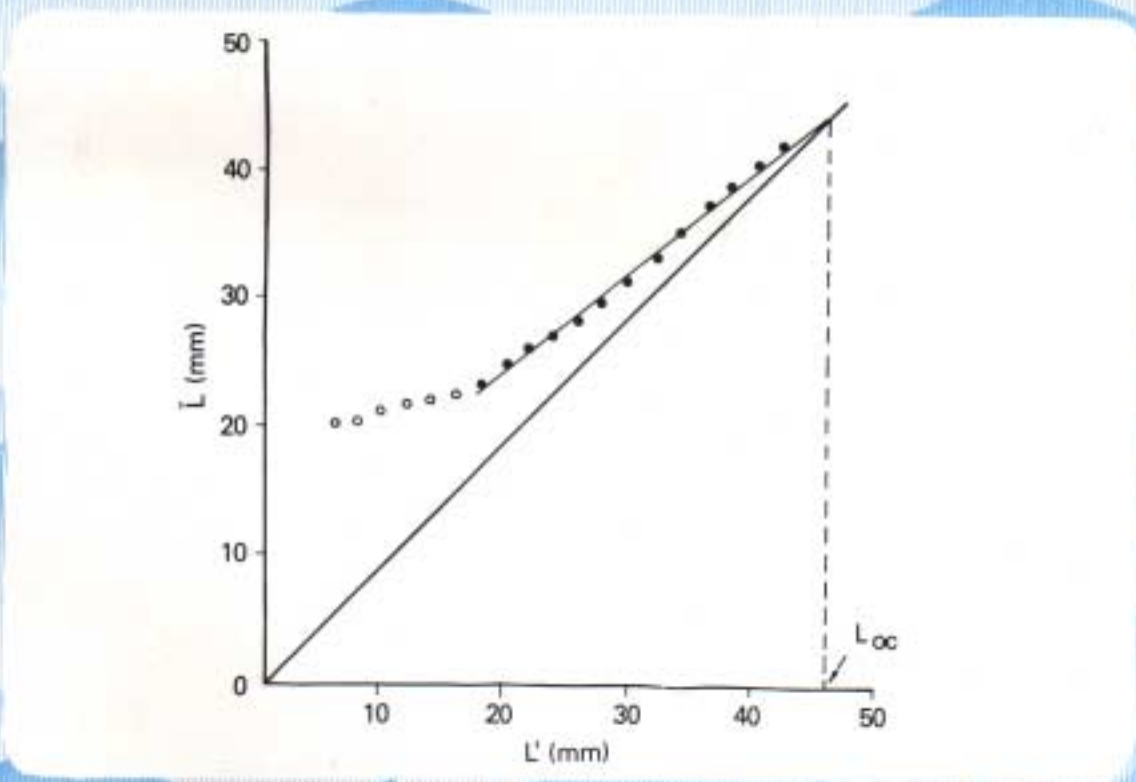


Separating mixtures of normal distributions: Basic programs for Bhattacharya's method and their applications to fish population analysis



BAY OF BENGAL PROGRAMME
Marine Fishery Resources Management

BOBP/MAG/4
(RAS/81/051)

SEPARATING MIXTURES OF
NORMAL DISTRIBUTIONS:
BASIC PROGRAMS FOR
BHATTACHARYA'S METHOD
AND THEIR APPLICATIONS
TO FISH POPULATION ANALYSIS

by H. Goonetilleke
Computer Technician, BOBP

K. Sivasubramaniam
Senior Fisheries Biologist BOBP

Executing Agency :
Food and Agriculture Organization
of the United Nations

Funding Agency :
United Nations
Development Programme

Marine Fishery Resources Management in the Bay of Bengal.
Colombo, Sri Lanka, November, 1987
Mailing Address : C/o FAO, P O Box 1505, Colombo 7, Sri Lanka.
Street Address : NARA Building, Crow Island, Mattakuliya, Colombo 15, Sri Lanka.
Cables : FOODAGRI Telex: 2203 A/B FAOR CE Phones:

This manual describes computer programs written in Microsoft BASIC (version 2.20B) for use on an Apple IIe micro-computer (with CP/M operating system) and EPSON RX-80 F/T printer.

These programs include a modification of Bhattacharya's method for identifying and separating normal distributions, in a mixture of distributions in length frequency data, and computation of Chi square statistics for testing goodness of fit, as described by Pauly and Caddy (1985). This package includes a program that plots the frequency distributions for several samples of data. It is useful for estimating growth parameters and calculates the catch at length for each of the mean length groups separated, and gives 95% and 99% confidence limits for each mean.

Growth parameters can also be estimated using length-at-age data by the Ford-Walford Plot and von Bertalanffy Plot programs. In cases where the ages of fish are not known and a series of sizes at relative ages is not available, growth parameters can still be estimated using data on length increase in time (such as also obtained from tagging-recapture data) by the Gulland and Holt Plot program.

This package also includes Sparre's, Jones' and Van Zalinge's method for estimating total mortality coefficients from length frequency data.

A new method for estimating the asymptotic length and mortality coefficient as proposed by J.A. Wetherall is also presented.

