short period $(t = h, t = h + \delta)$ is given by

$$\overline{B}_h = \frac{Y_h}{F_h}.$$

The surplus production during this period $(t = h, t = h + \delta)$ is

$$P_h = B_{h+\delta} - B_h + Y_h$$

When yield is equal to surplus production, the population is in equilibrium.

Parameter Estimation

It is assumed that the yield Y_t at equal time periods *t=1,...,T* are available. The following notations and assumptions are made for estimation purpose.

B_t	:	Population biomass at start of time *t*
Y_t	:	Yield in biomass during time *t*
P_t	:	Surplus production during time *t*
F_t	:	Fishing mortality rate during time *t*,assumed to be proportional to fishing effort rate.
f_t	:	Fishing effort rate during time *t*
q	:	Catchability coefficient

$$F_t = q\, f_t$$
$$\alpha_t = r - F_t$$

Parameters to be estimated are *r*, *K*, *q* and the initial biomass B_1.

Algorithm for estimation

The estimation procedure is by minimizing an objective function. With some starting guess estimates of the parameters compute the initial biomass and project through time estimating the yield for each time point *t=1,...,T*. The procedure is then iterative leading to the general function minimization procedure with the function to be minimized is

$$f(r,K,q,B_1) = \sum_{t=1}^{T} [\log(Y_t) - \log(\hat{Y}_t)]^2$$

where Y_t is the actual yield and $\hat{Y}_t$ is the corresponding yield estimated according to the model. Fishing mortality can also be estimated from recorded yield using the equation

$$F_t = \begin{cases} \dfrac{\beta\, Y_t}{\ln[\dfrac{\beta\, B_t e^{\alpha_t - 1}}{\alpha_t} + 1]} & when\, \alpha_t \neq 0 \\ \dfrac{\beta Y_t}{\ln(1 + \beta B_t)} & when\, \alpha_t = 0 \end{cases}$$

Pella and Tomlinson's Model: One problem with the Graham-Schaefer model is that the maximum sustainable yield *MSY* always occurs when the biomass is half the carrying capacity *K*. This is a direct consequence of the parabolic relationship between $\frac{dB_t}{dt}$ and B_t, which in turn follows from the linear relationship between per capita productivity and population size. Pella and Tomlinson (1969) proposed an alteration to the model for which uncouples B_{MSY} from *K*.

One form of this model is given by

$$\frac{dB_t}{dt} = \begin{cases} a\,B_t^n - b\,B_t & \text{for } 0 < n \le 1 \\ b\,B_t - a\,B_t^n & \text{for } n > 1 \end{cases}$$

Simple forms

1. The simple representation of Schaefer model is

$$(Y_t / f_t) = a + b\,f_t$$

For this model the catch per unit effort is considered as a linear function of effort and the linear relationship has negative slope and positive intercept. Under this model the catch per unit effort will be maximum when

$$f_t = \frac{-a}{b}$$

The maximum sustainable yield (*MSY*) for the model is

$$MSY = \frac{-a^2}{4b}$$

and the corresponding effort is

$$f_{MSY} = \frac{-a}{2b}$$

When we have time series data on catch and effort by a linear regression of catch per unit effort (Y_t / f_t) (CPUE) on effort f_t, we can estimate the coefficients *a* and *b* and calculate MSY using this estimates.

2. In the model suggested by Fox, exponential relationship between *CPUE* and effort is assumed. The model is given by

$$Y_t / f_t = e^{c + d\,f_t} \text{ or equivalently } \ln(Y_t / f_t) = c + d\,f_t$$

This function will have maximum value for the yield when

$$f_t = \frac{-1}{d}$$

and the maximum value of yield (*MSY*) is given by

$$MSY = \frac{-1}{d} e^{c-1}$$

Using time series data on catch and effort through a linear regression of logarithm of catch per unit effort $\ln(Y_t / f_t)$ on effort f_t, we can estimate the coefficients *c* and *d* and calculate MSY using this estimates.

Example: Given in the following table are total catch and total effort in standard boat days for the shrimp fishery during the years 1969 to 1978. Estimate MSY and corresponding fishing effort (F_{MSY}) based on Schaefer model and Fox model.

Year (i)	Yield Y(i)	Effort f(i)	Y/f	ln(Y/f)
1969	546.7	1224	446.65	6.102
1970	812.4	2202	368.94	5.911
1971	2493.3	6684	373.03	5.922
1972	4358.6	12418	350.99	5.861
1973	6891.5	16019	430.21	6.064
1974	6532.0	21552	303.08	5.714
1975	4737.1	24570	192.80	5.262
1976	5567.4	29441	189.10	5.242
1977	5687.7	28575	199.04	5.294
1978	5984.0	30172	198.33	5.290
Mean	4361.1	17285.7	305.20	5.7
Sd	2184.9	10657.0	97.60	0.3

Analysis for Schaefer Model

Regression Statistics	
Multiple R	0.8797919
R Square	0.7740338
Adjusted R Square	0.745788
Standard Error	51.85595
Observations	10

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept - (a)	444.45412	31.24687	14.22396	5.81E-07
f(i) - (b)	-0.008055	0.001539	-5.23484	0.000788

$$MSY = \frac{-a^2}{4b} = \frac{-(444.45412)^2}{4*(-0.008055)} = 6130923 \ Kg = 6131 \ tonnes$$

$$F\,MSY = \frac{-a}{2b} = \frac{-(444.45412)}{2*(-0.008055)} = 27588.6 \ boat\,days$$

Analysis for Fox Model

Regression Statistics	
Multiple R	0.8847168
R Square	0.7827238
Adjusted R^2	0.7555642
Standard Error	0.1757386
Observations	10

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	6.1499573	0.105895	58.07604	8.58E-12
f(i)	-2.8E-05	5.21E-06	-5.36838	0.000671

$$MSY = \frac{-\exp(c-1)}{d} = \frac{-\exp(6.1499573 - 1)}{-0.000028} = 6159160 \ Kg = 6159 \ tonnes$$

$$F\,MSY = \frac{-1}{d} = \frac{-1}{(-0.000028)} = 35721 \ boat\,days$$

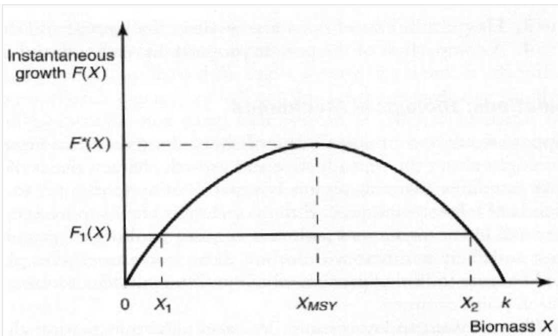
24 Estimation of Maximum Economic Yield

R. Narayanakumar

Fishery resources are renewable natural resource but are not inexhaustible extinct if the rate of harvest or exploitation is higher than the rate of regeneration or reproduction. Here the size of the stock (population) depends on the biological, economic and social considerations. Fisheries come under Common Property Resource (CPR), due to which a comprehensive management measure could not be exercised. "In an open access regime like fishery, negative externalities are many, which implies that uncontrolled fishery will bound to end up in what is called tragedy of commons.' (Grafton *et.al,* 2006)

Sustainable Fisheries Yield

The sustainable yield in fishing commonly referred to as "Maximum Sustainable Yield (MSY) is a biological phenomenon. MSY means that level of fish catch or yield that can be harvested from a given system in perpetuity without affecting the stock of the system (or the sea). In other words, a catch level is said to be sustainable whenever it equals the growth rate of the population since it can be maintained for ever. As long as the population size remains constant, the growth rate will remain constant as well.



