64594 PN-ABE-268

WORKING PAPER SERIES

FISHERIES STOCK ASSESSMENT

TITLE XII
Collaborative Research Support Program

Fisheries Stock Assessment CRSP Management Office
International Programs, College of Agriculture
The University of Maryland, College Park, Maryland 20742

In cooperation with the United States Agency for International Development (Grant No. DAN-4146-G-SS-5071-00) the Fisheries Stock Assessment CRSP involves the following participating institutions:

The University of Maryland—Center for Environmental and Estuarine Studies
The University of Rhode Island—International Center for Marine Resource Development
The University of Washington—Center for Quantitative Sciences
The University of Costa Rica--Centro de Investigación en Ciencias del Mar y Limnología
The University of the Philippines—Marine Science Institute (Diliman)—College of Fisheries (Visayas)

In collaboration with The University of Delaware; The University of Maryland—College of Business and Management; The University of Miami; and The International Center for Living Aquatic Resources Management (ICLARM).

WORKING PAPER SERIES

Working Paper No. 63

SAMPLE SIZE AND GROUPING OF DATA FOR LENGTH FREQUENCY ANALYSIS

by
Karin Erzini

Graduate School of Oceanography
The University of Rhode Island

August 1989

Fisheries Stock Assessment
Title XII
Collaborative Research Support Program

The Fisheries Stock Assessment CRSP (sponsored in part by USAID Grant No. Dan-4146-G-SS-5071-00) is intended to support collaborative research between U.S. and developing countries' universities and institutions on fisheries stock assessment and management strategies.

This Working Paper has been produced by The University of Rhode Island research component in collaboration with The University of the Philippines and in association with the International Center for Living Aquatic Resources Management. Additional copies are also available from:

International Center for Marine Resource Development
The University of Rhode Island
126 Woodeard Hall
Kingston, Rhode Island 02881

Sample size and grouping of data for length frequency analysis.

Abstract

A modified Bhattacharya method was used to analyse simulated length frequency distributions in four different sample sizes (100, 200, 500, 1000) and 20 interval sizes (2, 3,....., 20, and 50 mm). Life history characteristics ranging from a small, shortlived, fast growing, high mortality clupeid or engraulid type fish to a large, longlived, slow growing, relatively low mortality lutjanid or sparid type fish were represented. Results indicated that the best interval size for grouping the data was a function of sample size and biological parameters such as variability of length-at-age, recruitment pattern, growth rate, and maximum size. For the clupeid-engraulid type frequency distributions, samples of 500 to 1000 individuals and interval sizes giving about 6 intervals per constituent age group gave the best results. At the larger end of the size spectrum, sample sizes greater than 1000 were indicated. These results suggested that life history characteristics. particularly variability of length-at-age and recruitment should be taken into account when designing sampling strategies for length-frequency based analyses.

Introduction

Information on demographic parameters of fish and other animal populations can be obtained from length frequency samples by length frequency analysis (LFA). While mean length-at-age and the number of age classes present are the most commonly estimated parameters, LFA is also useful for providing information on recruitment levels and mortality. It is applied extensively in situations where aging by growth rings of hard parts is not possible and where tagging is not a viable alternative. Methods are available for analysing single distributions and/or time series of samples. They include graphical, computerized, parametric and nonparametric methods (Harding 1949, Cassie 1954, Hasselblad 1966, Bhattacharya 1967, Tomlinson 1971, MacDonald and Pitcher 1979, Schnute and Fournier 1980, Akamine 1985, Brey and Pauly 1986, and MacDonald and Green 1986).

The degree of success with any method is dependent in part on certain characteristics of the species and the sample. These include growth rate, mortality rate, recruitment pattern, gear selectivity, sample size and interval width. The life history characteristics determine the relative contribution of each age group to the population and the sample, the distance between the mean lengths-at-age, and the amount of overlapping.

The ability to separate mixtures of distributions is also affected by sample size and interval width, particularly in the case of older age groups where overlapping is most significant and

large individuals may not be adequately represented. When data is grouped, it is assumed that the midpoint of each class can represent the original measurements that fell within the boundaries of the class without significantly affecting subsequent analysis and identification of modes. By not making a distinction between measurements falling in the same interval, information is lost and the larger the interval or class size, the greater the loss. On the other hand, by increasing the number of intervals and decreasing interval size, more effort is required in sampling, homogeneity is decreased and errors in sampling may take on added importance (Guiasu 1986). In fisheries, a compromise often has to be made between measuring a small number of fish slowly and accurately and grouping larger numbers of measurements into wider intervals (Caddy 1986).