This manual and the programs have been prepared by the BOBP project "Marine Fishery Resources Management in the Bay of Bengal" (RAS/81/051). The project commenced January 1983 and terminated December 1986. It was funded by the UNDP.

This document and the programs have not been officially cleared by the governments concerned or by the FAO.

CONTENTS

	<i>Page</i>
1. Introduction	1
2. Getting started	2
3. Bhattacharya method	4
4. Chi square statistics for testing results	6
5. Some applications of Bhattacharya's method to fish population analysis	7
6. Total Mortality Estimation	10
7. References cited	13
 <i>Appendices</i>	
1. Detailed program listing	14
2-A. Percentage points of the X^2 distribution	42
B. Percentage points of the t distribution	43
3. Bhattacharya analysis	44
4. Chi square statistics	46
5. Plot of percentage frequencies	47
6. Ford-Walford plot	50
7. Gulland & Holt plot	51
8. Von Bertalanffy plot	51
9. Jones & Van Zalinge method	52
10. Sparre's catch curve method	53
11. Wetherall et al. method	54
<i>Publications of the Bay of Bengal Programme</i>	56

1. INTRODUCTION

Programs for applications of Bhattacharya's method, Ford-Watford Plot, Gulland and Holt Plot and the Jones and Van Zalinge method were first developed for use with HP-67/97 programmable calculators. (Pauly and Caddy 1985, Pauly 1984);

When dealing with large data sets, comprising a large number of samples, methods based on pocket calculators were found to be somewhat tedious.

With advances in technology — programmable calculators are already being replaced by micro computers -the need has arisen for translating some of these programs to micro-computer BASIC.

The user's instructions in this manual give a step by step instruction format, to enable users with no previous experience with computers or programmable calculators, to operate the programs without any difficulty.

The first two programs presented here, namely —

- (a) modification of the Bhattacharya method for separation of normally distributed groups from a mixture of normal distributions, and
- (b) computation of Chi square statistics to assess whether the groups identified generate a mixture of normal distributions that fit the original frequencies are discussed by Pauly and Caddy (1985). For the benefit of the users of this program, a brief description of the theory, assumptions and limitations is also presented.

The present program accepts the data for an entire frequency distribution, to identify all the normal distributions which exhibit a regression coefficient (r) with value equal to or higher than the critical r value chosen by the user. The program also displays r values for the remaining and unaccounted frequencies, to assist the user in choosing the critical values for a second and third run. The rest of the programs deal with the application of Bhattacharya's method to fish population analysis.

With these programs, separation of normal distributions, estimation of their mean length values (with 95% and 99% confidence limits) and calculation of catch at each mean length can be executed with length frequency data for 12 months or 4 quarters of a year or even for a number of years. This analysis is expected to provide evidence of seasonal or annual changes in the composition of size groups, modal progressions and parameters for growth and total mortality rate estimation. However, for the latter estimation, the length frequencies should be raised to the total catch of the species before entering the data for the analysis, and any adjustments for gear selectivity should also be incorporated.

The symbols used and derivations of the formulae are not explained in this user manual. For details of all the analyses carried out by the programs, users are requested to refer Pauly (1984),* Pauly and Caddy (1985)* and Sparre (1985).

A graphical plot of percentage frequencies for all samples is provided to enhance growth parameter estimation and to identify incorrect assignment of a mode to a component. Note that the graph will be printed only with an EPSON RX-80 F/T printer.

For estimating growth parameters, the program provides the user with three optional methods, namely Ford-Walford Plot, Gulland and Holt Plot, and von Bertalanffy Plot. For theoretical aspects of these methods, users are referred to Pauly (1984) and Sparre (1985).

Three graphical methods are also provided in these programs for estimating total mortality rates.

Detailed program listings (Appendix 1) in Microsoft BASIC language, Chi square and t-distributions and sample printouts are also included for the benefit of the users.

* Errata sheets for these two documents may be obtained by writing to ICLARM, MCC P.O. Box 1501, Manila, Philippines.

2. GETTING STARTED

The program will automatically execute when the Apple IIe is booted (powered on).

- (a) Turn on your video display unit.
- (b) Insert the program disk into drive A and close the drive door.
- (c) Press the power switch of the computer.

Within a few seconds, without the user having to enter any commands, the computer will display the following menu:

MAIN MENU

1. FILE MAINTENANCE
2. BHATTACHARYA'S METHOD
3. CHI SQUARE STATISTICS
4. ESTIMATING GROWTH PARAMETERS
5. TOTAL MORTALITY ESTIMATION
6. QUIT TO CP/M

Enter option ?

Remember that option "1" has to be selected before option "2", "3", item 1 of option "4" or "5", because the data entered in option "1" will be necessary for the options mentioned.

At the end of a run, remember to enter option 6, to exit from your programs. Then CP/M prompt A > will appear on the screen, indicating that you can remove the disk from the computer and turn the power off from the display unit and the computer.

2.1 Running instructions

Choose option "1" from main menu to display the following screen:

FILE MAINTENANCE MENU

1. CREATE A NEW DATA FILE
2. EDIT AN EXISTING DATA FILE
3. LIST/DISPLAY DATA FILE
4. DISPLAY DISK CATALOG
5. DELETE DATA FILES
6. INITIALISE NEW DATA DISK
7. REGROUPING OF LENGTH FREQUENCY DATA
8. RETURN TO MAIN MENU

if "1", "2", "3", "5" or "7" is entered, the prompt "enter file name (maximum 8 characters) ?" appears on the screen.