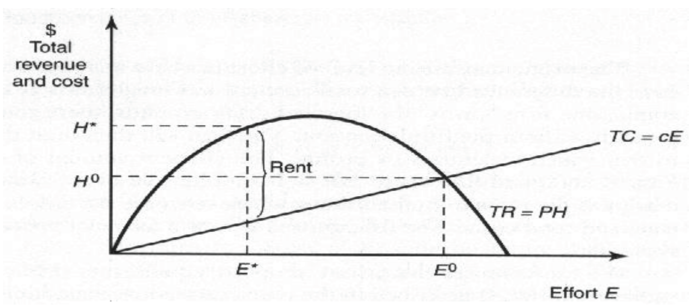
Sustainable Yield Curve

Economics of fisheries management

Economics play a vital role in fisheries management. In the earlier stages, fisheries management focused on controlling the effort to maintain the fish stocks. The common assumption is that if the control measures are strictly implemented, the further increase in effort is prevented and thus a sustainable harvest can be expected. But by 1970 it was found that such measures fail to control the fishing effort and capacity as the fishers substituted from regulated to unregulated inputs (Wilen, 1979) and further remedies suggested also failed to prevent the increase in fishing effort (Townsend, 1990).

"An economic perspective of fisheries management is that marine resources should not only be managed sustainably but also in a way that they contribute to and provide net benefits for the nation as a whole. Indeed the economists argue that sustainable and economically profitable fishery is complimentary. A level of harvest that maximizes the sustainable returns from fishing is often at a stock size that is greater than that which would maximize the overall yield from a fishery. Moreover, if there are other costs associated with fishing like habitat damage or environment loss etc., the economic optimum level of harvest that accounts for these costs would be even less, and the desirable fish stock even larger. In other words, a fishery that is economically optimum in the long run is also likely to be an ecologically sustainable fishery". (Grafton, *et al.*, 2006).

Maximum Economic Yield (MEY) is realized at that level of effort in which the sustainable net return from the fishery is maximum. The difference between the total revenue (TR) and the total cost (TC) is maximum. This difference is also referred to as **resource rent.**



Maximum Economic Yield

Total revenue (*TR*) = Price (*P*) × Catch (*H*)
TC = Unit cost (*c*) × Effort
Rent = TR – TC

The resource rent is maximized at the point E*.

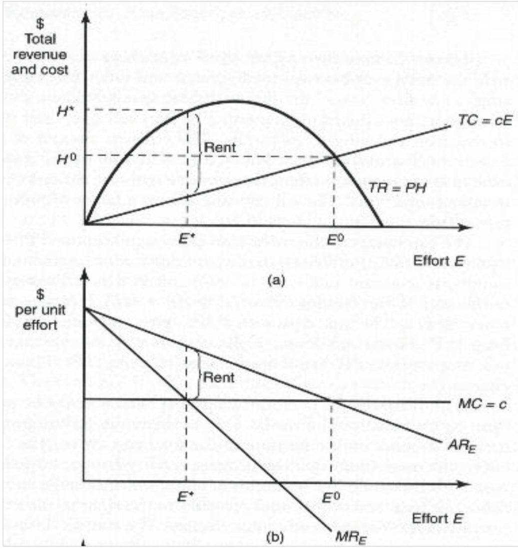
Here *MEY* is left of *MSY*

- ⊙ Optimal harvest (*H**) is less than the *MSY* harvest
- ⊙ But rent is larger than at *MSY*

The marginal analysis can show that the MEY occurs at the point where MC =MR. It is observed that for marginal unit of effort, marginal rent is = 0 and average rent >1.

The point **E* is that effort level at which the MEY occurs**. At this point of effort only the **difference between the total revenue from fishing and total cost of fishing is the maximum.** This difference is also referred to **as resource rent.** "Goal of traditional fisheries management: achieve *MSY*. However the economists aim for MEY in contrast to MSY.

AT MEY, compared to MSY, the fish catch is lower, fishing profit is higher, fishing effort is lower and the fish stock is higher. Thus MEY is where more fish is conserved and economic is the friend of conservation. (Dixon, 2005 and Grafton *et.al.*, 2006).



MEY is affected by the changes in price of fish and the costs of fishing. When the price of fish increases, the total revenue curve shifts upward at all effort levels, leaving the intercepts unchanged and the point of MEY moves closer to MSY but never beyond MSY so long as the cost of fishing increases with effort. On the other hand, if the cost of fishing increases, the total cost curve moves upward to the left, thus the new point of MEY is to the left of the previous MEY. This will lower the optimal fishing effort (E*) because with a more costly harvest, it pays more to have larger stocks from which to catch. In total, a fall in fish price or an increase in cost of fishing will lead to lower harvest with a less fishing effort and a larger stock size in order to maximize the economic profits (Grafton *et.al.*, 2006)

Estimation of Maximum Economic Yield

The Maximum Economic Yield (MEY) and the effort to harvest this Maximum Economic Yield (f_{mey}) can be estimated following Devaraj and Smita (1988) as mentioned below

$$p = a - by \quad (1)$$

where, p is the unit weight of fish, y is the annual yield

The average price per unit weight of fish (p) is generally a monotonically decreasing function of annual yield (y)

The profit is obtained as a difference between total revenue (TR) and total cost (TC), i.e.,

$$\Pi = TR - TC = (p-c)y \quad (2)$$

where 'c' is the cost of harvesting one unit weight of fish. From this, a cost function will be fit from the data collected. Then the MEY and f_{mey} will be obtained using the following equations

$$MEY = (a - c) / 2b \quad (3)$$

$$f_{mey} = [a +/- (a^2 - 4\ b\ MEY)]^{1/2} / 2b \quad (4)$$

where, a = intercept; b, c =regression coefficients

From $f_{mey,}$ the optimum fleet size is obtained by dividing 'b' by the average annual fishing days. Based on this the excess capacity and thus the capital investment (over and above the optimum fleet size) can be worked out.