The choices of sample and interval sizes are usually up to the researcher and there have been no hard and fast rules to date. Caddy (1986) suggested 300 to 800 individuals per sample. According to MacDonald and Pitcher (1979), each age group should be represented by a minimum of 50 individuals. Grant (1989) found that graphical methods required "large" sample sizes and clearly visible modes, while very large sample sizes (>1000) were necessary with computer based methods if modes were obscured (Grant et al. 1987). The distances between means of components as a function of the standard deviation of the components was an important factor influencing the degree of success in their study.

Interval sizes of 0.5 cm for small species (< 30 cm) and 1.0

or 2.0 cm for larger species are widely used. Wolff (1989) proposed an empirically derived formula based on maximum observed size and estimated number of age classes in the sample for selection of the optimal interval size. Caddy (1986) examined a number of LFA studies and found that interval size should be small enough so that successive peaks were separated by 5 or 6 size class intervals.

In a previous study, simulated samples with known life history parameters were used to investigate the ability of two LFA methods to estimate growth and mortality parameters (Castro and Erzini 1988). The objective of this study was to evaluate the optimal sample size and grouping of data for length frequency analysis for a wide range of life history characteristics. In this context, the optimal grouping is that partition or interval size which results in the most successful separation of component mixtures and estimation of mean lengths-at-age. The success of LFA can be judged on the basis of the correct identification of the number of components or age classes in a distribution and on the precision of the estimated means.

Methods

The program SIMULPOP was used to generate length frequency distributions for different life history and exploitation patterns (Castro 1988). Major features of the simulations include: 1) mean length-at-age is represented by the von Bertalanffy curve, 2) the catch equations model the exponential decline in

numbers of cohorts, 3) length-at-age is normally distributed, 4) natural and fishing mortality are random normal, 5) retention or selectivity is described by the logistic curve, and 6) recruitment is random for specified annual patterns (Castro 1988, Castro and Erzini 1988).

The choice of parameters was based on a review of the literature for 11 species of small, fast growing, short-lived, high mortality species of the Genus *Engraulis*, *Sardina*, *Sardinella*, *Sardinops* and *Sprattus* (Table 1), and 10 species of relatively long-lived, large, slow growing, low mortality snappers of the Genus *Lutjanus* and *Ocyurus* (Table 2) (Polovina and Ralston 1987, Erzini work in progress). In what follows, case 1 refers to the engraulid/clupeid type simulations while the lutjanid/sparid type simulations are case 2. Within each case there are 12 "situations" representing different life history and fishing patterns. Series of monthly length frequency distributions for each of 24 situations were simulated (Table 1,2). All the distributions consisted of 10,000 measurements grouped into 1 mm size classes. Fifty random samples of 100, 200, 500 and 1,000 fish were drawn for each of the 24 situations. These 4800 samples were regrouped into 19 new distributions of 2, 3, ..., 19, 20 mm interval size . The samples for the lutjanid-type fishes were additionally grouped into 50 mm interval size classes.

The LFA method used to analyse these distributions was that of Bhattacharya (1967). Many of the widely used packages such as MIX (MacDonald and Green 1985) or ELEFAN (Brey and Pauly 1986) are quite subjective. However, the Bhattacharya

method can be modified so that its implementation becomes fully automated and the estimation of mean lengths-at-age does not require any interaction with the user (Pauly and Caddy 1985). A program was written to implement this method.

The Bhattacharya method assumes that components of a distribution are normally distributed. The logarithms of the ratios of successive frequencies (log(fi+1/fi)) are plotted against the midpoints of the size classes (Li). The mean lengths corresponding to each age class can be calculated by the regression of log(fi+1/fi) on Li for series of three or more points which define a straight line with a negative slope. The mean is given by

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

where d is the interval size and a and b are the constant and regression coefficient terms respectively. Rather than visual selection of points making up a straight line with a negative slope on a graph, Pauly and Caddy (1986) proposed that the choice of points be based on the correlation coefficients calculated for the regressions of all series of 3 successive points and a critical value of r. Regressions which did not have a negative slope or did not exceed the critical value for r were rejected. For this study, the distributions were adjusted for selectivity and this approach was extended to all series of 3, 4, ..., 15 successive points at the 95% level critical value of r. Estimates of mean length-at-age were judged satisfactory if they were within 2.5% of the actual length-at-age. This small margin of error was used because larger boundaries such as ± 5% tended to overlap for the older age classes of some distributions. Fitting growth curves to mean

lengths-at-age is sensitive to small deviations from actual means. Therefore a small margin is also more realistic.