The file name can be a combination of letters and numbers but should begin with a letter. Do **not** enter a name with more than eight characters. Remember to enter unique names for all the

data files. File name cannot be blank. Do not use a dot (.) in the file name since CP/M uses a dot to differentiate between different kinds of files.

For creating a new data file (option "1") several parameters have to be input by the user. With each file, the user can enter any particulars describing the file at this stage. This description of the file will be printed in all the reports under the file name, enabling the user to easily identify the print-outs. This description can be up to 254 characters in length.

Enter the number of samples and number of classes. Then enter the lower and upper limits of the smallest size class for the program to calculate the interval size. Next, the program will display the following prompts for each sample with the sample number appearing on the screen.

Month ?

First midlength of this sample?

Last midlength of this sample?

To enable the user to distinguish between the same months of two different years (in the case of more than 12 samples), the computer will accept any numbers (even >12) for the month and will not perform any month checking here.

Then the frequencies corresponding to the sample can be entered from the first midlength to the last midlength. Note that zero frequencies will not be accepted. If "return" is pressed or zero is entered for frequency, then the computer will display "enter 0.01 for zero frequency."

Before these frequencies are written on the disk, a chance is given to the user to edit the data for that sample. Enter Y or N for the question "Do you wish to edit data for this sample (Y/N) ?". If "N" is entered the next sample data can be entered. If "Y" is entered, the computer asks for the number of class to be edited. Enter the number corresponding to the row in which data needs editing. If a number is input which is greater than the total number of length classes in that sample, the computer displays "no such length class" and asks for the number of class to be edited. If the user does not have any other class to edit, then enter zero to quit editing.

The data entering continues until entry of data for all the samples is completed. Then the computer writes these data records in the disk under the file name initially given by the user.

A data file has to be created in this manner before any of the Bhattacharya, Chi square or % frequency plot programs could be run. Once a data file is created, the user can list the contents of this file by using option "3" of the file maintenance menu. Once the name of the data file is entered the computer will prompt "ready the printer and press return to proceed". When "return" is pressed, the file contents will be displayed on the screen and printed as hard copy.

If any mistakes are found in the data listing or if the user wishes to adjust the previously entered data, option 2 of the file maintenance menu should be used. The description given in the original data file will be displayed on the screen followed by the prompt "O.K. (Y/N) ?". If "N" is entered, the computer asks for the "new description" to be entered. If Y or return is pressed, the contents of the original file are shown, sample by sample. If month needs to be edited, enter the no. of the new month. Else, press "return". Editing process is the same as editing data in the "creation of a new data file" stage.

Option "4" can be used to display the disk catalog. File names ending in .BAS and .COM denote program files and system files which need not concern the user. The rest of the files are the data files created by the user.

If at any stage of the operation a "disk full" message appears, the user has to format a new disk for data or delete some of the unnecessary data files from the disk. For deleting purposes, use the option "5" and give the file name to be deleted. Make sure that you want that file erased and enter Y as the answer to the prompt "do you wish to delete this file (Y/N) ?". If you enter a file name ending with .BAS or .COM the computer will reply "Sorry — program files cannot be erased".

Option "6" displays the following message. "If you need to prepare a new blank disk for use as data disks with Bhattacharya's program, you must initialise the disk first. This process is done at CP/M command level (A>) by typing the command COPY B:/D. Press return, then insert blank disk in drive B and press return again. Make sure that the disk you insert to drive B is blank, otherwise the contents of this disk will be erased. Note down this command and press return to quit to CP/M command level."

When the user presses "return", the program will exit to command level, clear the screen and display "A>" on the top of the screen allowing the user to type the COPY command for initializing new disks. If, however, the user wishes not to initialize disks, user can safely turn off the computer at this stage.

Option "7", in the file maintenance menu, is used for regrouping the length frequency data by widening the interval of the size groups. The original size interval of each group may be doubled, tripled, quadrupled etc. by inserting the corresponding multiplying factor of 2, 3, 4, etc. The program regroupes the frequencies accordingly and provides a printout of the regrouped frequency data. It should be noted that only the original length frequencies are saved and stored on the data diskette and not the regrouped data. Therefore, if the regrouped data are to be used in any analysis, they must be re-entered using the printout of the regrouped frequencies with a new file name.

Once file handling is over, use option "8" to enter the main menu to perform analysis.

3. BHATTACHARYA'S METHOD

Bhattacharya's (1967) method was proposed by its author as a graphical method for the separation of normally distributed groups from a mixture of normal distributions.

When the natural logarithm of ratios of successive frequencies is plotted against the mid-points of the corresponding classes, a straight line with a negative slope will be obtained when a group has been identified.

The straight line is of the form

$$\ln (f_{i+1}/f_i) = a + bx_i$$

where b denotes a negative slope.