Beyond Maximum Economic Yield

While estimating the MEY, three assumptions have been made: (i) zero discount rate; (ii) cost of fishing is a simple liner function of stock size; (iii) fishing costs rise proportionately with effort. The discount rate is that interest rate at which future catches are valued today. IF the discount rate is very high or large, the MEY will correspond to a bionomic equilibrium (Clark, 1990), because it will be profitable to harvest the stock today itself if the loss of future net returns are very heavily discounted. The fisheries with commonly used discount rates, MEY will be conservationist and generate stock sizes that are larger than those associated with MSY. This implies that maximizing economic viability of fisheries is compatible with economic sustainability of the fisheries.

The cost of fishing is an important component that decides the MEY. Generally the cost of fishing increases with a decrease in stock size at an increasing rate. This is the characteristics of the fishing practice. Under such a situation, it will be desirable to have a catch and effort level further to the left of the bionomic equilibrium.

In case of multi-species fishery, the estimation of MEY becomes complicated in many ways due to factors like biological interactions, apportioning of the cost of fishing, value of target versus by catches and splitting the efforts and related aspects. Though all these factors can be accommodated in the model, the computation of species-wise MEY becomes very difficult. The concept of uncertainty in fisheries makes the process of estimating MEY more complicated. The sources of uncertainties include lack of complete biological data to calculate the stock-recruitment relationship, inability to accurately measure the actual catch and effort of fishers and the current size of the fish stock. Price of fish and the precise cost of fishing are also the other sources of uncertainties facing the fishery. Such aspects call for a different approach to manage fisheries.

Despite such assumptions, the MEY is a good target reference point for fisheries management. Because MEY ensures that the stock levels in many fisheries are larger than those associated with the traditional MSY target; (ii) ensures that the major inputs like fuel and labour are utilized efficiently so as to maximize the profit (Grafton *et. al.* 2006). If the resources are used beyond the MEY target, it will result in excess fishing capacity, lower returns and thus lower profits. Hence it pays rich dividend to follow the MEY as an important component for aiming at a sustainable fishery.

References

Clarke, C.W., 1990. Mathematical bioeconomics: the Optimal Management of Renewable Resources, *John Wiley & Sons, New York*.

Devaraj, M and Smita, Paralkar.1988. *Economic performance of mechanised trawlers in the State of Kerala, India. Fisheries Research*, 6 (3): 271-286.

Grafton, R.Q., James Kirekly, Toom Kompas and Dale Squires, 2006. *Economics of Fisheries Management*. Ashgate Publishing Company, England, p.160

John A. Dixon**,** 2005. Fisheries and Aquatic Resources. Caspian EVE 2005/UNDP and WBI.

Townsend, R.E., 1990. Entry restrictions in the Fishery: A Survey of the Evidence, *Land Economics,* 66:359-378.

Wilen, J., 1979. Fishermen Behaviour and the Design of Efficient Fisheries Regulation Programmes. *Journal of Fisheries Research Board of Canada*, 36:855-858.

25

Trophic Modelling of Marine Ecosystems and Ecosystem based Fisheries Management

K. Sunilkumar Mohamed

Introduction

Fish populations are an integral part of marine ecosystems. Historically, fish population dynamics have been studied as single species, for example as mackerel, shrimp or sardine, and almost always in isolation from the system in which they exist. In recent years, however, there has been growing awareness that traditional approaches to managing fisheries are incomplete and partially unsuccessful. Sustainable use of living marine resources must consider both the impacts of the ecosystem on the living marine resources, and the impacts of fishery on the ecosystem. This holistic approach to fisheries management has been termed as '*ecosystem based fisheries management*'. 'The Principles of Ecosystem-Based Fisheries Management are: 1. Maintaining the natural structure and function of ecosystems, including the biodiversity and productivity of natural systems and identified important species, is the focus for management. 2. Human use and values of ecosystems are central to establishing objectives for use and management of natural resources. 3. Ecosystems are dynamic; their attributes and boundaries are constantly changing and consequently, interactions with human uses also are dynamic. 4. Natural resources are best managed within a management system that is based on a shared vision and a set of objectives developed amongst stakeholders. 5. Successful management is adaptive and based on scientific knowledge, continual learning and embedded monitoring processes.

A lot of attention has recently been directed at assessing the impacts of fisheries on whole marine ecosystems (ICES, 1998, 2000; Frid *et al*., 1999b; Hall, 1999a, b). This has in part been driven by the need to ensure conservation of biological diversity and sustainable use of the biosphere, key provisions of the convention agreed at the UN Rio summit (Tasker *et al*., 2000). The utilization of sound ecological models as a tool in the exploration and evaluation of ecosystem health and state has been encouraged and endorsed by the leading bodies in ecosystem-based fisheries research and management (NRC, 1999; ICES, 2000). The potential of the available dynamic ecosystem models to make measurable and meaningful predictions about the effects of fishing on ecosystems has not however been fully assessed.

Ecological Factors

Harvesting alters ecosystem structure in ways that are only beginning to be understood. It is argued that long-term heavy commercial harvesting is likely to shift the ecosystem to high-turnover species with low trophic levels (Pitcher and Pauly, 1998). The biological mechanism underlying species shifts is that the relatively large, long-lived fishes which have low mortality rates are more strongly affected by a given fishing mortality rate than are smaller fishes which are part of the same community. A second shift-inducing biological mechanism is habitat degradation caused by various fishing gears especially bottom trawls. Here, the effect is through destruction of bottom structure, depriving benthic fishes of habitats and prey.

Thirdly, the above and the fishery-induced reduction of predatory pressure by benthic fish, may then lead to an increase of small pelagic fish and squids, which becomes available for exploitation. This may mask the decline in catches of the demersal groups. In the Gulf of Thailand, in Hong Kong Bay and other areas of the South China Sea, extremely heavy trawl pressure has resulted in a shift from valuable demersal table fish such as croakers, groupers and snappers to a fishery dominated by small pelagics used for animal feed and invertebrates such as jellyfish and squids.

These mechanisms almost often lead, through a positive feedback loop, to a fourth biological mechanism: harvesting small pelagic fish species at lower trophic levels reduces the availability of food for higher trophic levels, which then decline further, releasing more prey for capture by a fishery that finds its targets even lower down the food web, a process now occurring throughout the world (Pitcher and Pauly, 1998). Some examples of such documented species shifts in exploited multispecies fish communities are shown in table.
Table 1: Examples of documented shifts towards smaller, high-turnover species in exploited multispecies communities (modified from Pitcher and Pauly, 1998)

Fishing grounds/ Stocks (period)	Documented species shift
Gulf of Thailand Demersal stocks (1960-1980)	Overall biomass reduced by 90%; residual biomass dominated by trash fish
Philippine shelf Small pelagics (1950-1980)	Gradual replacement of sardine-like fishes by anchovies
Carigara Bay, Philippines All fish (1970-1990)	Fish replaced by jellyfish, now an export item
North Sea	Halibut and small sharks extinct; cod and haddock threatened; demersal omnivores and small pelagics favoured

Fishing grounds/ Stocks (period)	Documented species shift
Humboldt Current, Chile	Large hake depleted, small pelagics favoured
North Pacific	First marine mammal depletions, followed by huge trawl fisheries: Pollock favoured
South China Sea, Hong Kong	Croakers and groupers almost extinct; small pelagics bulk of fishery

It has also been observed that fishes evolve or change their life histories in response to selective fishing mortality, for e.g., halving of the size of mature Chinook salmon. In this semelparous species early maturity means less time at risk of being caught and therefore, higher fitness. This species has been intensively managed for over 80 years using the best that single species quantitative science can offer, and yet Chinook salmon are on decline.