Results

Each distribution was analysed using the modified Bhattacharya method. The proportion of means successfully identified was recorded and the means for each 50 distributions of the same situation, sample size and interval width are given in Tables 3 and 4. To save space, only results for interval sizes of 2, 5, 10, 15 and 20 mm are shown.

For given interval sizes and life history parameters, more mean lengths-at-age were successfully identified with larger sample sizes. As illustrated in Figure 1 with four samples from situation 1 and an interval width of 5 mm, the smaller the sample, the less clearly defined the modes become. The range also increases with sample size and there are fewer gaps in the data. For a particular sample size, the problems of gaps in the distributions and of poorly defined modes were greater for case 2 situations because of the wider range.

In general, there was a trend towards a smaller optimal interval size with increasing sample size. If only interval sizes 2, 5, 10, 15, and 20 mm are considered, this trend is not readily apparent and the best interval size is often the same over the four sample sizes for each situation (Tables 3 and 4). In fact, the range of best interval sizes for case 1 situations decreased from 4 - 17 mm for n = 100 to 2 - 8 mm for n = 1000. For case 2 the trend

was not as clear. Optimal interval size ranged from 10 - 20 mm for n = 100. However, for n = 1000 optimal sample size ranged from 20 to 4 mm.

Estimation of mean length-at-age was more successful with case 1 situations, with a maximum proportion identified correctly of 0.56 for situation 1 samples (n=1000, interval sizes of 7 and 8 mm). However, on the average it was seldom possible to estimate more than half the mean lengths-at-age for distributions of either case. All other parameters being constant, the interval size giving the best results tended to be greater for situations with greater variability of length-at-age (the even numbered situations in Tables 3 and 4). This difference was particularly pronounced in case 2 situations.

Of the three recruitment patterns, the Bhattacharya analysis identified a greater proportion of the mean lengths-at-age for single spawning peak conditions. This is in agreement with the findings of Castro and Erzini (1988). The difference between bimodal and continuous recruitment patterns in terms of these results was not clear. Within the two cases, greater success was achieved for simulations with the slower growth rate (K=0.3 versus 0.6, K=0.1 versus 0.2).

Following Wolff (1989) but substituting the range for the maximum observed length, the parameter D for the following formula was estimated for all situations and sample sizes:

Optimal interval size = (D . Range) / Number of age groups

For case 1, the mean value of D was 0.20 (SD = 0.062) and for case 2 it was 0.29 (SD = 0.084).

The mean number of intervals per mode for the optimal interval size increased with sample size. For case 1, the mean numbers for n = 100, 200, 500, and 1000 were 4.4, 5.2, 6.1, 6.1. The corresponding means for case 2 were 3.6, 3.7, 3.5, 4.5.

Discussion

Several scientists have examined the reliability of various length frequency methods and their limitations using simulated distributions (Grant et al. 1987, Hampton and Majkowski 1987, Castro and Erzini 1989). These studies have demonstrated the difficulty of obtaining reliable age and growth information from length data. In most fish length frequency studies, the fish were measured to the nearest 1 cm, 2 cm or very rarely to the nearest 0.5 cm or 0.25cm. Use of these measurement intervals is often a compromise between precision and time and effort required in sampling. In this study, simulations representing a variety of life history and exploitation patterns were analysed by a modified Bhattacharya method using a wide range of sample and interval sizes.The results have a number of implications for LFA.

It has been shown that optimal interval size is determined largely by sample size and biological characteristics such as length-at-age variability, recruitment pattern, growth rate and maximum size. For example, if all other factors are equal, greater length-at-age variability requires the use of a larger interval for

grouping the data. Therefore, a simple empirically based method for determining the most suitable interval size for grouping length frequency data, such as that of Wolff (1989), may be useful for providing rough estimates. However, for certain species and sample sizes, the interval size may be quite wrong. Clearly, biological parameters other than maximum size, range and estimated number of age groups must be taken into consideration when designing a sampling scheme for length frequency studies.