Using the linear regression technique, a and b are obtained (for sets of 3 points each) using the following equations :

$$b = \frac{\sum xy - \sum x \sum y/n}{\sum x^2 - (\sum x)^2/n}$$

$$a = \sum y/n - b \sum x/n$$

When the absolute correlation coefficient calculated according to the equation

$$r = \frac{\sum xy - \sum x \sum y/n}{\sqrt{(\sum x^2 - (\sum x)^2/n)(\sum y^2 - \sum y)^2/n}}$$

is higher than or equal to the input value of the critical correlation coefficient, the mean (μ), std. deviation (σ) and number of items in the component distributions (N) are calculated using

$$\mu = (0.5d) - (a/b)$$

$$\delta = \sqrt{(1/b) - (d^2/12)}$$

$$N = \frac{\delta}{d} \exp \left[\frac{\sum_{i=1}^3 \ln (f_{i+1}/f_i)}{3} + \frac{\sum_{i=1}^3 (x_i - \mu)^2}{6(\delta^2 + (\delta^2/12))} + \frac{d^2}{24\delta^2} + \ln \sqrt{2\pi} \right]$$

where d is width of the class interval, and f_i , x_i are the values of the frequencies and mid-points of the three points under consideration respectively.

Unlike in the HP 67/97 program of Pauly and Caddy (1985) where 3 frequencies at a time are entered, in this program, any number of frequencies for any number of samples can be entered at once.

In view of overlapping distributions, a certain number of items between two adjacent components may be counted into both components resulting in the sum total of the number of items (N) allocated to the identified components being greater than the original number of items input. A correction for this is made by calculating a raising factor, using the ratio of original number of items input to the calculated total number of items in the components and multiplying the calculated number of items in each component, to give the number of items in each component. The size (after correcting) will appear in the percentage frequency plot program. (Example in Appendix 3.)

3.1 Running instructions

Choose option "2" from the main menu to run Bhattacharya's method. Enter the date. This can be skipped by pressing return but entering the date can be useful, as it will appear on the print-out along with the file name. Next enter the data file name. Return or blank will not be accepted here. If "/" is entered, the main menu will appear.

When the prompt "ready the printer and press return" is displayed, set the printer switch on and press return.

Next, enter a high critical value of correlation coefficient. The program will calculate, display and print the calculated correlation coefficient and the mid-points of each group (consisting of 3 consecutive points) with a negative slope. For cases where calculated correlation coefficient is more than the input value of critical correlation coefficient, the mean, std. deviation and the no. of items in the groups will also be printed.

This calculation and printing will be repeated for all the samples in the data file. Then, the prompt "re-run with different values (Y/N)?" appears on the screen. If "Y" is entered, the user can enter another value for the correlation coefficient and proceed as in the earlier case. If "N" is entered the main menu will appear.

By re-running this program several times with different and lower values of correlation coefficients the user will be able to identify more groups than those identified by higher correlation coefficients.

All these values can then be tested for goodness of fit using the Chi square statistics program.

Note that if the input value of correlation coefficient is too high, genuine distributions might not appear. On the other hand, if it is too low, irrelevant bumps may be misidentified as genuine components.

Using this program, the user may find two or more sets of mean, std. deviation and number of items for the same component because a component distribution which overlaps little with its neighbouring distributions might give rise to more than three points forming a straight line. (Refer to page 2 of Pauly and Caddy (1985).)

When there are several sets of δ , μ and N for mean values which are almost the same or very close to each other, the set of values corresponding to the highest r-value may be chosen, unless the user has reasons for choosing one of the other sets of values.

Zero frequencies will not be accepted by the computer because the logarithm of zero is not defined. Hence, the data will have to be re-grouped with a different interval size so that all the classes will have non-zero data. Alternatively, the user can enter 0.01 as frequency at the data entry stage in the file maintenance routine.

4. CHI SQUARE STATISTICS FOR TESTING RESULTS

For each of the groups identified by Bhattacharya's method, the Chi square statistics can be calculated to assess whether the groups identified generate a mixture of normal distributions that fits the original frequency data.

The standard measure for goodness of fit, the Chi squared criterion, is defined as

$$\chi^2 = \sum (f_i - \hat{f}_i)^2 / \hat{f}_i$$

where the predicted frequencies \hat{f}_i are calculated from the equation

$$\hat{f}_i = \sum_j \frac{N_j d}{\delta_j \sqrt{2\pi}} \cdot \exp \left[-0.5 \left\{ (x_i - \bar{x}_j) / \delta_j \right\}^2 \right]$$

where d = class interval width
 x_i = mid-point of class corresponding to frequency
 x_j = mean of j th component identified
= std. deviation of j th component identified
 N_j = no. of items (size) in j th component identified

and $i < j < 6$.

The predicted frequencies deducted from the raw frequencies provide the residual frequencies i.e. $R = f_i - \hat{f}_i$. The results may be used as input data for Bhattacharya's method, if they suggest that one or more components had been overlooked. For this purpose, when running the Chi square program for the first time, the user can give another file name to be created automatically with the residuals of this analysis. The Chi square values calculated for each x_i value are added up and the sum obtained is compared with the critical value in the χ^2 distribution table (Appendix 3) corresponding to the number of degrees of freedom. Here, the number of degrees of freedom is equal to the number of classes considered minus one, minus twice the number of component distributions used in generating the expected frequency values (j), or deg. of freedom = $n - (1 + 2j)$. (Example in Appendix 4.)