Socio-Economic Factors

One of the main socio-economic mechanisms, which contribute to species shift, is increasing prices, both for traditional high-value species and for trash species. Such price increases are effective in masking the economic consequences of fishing at lower trophic levels.

Single Species Assessments

The tools developed for single species population dynamics are an essential part of any new methodology. Detailed information on growth, mortality and recruitment schedules and their associated errors and uncertainties are essential for the implementation of the ecosystem approach advocated in the Rio summit. When considering the management of single components of the ecosystem, such as the target fish stocks, it is possible to set target and limit reference points for particular measurable properties of the species. For example, the implementation of precautionary fisheries management in the North Atlantic has progressed through the setting of reference points for various measures of the status of the exploited species, e.g. the spawning stock biomass (SSB). two types of reference point are considered - a limit reference point and a target reference point (Fig.1).

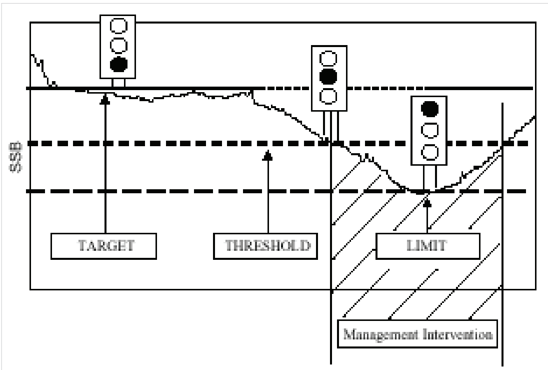


Fig.1. Illustration of target, threshold and limit reference points with regard to spawning stock biomass (from Hall and Mainprize, 2004)

Management measures are aimed at achieving the target reference point in the medium term and ensuring that the limit reference point is never exceeded. In theory, it should be possible to apply reference points to any or all taxa in the ecosystem. ICES (2000) have contended that even if this was practical for a significant number of taxa, it may not ensure adequate protection of all the ecosystem components at risk. There is a need, therefore, to develop reference points for system level emergent properties as a measure of ecosystem health (Hall, 1999a; Gislason et al., 2000).

Ecosystem Modelling

There are many recent developments in building of trophic models of aquatic ecosystems. Such modelling can now be performed more rapidly and rigorously than ever before, providing a basis for viable and practical simulation models that have real predictive power (Christensen and Pauly, 1993; Walters et al., 1997). This was made possible by the development of ECOPATH (Polovina, 1984; Christensen and Pauly, 1992), for construction of mass-balance models of ecosystems, based mainly on diet composition, food consumption rates, biomass and mortality estimates. Such ecosystem models can describe the biomass flows between the different elements of the exploited ecosystems, and can provide answers to 'what if' questions regarding the likely outcome of alternate fishing policies. The ECOPATH suite of software has now been modified (Walters et al., 1997, 2000) to include ECOSIM (simulation module) and ECOSPACE (spatial module). These new routine have not only increased the quantitative power of the approach, but have also allowed qualitatively new questions to be asked. Ecopath applications to ecosystems, ranging from low latitude areas to the tropics, and from ponds, rivers, and lakes to estuaries, coral reefs, shelves, and the open sea, but all using the same metrics, allowed identification of several general features of aquatic ecosystems.

Multivariate comparisons demonstrated the basic soundness of E. P. Odum's (1969) theory of eco-system maturation (Christensen, 1995b), including a confirmation of his detailed predictions regarding ecosystems near carrying capacity (Christensen and Pauly, 1998). Conversely, this theory can now be used to predict the effect of fisheries on ecosystems, which tend to reduce their maturity, as illustrated by the comparison of Ecopath models for the Eastern Bering Sea in the 1950s and early 1990s (Trites *et al*., 1999a, b), and to guide ecosystem rebuilding strategies implied in ''Back to the Future'' approaches (Pitcher, 1998; Pitcher *et al.,* 2000).

The importance (relative to fishing) of predation by fish and marine mammals within marine ecosystems as suggested by complex models in a few areas (North Sea – Andersen and Ursin, 1977; North Pacific – Laevastu and Favorite, 1977) was confirmed globally by Ecopath models (Christensen, 1996; Trites *et al.,* 1997).

Identification of trophic levels as functional entities rather than as concepts for sorting species (Lindeman, 1942; Rigler, 1975) implied the use of non-integer values (computed as 1+ the mean trophic level of the preys, as proposed by Odum and Heald, (1975) that express degree of omnivory (Christensen and Pauly, 1992a), i.e., the extent to which feeding occurs at different trophic levels (Pimm, 1982). Also, trophic level estimated from analyses of stable isotopes of nitrogen has been shown to correlate well with estimates from Ecopath models (Kline and Pauly, 1998). Estimates of transfer efficiencies between trophic levels (Christensen and Pauly, 1993b; Pauly and Christensen, 1995), previously a matter of conjecture usually pertaining to single-species populations or even to studies of a few individual animals (Slobodkin, 1972), differed radically from earlier guesses by ecosystem types (Ryther, 1969) used for inferences on the potential yields of fisheries (Pauly, 1996), even though the mean was unsurprising (about 10%; Morowitz, 1991).

Performance Measures

It is generally agreed that reductions in single species fishing mortality levels is perhaps the most significant step one could take towards ensuring the persistence of marine ecosystems (Hall and Mainprize, 2004). It is also clear that ecosystem based fisheries management is still in its formative years, although substantial developments have been seen in some countries and regions. Among these, North America, Antarctica, Europe, Australia and New Zealand are the most notable.

Table 2. The six principles for an ecosystem based fisheries management approach (adapted from Inter-agency Marine Fisheries Working Group, 2002)

Principle	Description
Ecosystem identification	The ecosystem that fisheries will be managed within need to be defined on the basis of the main physical, biological and human dependency relationships
Clear objectives	Objectives for fisheries management shall have regard to local and national needs, and management should be decentralized to the maximum extent possible
Long term benefits	Ecosystem based management should aim for long term benefits – management should look to restore stocks to levels that are capable of delivering optimal yields over the long term; and achieving such yields should not compromise other marine species and habitats. Management should also aim to support biological biodiversity
Incentives aligned with and ecosystem based approach	Incentives should be realigned to support aims of the ecosystem based approach – incentives and financial support needs to be redirected from fisheries that aim at increasing fishing efficiency to those that make concerted efforts to those that promote the restoration of fish stocks to optimal yield levels and which support responsible fishing practices in sensitive marine areas.