The Caddy (1986) suggestion that the interval size chosen should be that which allows 6-7 intervals per mode for sample sizes of 300 - 800 was supported by some of the results. For case 1, the mean number of intervals per mode corresponding to the optimal grouping sizes levelled out at 6.1 for sample sizes of 500 and 1000. Improvements with increased sample size for case 1 suggest that sample sizes greater than 1000 would not result in a significantly larger proportion of mean lengths-at-age estimated correctly. It indicates that this may be a useful rule to guide the selection of interval sizes for LFA. For case 2 situations, the improvement in results with an increase in sample size from 500 to 1000 suggests that it may be worthwhile to use samples of more than 1000 individuals. The mean number of intervals per mode for n = 1000 was only 4.5, but with the observed trend of decreased optimal interval size with increased sample size, it is highly probable that this value would increase with a sample size greater than 1000.

This implies that in practice more effort and time may be required to sample larger numbers of fish with increased

precision. The problems of effort and precision in sampling may be overcome by new developments for measuring and recording data (e.g. Trent et al. 1987, Bayley and Illyes 1988). However, the problem of obtaining samples which are large enough so that all age classes are adequately represented and there are no significant gaps in the distribution remains.

Contrary to usual expectations, it was found that identification of modes for large, long lived, slow growing species can be accomplished with as much success as for short lived, fast growing species if sample size is large and the data is grouped into appropriate classes. Successful length frequency analysis to separate mixtures of distributions may be more dependent on recruitment pattern, length-at-age variability and sample size than on growth rate, maximum size and maximum age.

This study was based on distributions which were corrected for gear selectivity. This is not realistic in terms of usual fisheries data but the purpose of this work was to examine the effect of size class and sample size on estimation of mean length-at-age using one particular method of LFA under ideal conditions. The criteria used to judge successful estimation of mean lengths-at-age were deliberately strict and undoubtedly resulted in a lower proportion of means identified than would have been by some other methods of LFA. Nevertheless, the conclusions of this study can be extended to real life data and to other length frequency methods. In summary, length frequency analysis can be improved by taking into account biological characteristics of the species when selecting sample size and grouping interval.

Acknowledgements

I would like to thank Margarida Castro, Saul Saila and Conrad Recksiek for their constructive criticism and help with this work. This work was sponsored in part by USAID Grant No. DAN-4146-G-SS-5071-00 (Fisheries Stock Assessment CRSP).

References

Akamine, T. 1985. Consideration of the BASIC programs to analyse the polymodal frequency distributions into normal distributions. Bulletin of the Japanese Sea Fisheries Research Laboratory, 35:129-160.

Bayley, P.A. and R.F. Illyes. 1988. Measuring fish and recording data in the field and laboratory: a microcomputer application. Fisheries, 13:15-18.

Bhattacharya, C.G. 1967. A simple method of resolution of a distribution into Gaussian components. Biometrics, 23:115-135.

Brey, T. and D. Pauly. 1986. Electronic Length Frequency Analysis. A revised and expanded users guide.Berichte aus dem Institut fur Meereskunde an der Christian-Albrechts-Universitat Kiel Nr. 149, 76p.

Caddy, J.F. 1986. Size frequency analysis in stock assessment - some perspectives, approaches and problems. Proceedings of the 37 th Annual Gulf and Caribbean Fisheries Institute,

pp.212-238.

Cassie, R.M. 1954. Some uses of probability paper in the analysis of size frequency distributions. Australian Journal of Marine and Freshwater Research, 5:513-522.

Castro, M. 1988. User's manual for SIMULPOP, a program to simulate series of length frequency distributions of an exploited fish stock. Fisheries Stock Assessment Collaborative Research Support Program Working Paper No. 44, 44p.

Castro, M. and K. Erzini. 1988. Comparison of two length frequency based packages for estimating growth and mortality parameters using simulated samples with varying recruitment patterns. Fisheries Bulletin, 86:645-653

Grant, A. 1989. The use of graphical methods to estimate demographic parameters. Journal of the Marine Biological Association of the United Kingdom, 69:367-371.

Grant, A., P.J. Morgan and P.J.W. Olive. 1987. Use made in marine ecology of methods for estimating demographic parameters from size/frequency data. Marine Biology, 95:201-208.

Guiasu, S. 1986. Grouping data by using the weighted entropy. Journal of Statistical Planning and Inference, 15:63-69.

Hampton, J. and J. Majkowski. 1987. An examination of the reliability of the ELEFAN computer program for length-based stock assessment. pp 203-216 In: D. Pauly and G.R. Morgan (eds.) Length-based methods in fisheries research. ICLARM Conference Proceedings 13.

Harding, J.P. 1949. The use of probability paper for the graphical

analysis of polymodal frequency distributions. Journal of the Marine Biological Association of the United Kingdom, 28:141-153.