When many groups of normal distribution are identified through this method, the degrees of freedom tend towards zero or even a negative value, and the observed Chi square values also tend to be abnormal. In such situations, the user must use her/his judgement to consider which of these groups should be deleted, keeping in mind the relative correlation coefficients obtained and the possible error in sampling.

4.1 Running instructions

Choose option "3" from main menu to obtain the Chi square statistics program. Enter original data file name containing the frequency data. Return or blank will not be accepted for file name. If "I" is entered, the user can exit to main menu.

Next the question "Do you wish to create another file with the residuals of this analysis (Y/N)" appears on the screen. If Y is entered or return is pressed, enter new file name followed by the description of this new file. If N is entered, the program will not create an additional file with the residuals.

Next set the printer on and press return. The user can select sample by sample and run the Chi square test if necessary. If so, enter Y for the question. "Do you wish to select sample by sample for the Chi square test (Y/N) ?". Then enter the sample number. To quit to main menu, enter zero for sample no. Else, for each sample in the data file, repeat the following procedure.

Enter the number of components to be tested. Note that only up to six sets can be used. Then enter separately the means, standard deviations and sizes of each component. The program will then display and print the values of observed frequencies, predicted frequencies, and

residuals (deviations from prediction). For the Chi square statistics, as a general rule, predicted frequencies less than 5 should be avoided. Therefore, the program will lump classes by adding them with frequencies in contiguous classes to create classes of adequate size and lump the corresponding classes of observed frequencies, then print the set of new residual values along with the Chi square statistics.

Compare this sample χ^2 value with a χ^2 distribution with the appropriate degrees of freedom (given as output by the program). If the sample χ^2 value is greater than the value from the tables (refer to Appendix 2a for χ^2 tables) with the appropriate degrees of freedom the null hypothesis that the sample data are described by the y , S and N values entered is therefore rejected (see "Biometry" by Robert R. Sokal and F. James Rohlf, Chapter 16, page 568 and 569 or any other good statistics text). If there are negative residuals, the program will automatically subtract the largest negative residual from all residuals and then save the resultant positive series of numbers as "frequencies" in a new data file.

Using this new data file, the user can rerun Bhattacharya's method. Note that the number in a component (N) found this way may not be reliable.

5. SOME APPLICATIONS OF BHATTACHARYA'S METHOD TO FISH POPULATION ANALYSIS

When Bhattacharya's method for the analysis of length frequencies distributions is applied to separate possible constituent components and to estimate the mean values and their variances, it may serve primarily to identify age groups, the assumption being that modal groups in the length frequency distribution represent year classes or cohorts.

5.1 Identification of age groups

Petersen's method (1892) for the analysis of length frequency data involves the attribution of assumed ages to the distinct peaks of a single, multi-peaked length frequency sample. There are problems in identifying the real peaks representing age groups, and various authors have suggested methods to separate these peaks on the assumption that they are normally distributed. The Bhattacharya method of separating a mixture of normal distribution is also being increasingly applied, for the same purpose, with the mean length estimated for each modal group and their variances. However, the reliability of results depends very heavily on the unbiased samples of length frequency. In this respect, length frequency samples obtained from catches made by non-selective gears are preferred but if samples are available only from highly size-selective gears, then the data must be adjusted for selectivity before analysis. In many cases, where particular species are caught by more than one type of gear e.g. Indian mackerels caught by purse seines and bottom trawls, the sample must be raised to the catch by each type of gear before adding up their length frequencies. In the case of pelagic species schooling by size, the length frequencies may be different between crafts or vessels operating from the same base and at the same time. In such cases, the length frequency sub-samples from individual boats may have to be raised to their catch before adding them to produce the day's sample.

Item 1 in option "4" of the programs calculates the confidence limits for the mean values (X) which may be considered for allocating the size ranges for the age groups (Appendix 4). The results present the percentage frequency distribution of the sample graphically, the mean values and the 95% and 99% confidence limits for these values. The table of values of the "t" statistic used in the program, for 95% and 99% levels, are presented in Appendix 2 (b).

$$\text{Confidence limits} = \bar{x} \pm t_{n-1} \cdot s/\sqrt{n}$$

There will be difficulties in attributing absolute ages to the modal size groups identified by this method. Even if the age (t) value for the first modal group cannot be determined, the probable age values for the succeeding modal groups can be assigned by using $t+1$, $t+2$, $t+3$ etc. when the intervals between time series data are known to be one month or one year. The growth curve

already determined will help in identifying the modal groups belonging to a cohort, i.e. the same cohort in successive time intervals or even successive cohorts in the same time period.

5.2 Growth parameters

Separation of modal groups in monthly samples of length frequencies for one year series can bring out the monthly modal progression more clearly than the length frequency distributions based on raw data, and growth curves can be constructed with the help of monthly shifts in the mean lengths of the modal groups. Further, by plotting these calculated values of mean length and their variances along with the graphical representation of this frequency distribution, it may be possible to detect modes that may have been incorrectly identified during the separation analysis. A print out of this analysis by the program, for these purposes, is the same as that in 5.1, but will be in series corresponding to the number of months for which the sample data were provided.