Principle	Description
Easily assessed information and alternate management options	Information necessary to implement the ecosystem based approach should be made available to all. Where information is insufficient, adaptive management and the precautionary approach should be followed. If the outcome falls short of what was intended the management decisions should be suitably altered – proactive management

Unfortunately, despite the legislative imperative and clearly articulated principles (Table 2), arriving at an operational framework for an ecosystem-based approach to fisheries management is fraught with difficulties. This difficulty is due, not only to the inherent challenge in establishing and quantifying the effects of fishing at an ecosystem level, but also due to the social and political dimensions associated with harvesting fisheries at an environmentally sustainable level.

An Overview of Ecopath & Ecosim

The Ecopath software is a simple approach for analyzing trophic interactions in fisheries resources systems (Christensen and Pauly 1992a,b, 1995). Ecopath is based on the earlier work of Polovina (1984), and is being widely applied to aquatic systems (Christensen and Pauly 1993, Pauly and Christensen 1995). It is a mass-balance approach that describes an ecosystem at steady-state for a given period. Further development of this steady-state model has resulted in a dynamic ecosystem model called Ecosim that is capable of simulating ecosystem changes over time (Walters et al., 1997). Ecopath and Ecosim represent all of the major components of the ecosystem, and their feeding interactions, but are relatively simple. These kinds of models readily lend themselves to answering simple, ecosystem wide questions about the dynamics and the response of the ecosystem to anthropogenic changes. Thus, they can help design policies aimed at implementing ecosystem management principles, and can provide insights into the changes that have occurred in ecosystems over time. Ecopath models rely on the truism that:

biomass accumulation + **fisheries catch** + **mortality due to predation** + **other mortality** + **loss to adjacent systems** = **Production**

This applies for any producer (e.g., a given fish population) and time (e.g., a year or season). Groups are linked through predators consuming prey, where:

production + **non-assimilated food** + **respiration** = **Consumption**

The implication of these two relationships is that the system or model is mass balanced (i.e., biomass is 'conserved', or accounted for in the ecosystem). This principle of mass conservation provides a rigorous framework – formalized through a system of linear equations – through which the biomass and trophic fluxes among different consumer groups within an ecosystem can be estimated (Christensen and Pauly 1995). Constructing an Ecopath model emphasizes ecological relationships rather than mathematical equations. All that is required are the types of data that are routinely collected by fisheries scientists and marine biologists. The model can incorporate and standardize large amounts of scattered information – information that might have otherwise languished in scattered journals, reports and filing cabinets (Christensen and Pauly 1995).

Ecopath is essentially a large spreadsheet that is simultaneously keeping track of all the species and all the feeding interactions occurring within the ecosystem. It describes the ecosystem at one point in time. Ecosim, which is based on the Ecopath equation, simulates how a change in one or more components might affect the ecosystem over time.

Ecopath and Ecosim have been widely applied in recent years. More than 80 Ecopath systems have so far been published world-wide. They span a diversity of systems including upwelling, shelves, lakes and ponds, rivers, open oceans and even terrestrial farming systems (see Christensen and Pauly 1992a,b, 1995; Walters et al. 1997; and the Ecopath home page at http://www.ecopath.org)

Principles of the Ecopath Model

The core routine of Ecopath is derived from the Ecopath program of Polovina (1984), and since modified to make superfluous its original assumption of steady state. Ecopath no longer assumes steady state but instead bases the parameterization on an assumption of mass balance over an arbitrary period, usually a year. In its present implementation Ecopath parameterizes models based on two master equations, one to describe the production term and one for the energy balance for each group.

The first Ecopath equation describes how the production term for each group (i) can be split in components. This is implemented with the equation,
Production = catches + predation mortality + biomass accumulation + net migration + other mortality;

or, more formally, $P_i = Y_i + B_i \cdot M2_i + E_i + BA_i + P_i \cdot (1 - EE_i)$ Eq.1

where Pi is the total production rate of (i), Yi is the total fishery catch rate of (i), M2i is the total predation rate for group (i), Bi the biomass of the group, Ei the net migration rate (emigration - immigration), BA i is the biomass accumulation rate for (i), while M0i = Pi · (1-EEi) is the .other mortality. rate for (i).

This formulation incorporates most of the production (or mortality) components in common use, perhaps with the exception of gonadal products. Gonadal products however nearly always end up being eaten by other groups, and can be included in either predation or other mortality.

Eq. 1 can be re-expressed as

$$B_i \cdot (P/B)_i \cdot EE_i - \sum_{j=1}^{n} B_j \cdot (Q/B)_j \cdot DC_{ji} - Y_i - E_i - BA_i = 0 \qquad \text{Eq. 2}$$

where: *P/ Bi* is the production/biomass ratio, Q/Bi is the consumption / biomass ratio, and DCj i , is the fraction of prey (i) in the average diet of predator (j).

Of the terms in Eq. 2 the production rate, Pi, is calculated as the product of Bi, the biomass of (i) and Pi/Bi, the production/biomass ratio for group (i). The Pi/Bi rate under most conditions corresponds to the total mortality rate, Z, see Allen (1971), commonly estimated as part of fishery stock assessments. The other mortality is a catch-all term including all mortality not elsewhere included, e.g., mortality due to diseases or old age, and is internally computed from,

$$M0_i = P_i \cdot (1 - EE_i)$$

where EEi is called the ecotrophic efficiency of (i), and can be described as the proportion of the production that is utilized in the system. The production term describing predation mortality, M2, serves to link predators and prey as,

$$M2_i = \sum_{j=1}^{n} Q_j \cdot DC_{ji} \qquad \text{Eq. 3}$$

where the summation is over all (n) predator groups (j) feeding on group (i), Qj is the total consumption rate for group (j), and DCji is the fraction of predator (j) diet contributed by prey (i). Qj is calculated as the product of Bj, the biomass of group (j) and Qj/Bj, the consumption/biomass ratio for group (j).

An important implication of the equation above is that information about predator consumption rates and diets concerning a given prey can be used to estimate the predation mortality term for the group, or, alternatively, that if the predation mortality for a given prey is known the equation can be used to estimate the consumption rates for one or more predators instead.

- biomass
- production/biomass ratio
- consumption/biomass ratio or
- ecotrophic efficiency

For parameterization, Ecopath sets up a system with (at least in principle) as many linear equations as there are groups in a system, and it solves the set for one of the parameters for each group depicted in the infographic.