Hasselblad, V. 1966. Estimation of parameters for a mixture of normal distributions. Technometrics, 8:431-444.

MacDonald, P.D.M. and P.E.J. Green. 1986. User's guide to program MIX. An interactive program for fitting mixtures of distributions. Release 2.2, July 1985. Ichthus Data Systems, Ontario, Canada, 28p.

MacDonald, P.D.M. and T.J. Pitcher. 1979. Age groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. Journal of the Fisheries Research Board Canada, 36:987-1001.

Pauly, D. and J.F. Caddy. 1985. A modification of Bhattacharya's method for the analysis of mixtures of distributions. FAO Fisheries Circular No. 781, 16pp.

Polovina, J.J. and S. Ralston (eds.) 1987. Tropical snappers and groupers: biology and fisheries management. Westview Press, 659pp.

Schnute, J. and D. Fournier. 1980. A new approach to length-frequency analysis: Growth structure. Canadian Journal of Fisheries and Aquatic Sciences, 37:1337-1351.

Trent, L., B.J. Palko, and A. Trimble. 1987. Improved method of measuring lengths of large samples of fish. The Progressive Fish-Culturist, 49:310-312.

Tomlinson, P.K. 1971. NORMSEP: normal distribution separation, in N.J. Abramson (ed.), Computer Programs for Fish Stock

Assessment. FAO Fisheries Technical Paper, 101.

Wolff, M. 1989. A proposed method for standardization of the selection of class intervals for length frequency analysis. Fishbyte, 7:5.

TABLE 1.

Parameters for case 1 simulations (Engraulid/clupeid type species: short lived, fast growth and relatively high mortality.

SITUATION	K	L_∞	SD length-at-age (mm) *	Recruitment Pattern **	M	SD	F	SD
1	0.3	200	10	1	0.5	0.025	0.5	0.025
2	0.3	200	20	1	0.5	0.025	0.5	0.025
3	0.3	200	10	2	0.5	0.025	0.5	0.025
4	0.3	200	20	2	0.5	0.025	0.5	0.025
5	0.3	200	10	3	0.5	0.025	0.5	0.025
6	0.3	200	20	3	0.5	0.025	0.5	0.025
7	0.6	200	10	1	0.6	0.030	0.6	0.030
8	0.6	200	20	1	0.6	0.030	0.6	0.030
9	0.6	200	10	2	0.6	0.030	0.6	0.030
10	0.6	200	20	2	0.6	0.030	0.6	0.030
11	0.6	200	10	3	0.6	0.030	0.6	0.030
12	0.6	200	20	3	0.6	0.030	0.6	0.030

For all situations:		
	Oldest age class in fishery	4 years
	Age when recruited to fishery	6 months
	Mesh size (stretched)	17.5 mm
	Selectior L(25%)	70 mm
	L(50%)	80 mm
	L(75%)	90 mm

* SD of length-at-age constant for all ages

** Recruitment patterns : 1-one peak per year, 2-bimodal, 3-continuous

TABLE 2.

Parameters for case 2 simulations (Lutjanid/sparid type species: long lived, slow growth, and relatively low mortality.

SITUATION	K	L∞	SD length-at-age (mm) *	Recruitment Pattern **	M	SD	F	SD
1	0.1	700	15	1	0.3	0.015	0.3	0.015
2	0.1	700	25	1	0.3	0.015	0.3	0.015
3	0.1	700	15	2	0.3	0.015	0.3	0.015
4	0.1	700	25	2	0.3	0.015	0.3	0.015
5	0.1	700	15	3	0.3	0.015	0.3	0.015
6	0.1	700	25	3	0.3	0.015	0.3	0.015
7	0.2	700	15	1	0.4	0.020	0.4	0.020
8	0.2	700	25	1	0.4	0.020	0.4	0.020
9	0.2	700	15	2	0.4	0.020	0.4	0.020
10	0.2	700	25	2	0.4	0.020	0.4	0.020
11	0.2	700	15	3	0.4	0.020	0.4	0.020
12	0.2	700	25	3	0.4	0.020	0.4	0.020

For all situations:		
	Oldest age class in fishery	8 years
	Age when recruited to fishery	6 months
	Mesh size (stretched)	60.0 mm
	Selectior L(25%)	170 mm
	L(50%)	210 mm
	L(75%)	250 mm

* SD of length-at-age constant for all ages

** Recruitment patterns : 1-one peak per year, 2-bimodal, 3-continuous

TABLE 3.
Mean proportion of modes successfully identified for each case 1 situation for 4 sample sizes and 5 interval sizes.
n = 50 for each mean and standard deviation.