5.2.1 Estimating growth parameters

The asymptotic length L_{∞} and growth constant K will be calculated by three different methods using the length at age data and data on length increase in time. This data for Ford-Walford plot, Gulland and Holt plot and von Bertalanffy plot will not be written on the disk and stored as a data file. As the program uses the generalized VBGF, it is better to enter the appropriate value of surface factor D if available. Use $D=1$ if you want to work with the standard von Bertalanffy equation.

Pauly (1984) should be consulted for details and further reference on the surface factor (D). Also note that the three methods given below for estimating L_{∞} and K did not include D in their original form, and that D effectively disappears from the relevant equations when $D=1$.

5.2.2 Running instructions

Enter "4" from the main menu to obtain the following screen:

ESTIMATION OF GROWTH PARAMETERS

1. PLOT OF % FREQUENCIES
2. FORD-WALFORD PLOT
3. GULLAND AND HOLT PLOT
4. VON BERTALANFFY PLOT

5.2.3 Plot of percentage frequencies

Before running this program make sure that Bhattacharya's analysis has been done because this requires the user to enter the mean values, standard deviations and sizes obtained from that analysis.

Enter the name of the data file. If "/" is entered, the user can exit to main menu. Set the printer on and press return. The program will print the percentage frequency distributions for all the samples in the data file. In the same manner as in the Chi square test program for each sample, choose groups with high "r" values from the results of Bhattacharya's analysis and enter the number of groups chosen and their means, standard deviations and sizes. Then the program will print the confidence limits for the mean values for each of the samples, and the sizes after applying the correction involving the raising factor as mentioned in Section 3 above. For plotting the length frequency distributions, the percentage frequencies have been rounded to the nearest integer and these values are also printed along with the percentage frequencies.

5.2.4 Ford- Walford plot

This method introduced by Ford, 1933 and Walford, 1946, is based on the equation

$$L_{t+1}^D = a + b L_t^D$$

where L_t^D and L_{t+1}^D correspond to lengths separated by a constant time interval.

In the case of a single cohort, when one set of lengths (L_t) for a series of ages is entered, the computer will automatically create the consecutive length (L_{t+1}). But for multiple broods, the

user is given the chance to enter both sets of L_t and L_{t+1} values for the different cohorts combined. In this case, however, the L_∞ and K values obtained will represent mean values for the different cohorts.

According to the above equation, both x and y axis are measured with the same error since they are indeed the same data. Hence, a geometric mean (or type II) regression is used to obtain a' and b' . If a and b denote the arithmetic mean regression values then,

$$b' = b/r$$

$$a' = y - b'x$$

$$L_\infty = [a' / (1 - b')]^{1/D}$$

$$K = -\ln b' / D$$

For single broods, enter "S" and press "return". Enter the no. of age classes and the lengths (L_t). The computer will automatically generate the table of L_t and L_{t+1} values.

For multiple broods enter M and press return. Enter the no. of pairs of consecutive lengths and the values of L_t and L_{t+1} .

Next enter the surface factor (D) where $0 < D \leq 1$. In the case of fish capable of reaching large sizes use the appropriate value of D if possible. **If the correct value of D is not known enter 1.** Set the printer ready and press return. The program will print the table of L_t and L_{t+1} and output values of a' , b' , r, L_∞ and K. (Example in Appendix 6.)

52.5 Gulland and Holt plot

This method introduced by Gulland and Holt (1959), is based on the equation:

$$\frac{L_2^D - L_1^D}{t_2 - t_1} = a - K D \left(\frac{L_1^D + L_2^D}{2} \right)$$

where L_1 and L_2 are successive lengths corresponding to times t_1 and t_2 . This gives:

$$K = -b/D$$

$$L_\infty = (a/KD)^{1/D}$$

Here too data will not be saved on the disk and kept as a data file.

Enter the number of pairs of L_1, L_2 and then enter separately $L_1, L_2, t_2 - t_1$ values.

Next enter the surface factor (D), when $0 < D \leq 1$. Set the printer on and press return. The program will print the table of values of $L_1, L_2, t_2 - t_1$ and the coordinates x and y. The intercept (a), slope (b), correlation coefficient (r), L_∞ and K values will be produced.

The next user can rerun the program with different values of D if necessary. (Example in Appendix 7.)

52.6 Von Bertalanffy plot

The generalized Von Bertalanffy growth formula proposed by Von Bertalanffy (1934) has the form

$$L_t^D = L_\infty^D \left(1 - e^{-KD(t-t_0)} \right)$$

which can be rewritten in the form:

$$-\ln [1 - (L_t/L_\infty)^D] = -KD(t_0 - t) + KDt$$

This equation defines a linear regression with t as x axis and

$$-\ln [1 - (L_t/L_\infty)^D] \text{ as y axis.}$$

If an estimate of L_∞ is available then

$$K = b/D$$

$$\text{and } t_0 = -a/b$$

Enter the number of pairs of t and L_t values. Zero will not be accepted here. Then enter the values of t and L_t . Enter D and L_∞ . Set the printer ready and press return. If D is unknown, enter $D=1$.