While the other three parameters along with parameters given in the infographic must be entered for all groups. It was indicated above that Ecopath does not rely on solving a full set of linear equations, i.e., there may be less equations than there are groups in the system. This is due to a number of algorithms included in the parameterization routine that will try to estimate iteratively as many missing parameters as possible before setting up the set of linear equations.

- catch rate
- net migration rate
- biomass accumulation rate
- assimilation rate and
- diet compositions

ECOSIM – Dynamic mass-balance approach for Ecosystem Simulation

By converting the linear equations of Ecopath models to differential equations, Ecosim provides a dynamic mass-balance approach, suitable for simulation (Walters et. al. 1997). Constructing a dynamic model from equation (1) there are three changes viz; (a) replace the left side with a rate of change of biomass; (b) for primary producers, provide a functional relationship to predict changes in (P/Bi) with biomass Bi (representing competition for light, nutrients and space); and (c) replace the static pool-pool consumption rates with functional relationships predicting how consumption will change with changes in biomass of Bi and Bj. The basics of ECOSIM consist of biomass dynamics expressed through a series of coupled differential equations. The equations are derived from the ECOPATH master equation (Eq.1), and take the form

$$dB_i / dt = g_i \sum_j C_{ji} - \sum_j C_{ij} + I_i - (M_i + F_i + e_i) B_i \qquad \text{Eq.4}$$

where dBi/dt represents the growth rate during the time interval dt of group (i) in terms of its biomass, Bi, gi is the net growth efficiency (production/consumption ratio), Mi the non-predation (other) natural mortality rate, Fi is fishing mortality rate, ei is emigration rate, Ii is immigration rate, (and ei·Bi-Ii is the net migration rate). The two summations estimates consumption rates, the first expressing the total consumption by group (i), and the second the predation by all predators on the same group (i). The consumption rates, Cji, are calculated based on the foraging arena concept, where Bi.s are divided into vulnerable and invulnerable components (Walters et al. 1997), and it is the transfer rate (vij) between these two components that determines if control is top-down (i.e., Lotka-Volterra), bottom-up (i.e., donor-driven), or of an intermediate type. The set of differential equations is solved in Ecosim using (by default) an Adams-Basforth integration routine or (if selected) a Runge-Kutta 4th order routine.

Using previously constructed Ecopath models, Ecosim calculates corresponding changes in biomass of each component when the fishing mortality of any particular group is altered. These dynamic simulations are plotted as coloured biomass curves. The scale differs for each curve. By altering the rate of flow between vulnerable and non-vulnerable prey different functional relationships for predators and prey can be considered. These can range from pure donor control, where the prey availability governs interactions, to top-down control where predation pressure dominates. Using equilibrium simulations, where equilibrium biomass is plotted over a range of F values, Ecosim provides the facility to predict the potential equilibrium yield for the fished group.

Trophic Modelling Studies in India

Trophic modelling studies in Indian aquatic ecosystems are few. The first preliminary attempt was made in small ecosystem in Veli Lake near Thiruvanathapuram. Subsequently another preliminary attempt was made to model the southwest coast ecosystem using already existing data and many assumptions (Vivekanadan et al. 2003). The first major targeted attempt to study was that of the model for the Arabian Sea off Karnataka (Mohamed et al. 2008; Mohamed and Zacharia, 2009). This Ecopath model had a pedigree index of 0.521 (scale from 0 for data that is not rooted in local data up to a value of 1 for data that are fully rooted in local data). The Karnataka model encompassed an area of 27,000 km^2 (from the shore to the edge of the continental shelf) and had 24 functional ecological groups (species assemblages) of which 23 were living groups and one dead group (detritus). Ecological groups ranged from apex predators like marine mammals, sharks and tunas to micro zooplankton and phytoplankton.

A comparison of ecosystem parameters from other parts of the world is given in table below (modified from Trites *et al.*, 1999) above. The total throughput for the Arabian Sea ecosystem of Karnataka ranks third after Peru and Monterey bay and is double that of Bering Sea and Venezuela upwelling ecosystem. The gross efficiency of the fishery (catch/PP) value obtained for Karnataka is close to that of the Peruvian ecosystem, which is also an upwelling ecosystem, harvesting fishes low in the food chain. The omnivory index is quite high comparatively for the Karnataka ecosystem indicating the complex feeding interactions in the ecosystem. The estimated ascendancy values for the Arabian Sea ecosystem of Karnataka indicate that it has not reached its full development capacity, unlike the Yacutan and Monterey bay ecosystems. The recycling capacity of the ecosystem throughput as indicated by the cycling index shows that recycling in Arabian Sea ecosystem of Karnataka is only moderate as compared to ecosystems like Brunei and Bering Sea.

Ecosystems	Through put	Catch /PP	PP/B	B/T	Net syst. prod.	Omnivory Index	Ascen-dency	Cycling Index	Path length
Yacutan	2362	0.0029	27.4	0.036	370	0.134	44.0	2.8	2.84
N. Gulf of Mexico	1790	0.0002	7.0	0.015	19	0.195	39.1	2.1	3.03
Venezuela (upwell.)	5309	0.0016	27.0	0.023	831	0.135	39.9	2.2	4.05
Brunei, SE Asia	1816	0.0008	28.6	0.018	300	0.201	29.4	16.3	2.80
Peru 70 (upwell.)	18800	0.0017	87.5	0.012	14709	0.169	38.1	8.7	3.63
Monterey	17513	0.0012	1.2	0.012	2208	0.324	66.2	4.4	3.63
Alaska Gyre	5946		38.1	0.015	407	0.103	42.3		2.03
British Columbia Shelf	1237		21.1	0.180	4106	0.140	40.1		2.03
Bering Sea 50's	6535	0.0002	5.9	0.050	-115	0.183	32.5	13.2	3.47
Bering Sea 80's	5692	0.0021	4.9	0.050	-356	0.157	30.9	11.1	3.51
Karnataka Arabian Sea	**11522**	**0.0016**	**29.9**	**0.012**	**904**	**0.299**	**33.0**	**6.03**	**2.81**

References

Christensen, V. 1995b. Ecosystem maturity - towards quantification. Ecol. Modelling 77: 3-32.

Christensen, V. 1996. Managing fisheries involving predator and prey species. *Reviews in Fish Biology and Fisherie*s, 6:417-442.

Christensen, V. and D. Pauly, (eds.) 1993. Trophic models of aquatic ecosystems. ICLARM Conference Proceedings No. 26. ICLARM Manila, Philippines, 390p.

Christensen, V. and D. Pauly. 1992a. ECOPATH II - a software for balancing steady-state ecosystem models and calculating network characteristics. Ecol. Modelling 61:169-185.

Christensen, V. and D. Pauly. 1992b. A guide to the ECOPATH II program (version 2.1). ICLARM Software 6, 72 p.

Christensen, V. and D. Pauly. 1993. Flow characteristics of aquatic ecosystems, p. 338-352. In V. Christensen and D. Pauly (eds.) Trophic models of aquatic ecosystems. ICLARM Conference Proceedings No. 26.

Christensen, V. and D. Pauly. 1995. Fish production, catches and the carrying capacity of the world oceans. Naga, The ICLARM Quarterly 18(3):34-40.

Christensen, V. and Pauly, D. 1998. Changes in models of aquatic ecosystems approaching carrying capacity. Ecological Applications, 8(1), (Suppl): 104-109.

Christiansen, V., C.J. Walters, and D. Pauly, 2000. ECOPATH with ECOSIM: A User's Guide, October 2000, Fisheries Centre, University of British Columbia, Vancouver, Canada and International Center for Living Aquatic Resources Management (ICLARM), Penang, Malaysia. 125p.

Finn, J.T., 1976. Measures of ecosystem structure and function derived from analysis of flows. J. Theor. Biol. 56: 363-380.

Frid CLJ, Hansson S, Ragnarsson SA, Rijnsdorp A, and SA. Steingrimsson. 1999b.Changing levels of predation on benthos as a result of exploitation of fish populations. *Ambio* 28: 578 -582.

Hall, S.J. 1999a. *The effects of fishing on marine ecosystem and communities*. Blackwell Science, Oxford.

Hall, S.J. 1999b. Managing fisheries within ecosystems: can the role of reference points be expanded? *Aquatic Conservation*: *Marine and Freshwater Ecosystems* 9. 579-583.

Hall, S. J and B. Mainprize. 2004. *Towards ecosystem-based fisheries management* Blackwell Publishing Ltd. Fish and Fisheries, 5, 1-20.

Hannon, B. and C. Joiris. 1989. A seasonal analysis of the southern North Sea ecosystem. Ecology 70(6):1916-1934.

Hannon, B., 1973. The structure of ecosystems. J. Theor. Biol. 41:535-546.

Hurlbert, S. H. 1978. The measurement of niche overlap and some relatives. Ecology 59:67-77.

ICES.1998. Report of the Working Group on the Ecosystem Effects of Fishing Activities ICES, Copenhagen, ICES CM 1998/ACFM/ACME: 01 Ref.: E, 263pp.

ICES 2000. Ecosystem effects of fishing. *In: ICES Marine Science Symposia*. 210. ICES Journal of Marine Science 57, 791.

Ivlev, V.S. 1961. Experimental ecology of the feeding of fishes (Transl. by D. Scott). Yale University Press, New Haven, 302 p.

Kavanagh, P. 2002. Automated Mass Balance Procedure for ECOPATH Ecosystem Models User's Guide. Fisheries Centre, University of British Columbia. 28p.

Kline, T.C. Jr, and D. Pauly. 1998. Cross-validation of trophic level estimates from a mass-balance model of Prince William Sound using 15N/14N data. In Fishery Stock Assessment models, pp.693-702. Ed. by T.J. Quinn II, F. Funk. J. Heifetz, J. N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.I. Zhang. Alaska Sea Grant, Fairbanks

Laevastu, T., and F. Favorite. 1977. Preliminary report on the dynamical marine ecosystem model (DYNUMES) for Eastern Bering Sea. Manuscript Report, National Marine Fisheries Service, Northwest and Alaska Fisheries Center, Seattle.

Leontief, W.W., 1951. The structure of the U.S. economy, 2nd ed. Oxford University Press, New York.

Lindeman, R.L. 1942. The trophic-dynamic aspect of ecology. Ecology 23: 399-418.

Loman, J. 1986. Use of overlap indices as competition coefficients: tests with field data. Ecol. Modelling, 34:231-243.

Majkowski, J. 1982. Usefulness and applicability of sensitivity analysis in a multispecies approach to fisheries management, p. 149-165. In D. Pauly and G.I. Murphy (eds.) Theory and management of tropical fisheries. ICLARM Conf. Proc. 9. 141p.

Manickchand-Heileman, S., F. Arreguin-Sanchez, A Lara-Dominguez and L.A. Soto. 1998. Energy flow and network analysis of Terminos Lagoon, SW Gulf of Mexico. J. Fish. Biol. 58: 179-197.

Mendoza, J.J. 1993. A preliminary biomass budget for the northeastern Venezuela shelf ecosystem, p. 285-297. In: V. Christensen and D. Pauly (eds) Trophic models of aquatic ecosystems. ICLARM Conf. Proc. 26, 390p.

Mohamed K S, Zacharia P U, Muthiah C, Abdurahiman K P & Naik T H, (2008) A trophic model of the Arabian Sea ecosystem off Karnataka and simulation of fishery yields for its multigear marine fisheries, Bull. Cent. Mar. Fish. Res. Inst. 51: 140p.

Mohamed, K.S. and P.U. Zacharia. (2009). Prediction and modelling of marine fishery yields from the Arabian Sea off Karnataka using Ecosim. Indian J. Mar. Sci. 38 (1): 69-76.

Nee, S. 1990. Community construction. Trends Ecol. Evol. 5(10):337-339.

NRC 1999. Sustaining marine fisheries. National Research Council, National Academy Press, Washington, D.C. 164pp

Odum, E.P. 1969. The strategy of ecosystem development. Science 104:262-270.

Odum, W.E. and E.J. Heald. 1975. The detritus-based food web of an estuarine mangrove community, p. 265-286. *In* L.E. Cronin (ed.) Estuarine research, Vol. 1. Academic Press, New York.

Pauly, D. 1996. One hundred million tonnes of fish, and fisheries research. Fisheries Research, 25(1): 25-38.

Pauly, D. and V. Christensen. 1995. Primary production required to sustain global fisheries. Nature 374: 255-257.

Pauly, D., M. Soriano-Bartz and M.L. Palomares. 1993. Improved construction, parametrization and interpretation of steady-state ecosystem models. p. 1-13, In V. Christensen and D. Pauly (eds.) Trophic models of aquatic ecosystems. ICLARM Conference Proceedings No. 26.

Pauly, D., V. Christensen and V. Sambilay. 1990. Some features of fish food consumption estimates used by ecosystem modellers. ICES Counc. Meet. 1990/G:17, 8 p.

Pianka, E. R. 1973. The structure of lizard communities. Ann. Rev. Ecol. Syst. 4: 53-74.

Pimm, S. L. 1982. Food webs. Population and Community Biology Series. Chapman and Hall, London. 219pp.

Pinkas, L., M.S. Oliphant and I.L.K. Iverson. 1971. Food habits of albacore, bluefin tuna and bonito in California waters. Calif. Fish Game 152: 1-105.

Pitcher, T.J. 1998. Back to the future: a novel methodology and policy goal in fisheries. In Back to the future: Reconstructing the Strait of Georgia Ecosystem, pp. 4-7. Ed. by D.Pauly, T.J. Pitcher, and D. Preikshot. Fisheries Center Research Reports, 6(5).