The table of values of t , L_t and the points of the y axis (x axis is t) will be output. a , b , r , t_0 and K will be calculated and printed. This program can then be rerun with different values of D and L_∞ (Note that the method of Wetherall (see below) can be used to obtain estimate of L_∞)

6. TOTAL MORTALITY ESTIMATION

If the modal groups identified are assumed to represent age groups, then the number of items (N) in the modal groups which can be linked through modal shift, over a specified period of time, may be used to estimate the total mortality rates. For such an analysis with this program, each length frequency should be raised to the total catch of that species, going through the raising process mentioned in 5.1. An estimate of total mortality, as explained in Pauly and Caddy (1985) may be obtained through the formula

$$Z = \frac{\ln(N_t - N_{t+\Delta t})}{\Delta t}$$

where N_t is the number of items in a modal group at first instance

and $N_{t+\Delta t}$ is the number in corresponding modal group after the time interval Δt expressed as a fraction of a year.

A log-linear regression of the numbers in a given component against time provides Z from its slope. N_t , N_{t+1} , N_{t+2} , may also be used to calculate the total mortality rates using the formula

$$Z = - \ln \left[\frac{2N_{t+1} + 2N_{t+2} + \dots}{iN_{t+1}N_{t+1} + \dots} \right]$$

for larger sets of data combined to give a single estimate or

$$Z = - \ln \frac{N_{t+1}}{N_t}$$

for a single group. Alternatively, the data for a long series of years may be analyzed separately for different years and ages to obtain a series of Z values which may exhibit any systematic changes in mortality (Gulland, 1983).

If it is successive cohorts in the same time period, instead of the same cohort in successive time intervals, then it must be assumed that recruitment is not subject to appreciable changes. In this application it is necessary that the numbers refer to fully recruited cohorts.

6.1 Running instructions

When "5" is entered from the main menu, the following screen will be displayed:

```

TOTAL MORTALITY ESTIMATION
1. JONES & VAN ZALINGE METHOD
2. SPARRE'S CATCH CURVE METHOD
3. WETHERALL ET AL METHOD
4. EXIT TO MAIN MENU
ENTER YOUR CHOICE?

```

All these three sub-programmes operate on the data file originally created by the file maintenance program. If a data file has not been created already, enter "4", return to main menu and create a data file preferably with total catch figures entered as frequencies. (Example in Appendix 8.)

6.2 Jones and van Zaiinge method

The inverse Von Bertalanffy equation

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

and the cumulated catch curve equation lead to the Jones and van Zalinge equation (1981) :

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

Here the y axis represents the natural logarithm of the cumulated catch of fish of length L and longer (C). When Z, L_∞ and K are assumed to be constants with age, this equation gives a linear regression with x axis as log (L_∞-L).

$$\text{slope } b = Z/KD$$

Thus total mortality rate Z=b.K.D.

In practice, the relationship between $\ln(L_\infty^D - L^D)$ and $\ln(C(L, L_\infty))$ is linear only in the middle part of the plot. This is because the first few length groups which form the ascending part of the curve consist of fish not yet fully recruited to the fishery. Also the last few age groups, which are close to L_∞, give rise to an uncertain relationship between age (t_L) and length (L). Hence, the regression analysis should be limited to the central, straight portion of the curve.

This program utilizes the data created and saved as a data file under file maintenance routine. Enter the name of the data file. In this case the frequencies should have been raised to the total catch. If "/" is entered, the main menu can be obtained. Enter an estimated value of L_∞, D (or set D =1) and K. Set the printer on and press "return". The table of values of length intervals, catch frequencies, cumulated catch figures, and x and y coordinates will be printed. The screen will also display the scatter diagram for the points to help the user in choosing the points for the regression analysis. However, users are strongly urged to plot the points on graph paper for choosing the points, as the screen display will tend to be distorted. For this reason, the following message appears on the screen. "If you wish to display the above scattergram on the screen again, enter Y or else press return". If return is pressed, the prompt appears: "Please draw a scattergram with the printed values, Then select points belonging to the straight segment of the plot. Enter the starting and ending numbers of the points selected for regression". When the user enters the starting number and the ending number of the straight portion of the graph, the program will execute the regression analysis to produce a series of Z/KD values, initially with the first three selected points and then repeating the analysis after including each of the successive points selected. Then the Z values are calculated using the K and D values input and then printed. Their 95% confidence limits are also automatically calculated and printed, using the variances (S_b²) and the corresponding t_{n-2} values.

$$S_b^2 = \frac{1}{n-2} [(S_y / S_x)^2 - b^2]$$

$$\text{Confidence limit of } Z = \pm Z \text{ KD } t_{n-2} \cdot S_b$$

This procedure will be repeated for all samples in the data file. (Example in Appendix 9.)

6.3 Sparre's (1985) catch curve method based on length composition data

The method of Sparre (1985) is a variant of the method of Jones and van Zalinge (1981) which also starts with the inverse von Bertalanffy equation, which is used to estimate the time taken.