Pitcher, T.J., Courtney, A., Watson. R. and D. Pauly. 1998. Assessment of Hong Kong's inshore fishery resources. Fisheries Centre Reports, Vancouver. 6(1), 149pp

Pitcher, T.J., Watson, R., Haggan, N., Guenette, S.Kennish, R., and Sumaila, R. 2000. Marine Reserves and the Restoration of Fisheries and Marien Ecosystems in the South China Sea. Proc. Mote Symposium on Marine Protected Areas, Sarasota Springs, Florida. November 1998. Bulletin of Marine Scienes, 66(3):

Polovina, J.J. 1984a. Model of a coral reef ecosystems I. The ECOPATH model and its application to French Frigate Shoals. Coral Reefs 3(1):1-11.

Polovina, J.J. 1984b. An overview of the ECOPATH model. Fishbyte 2(2):5-7.

Qasim, S.Z., 1972. The dynamics of food and feeding habits of some marine fishes. Indian J. Fish., 19 (1& 2): 11-28.

Rigler, F.H. 1975. The concept of energy flows and nutrient flows between trophic levels. In Unifying Concepts in Ecology, pp.15-26. Ed. by W.H. van Dobben, and R.H. Lowe-McConnel. Dr. W. Junk B.V. Publishers, The Hague.

Ryther, J. 1969. Photosynthesis and fish production in the sea. Science, 166: 72-76.

Slobotkin, L. B. 1972. Growth and Regulation of Animal Populations. Holt, Rhinehart and Winston, New York. 184pp.

Sukumaran, K.K., V.D. Deshmukh, G.S. Rao, K. Alagaraja and T.V. Sathianandan 1993. Stock assessment of the penaeid prawn *Metapenaeus monoceros* Fabricius along the Indian coast. Indian J. Fish. 40: 20-34.

Tasker M.L, P.A. Knapman and D. Laffoley .2000.Effects of fishing on non-target species and habitats: identifying key nature conservation issues. In: Effects of Fishing on Non-target Species and Habitats, Kaiser M.J, de Groot S.J (eds). Blackwell Scientific Publishers: Oxford; 281 -289.

Trites, A. W., Christensen, V. & D. Pauly. 1997. Competition between fisheries and marine mammals for prey and primary production in the Pacific Ocean. *Journal of Northwest Atlantic Fishery Science* 22, 173–187.

Trites, A. W., Livingston, P. A., Vasconcellos, M. C., Mackinson, S., Springer, A., and D. Pauly.1999a. Ecosystem consideration and the limitations of Ecosim models in fisheries management: insights from the Bering Sea. In Ecosystem Approaches for Fisheries Management, pp. 609-618. Ed. by S. Keller. University of Alaska Sea Grant, Fairbanks.

Trites, A.W, P.A. Livingston, S. Mackinson, M.C. Vasconcellos, A.M. Springer and D. Pauly. 1999b. Ecosystem change and the decline of marine mammals in the eastern Bering

Sea: testing the ecosystem shift and commercial whaling hypothesis. Fisheries Centre Research Reports. Vol 7: 98p.

Ulanowicz, R.E. 1986. Growth and development: ecosystem phenomenology. Springer Verlag, New York. 203p.

Ulanowicz, R.E. 1995. The part-whole relation in ecosystems. p. 549-560. In B.C. Patten and S.E. Jørgensen (eds.) Complex ecology. Prentice-Hall, Englewood Cliffs, New Jersey.

Ulanowicz, R.E. and C.J. Puccia. 1990. Mixed trophic impacts in ecosystems. Coenoses 5:7-16.

Ulanowicz, R.E. and J.S. Norden. 1990. Symmetrical overhead in flow and networks. Int. J. Systems Sci. 21(2):429-437.

Vitousek, P.M., P.R. Ehrlich, A.H. Ehrlich, and P.A. Matson. 1986. Human appropriation of the products of photosynthesis. BioScience, 36:368-373.

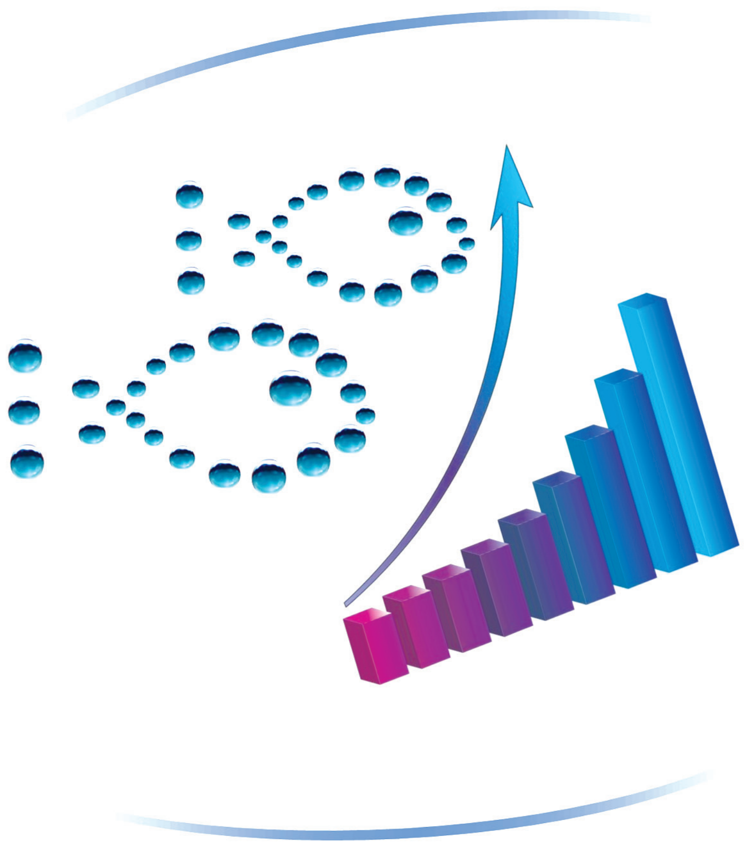
Walters, C.J., V. Christensen and D. Pauly. 1997. Structuring dynamic models of exploited ecosystems from trophic mass-balance assessments. Rev. Fish Biol. Fish. 7, 139-172.

Walters, C. 2000. Impacts of dispersal, ecological interactions, and fishing effort dynamics on efficacy of marine protected areas: how large should protected areas be? Proc. Mote Symposium on Marine Protected Areas, Sarasota Springs, Florida, November 1998. Bulletin of Marine Science 66 (3):

Wulff, F. and Ulanowicz, R.E., 1989. A comparative anatomy of the Baltic Sea and Chesapeake Bay ecosystems, p. 232-256. In F. Wulff, J.G. Field and K.H. Mann (eds.) Network analysis in marine ecology - methods and applications. Coastal and Estuarine Studies, Vol. 32. Springer-Verlag, New York.

Training Manual

Fish Stock Assessment and Management



ISBN 978-93-82263-07-4