The time taken for an average fish to grow from length L₁ to length L₂ i.e.

$$\Delta t = t(L_2) - t(L_1) = \frac{1}{K} \ln \frac{L_\infty - L_1}{L_\infty - L_2}$$

The age interval mid-point is given by the equation.

$$t \left[\frac{(L_1 + L_2)}{2} \right] = t_0 - \frac{1}{K} \ln \left[1 - \frac{L_1 + L_2}{2L_\infty} \right]$$

Thus based on the equation

$$\ln \left(\frac{C(L_1, L_2)}{\Delta t} \right) = C - Z t_0 - Z \frac{1}{K} \cdot \ln \left[1 - \frac{(L_1 + L_2)}{2} / L_\infty \right]$$

we obtain

$$\ln \left(\frac{C(L_1, L_2)}{\Delta t} \right) = C - Z \left[t \left((L_1 + L_2) / 2 \right) \right]$$

which is of the form, $y=a+bx$ with slope $b=Z$.

The program presents a scatterplot of $t \left((L_1 + L_2) / 2 \right)$ vs. $\ln \left(C(L_1, L_2) / \Delta t \right)$ which is used to identify the straight portion of the catch curve. The calculations and the regression for the selected points are as in the previous method but in this case, the slope $b=Z$. As in the previous case (6.1) some points are excluded from the regression analysis, for the same reasons. The calculation of a series of Z values and their confidence limits and printing of the results, follow the same procedure as in 6.1. (Example in Appendix 10.)

6.4 Wetherall method for estimating growth and mortality parameters from length frequency data

This simple method estimates asymptotic length (L_∞) and the ratio of the coefficients of mortality and growth (Z/K), using only the length frequency data (Wetherall 1986).

The principle is based on Beverton and Holt's formula

$$Z/K = (L_\infty - \bar{L}) / (\bar{L} - L_c)$$

which can be rearranged to

$$\bar{L} = \frac{L_\infty}{1 + Z/K} + L_c \cdot \frac{Z/K}{1 + Z/K}$$

to show that the mean length of the selected fish (\bar{L}) is a linear function of the selection length (L_c). However application of this method assumes the sampled fish population to have

- (a) constant annual recruitment
 - (b) von Bertalanffy growth
 - (c) continuous mortality occurring at a uniform rate,
- which are rarely fully satisfied in practice.

For better results it is expected that the length frequency samples be raised to the annual catch. For a series of arbitrary knife-edge selection lengths (L_c), the corresponding mean lengths (\bar{L}) of size groups exceeding the selection lengths (L_c) are calculated, ignoring the size groups below the selection length. These when graphically plotted, present a linear relationship according to the above equation with

$$\text{intercept } a = \frac{L_\infty}{1 + Z/K}$$

$$\text{slope } b = \frac{Z/K}{1 + Z/K}$$

$$\text{From these } L_\infty = \frac{a}{1 - b}$$

$$\text{and } Z/K = \frac{b}{1 - b}$$

$$\text{and } Z = \left(\frac{b}{1 - b} \right) K$$

(See example 6.4 in Appendix.)

For the knife-edge selection length (L_c) from the length frequency distribution, the lower limits of the size groups have been chosen.

This program utilizes the data created and saved as a data file under file maintenance routine. Enter the name of the data file. If "/" is entered main menu can be obtained. Enter estimated value of K, if available. The program estimates the mean lengths for the series of corresponding arbitrary selection length values identified by the user. Next the values are printed and a scatterplot

appears on the screen, to help the user to identify the lowest value of the selection length, to be used in the linear regression analysis. When the points for the regression analysis are entered, the program will print out estimates of L_{∞} Z/K and also Z if a K value had been entered, initially.

7. . REFERENCES CITED

- Bhattacharya, C.G. (1967)
A simple method of resolution of a distribution into gaussian components. *Biometrics*, 23 : 115-35.
- Gulland, J.A. (1983)
Fish stock assessment. A manual of basic methods, FAO/John Wiley Et Sons. 223 p.
- Pauly, D. (1984)
Fish population dynamics in tropical waters: a manual for use with programmable calculators *ICLARM Studies and Reviews* 8, Manila 325 p.
- Pauly D. and J.F. Caddy (1985)
Method for the analysis of mixtures of normal distributions. *FAO Fish. Circ. No. 787*; 16 p. (+errata).
- Petersen, J. (1892)
Fiskenbiologiske forhold i Holboek Fjord 1890-I 891. *Beretning fra de Danske Biologiske Station for 7980 (91) 1*, 121-183.
- Ng Fong Oon (1986)
Growth and mortality of the Malaysian cockle (*Anadara granosa*. L) under commercial culture. *BOBP/WP/47*, 21 p.
- Sanders, M.J. (1984)
Compilation of Hewlett Packard 97 programs for fishery stock assessment applications, *RAS/83/023/MAN/1*. (FAD) 158 p.
- Sparre, P. (1985)
Introduction to tropical fish stock assessment. FAO/DANIDA Project, training in fish stock assessment GCP/INT/392/DEN. Manual 1. Publ. by FAO. 338 p.
- Wetherall, J.A. (1986)
A new method for estimating growth and mortality parameters from length frequency data : *Fishbyte* 4 (1); 12-14.
- White, J. *et al* (1974)
Tables for Statisticians, Pitman Press, London; 61 p.
- Jones, R. and N. van Zalinge (1981)
Estimation of mortality rate and population size for shrimps in Kuwait waters. *Kuwait Bull. Mar. Sci.* 1981 (2) ; 273-288.
- BOBP (1987)
Hilsa investigations in Bangladesh. *BOBP/REP/36*.