SAMPLE SIZE	INTERVAL (mm)	SITUATION 1	2	3	4	5	6	7	8	9	10	11	12
100	20	0.01 (0.05)	0.13 (0.13)	0.01 (0.05)	0.08 (0.12)	0.02 (0.07)	0.16 (0.13)	0.12 (0.13)	0.10 (0.12)	0.00 (0.00)	0.04 (0.09)	0.05 (0.10)	0.17 (0.13)
	15	0.04 (0.09)	0.13 (0.13)	0.05 (0.10)	0.13 (0.15)	0.08 (0.13)	0.17 (0.12)	0.17 (0.14)	0.14 (0.13)	0.03 (0.08)	0.07 (0.11)	0.11 (0.13)	0.19 (0.11)
	10	0.13 (0.16)	0.15 (0.13)	0.04 (0.10)	0.15 (0.16)	0.13 (0.13)	0.18 (0.15)	0.22 (0.10)	0.15 (0.14)	0.07 (0.12)	0.09 (0.12)	0.17 (0.12)	0.17 (0.16)
	5	0.17 (0.16)	0.03 (0.08)	0.07 (0.12)	0.06 (0.13)	0.16 (0.13)	0.05 (0.11)	0.24 (0.10)	0.09 (0.12)	0.15 (0.14)	0.05 (0.13)	0.16 (0.15)	0.08 (0.12)
	2	0.08 (0.13)	0.04 (0.09)	0.03 (0.08)	0.05 (0.11)	0.03 (0.08)	0.06 (0.12)	0.09 (0.14)	0.06 (0.11)	0.04 (0.09)	0.00 (0.00)	0.04 (0.09)	0.05 (0.12)
200	20	0.00 (0.00)	0.19 (0.10)	0.01 (0.05)	0.14 (0.14)	0.02 (0.08)	0.24 (0.05)	0.14 (0.13)	0.10 (0.13)	0.03 (0.08)	0.04 (0.09)	0.07 (0.12)	0.19 (0.13)
	15	0.08 (0.13)	0.25 (0.13)	0.11 (0.13)	0.20 (0.13)	0.09 (0.12)	0.24 (0.07)	0.19 (0.11)	0.17 (0.12)	0.03 (0.08)	0.07 (0.12)	0.16 (0.13)	0.27 (0.13)
	10	0.25 (0.19)	0.28 (0.15)	0.19 (0.20)	0.26 (0.14)	0.23 (0.16)	0.28 (0.14)	0.25 (0.05)	0.23 (0.16)	0.14 (0.14)	0.10 (0.12)	0.28 (0.10)	0.24 (0.11)
	5	0.32 (0.14)	0.05 (0.12)	0.19 (0.17)	0.10 (0.14)	0.28 (0.13)	0.12 (0.15)	0.29 (0.11)	0.19 (0.17)	0.23 (0.17)	0.09 (0.13)	0.28 (0.11)	0.12 (0.15)
	2	0.17 (0.18)	0.14 (0.17)	0.12 (0.14)	0.12 (0.16)	0.08 (0.15)	0.10 (0.15)	0.19 (0.14)	0.05 (0.10)	0.14 (0.16)	0.06 (0.12)	0.09 (0.13)	0.14 (0.22)
500	20	0.00 (0.00)	0.25 (0.04)	0.05 (0.12)	0.17 (0.14)	0.03 (0.09)	0.23 (0.06)	0.20 (0.10)	0.15 (0.14)	0.02 (0.07)	0.09 (0.12)	0.02 (0.07)	0.22 (0.12)
	15	0.12 (0.15)	0.26 (0.09)	0.09 (0.14)	0.22 (0.12)	0.17 (0.19)	0.28 (0.08)	0.23 (0.10)	0.25 (0.09)	0.03 (0.08)	0.13 (0.14)	0.17 (0.13)	0.28 (0.10)
	10	0.36 (0.18)	0.31 (0.11)	0.29 (0.18)	0.28 (0.15)	0.34 (0.20)	0.32 (0.12)	0.31 (0.11)	0.28 (0.13)	0.19 (0.18)	0.17 (0.13)	0.31 (0.11)	0.35 (0.16)
	5	0.43 (0.18)	0.20 (0.21)	0.22 (0.15)	0.17 (0.19)	0.35 (0.13)	0.16 (0.17)	0.41 (0.13)	0.30 (0.15)	0.33 (0.16)	0.19 (0.15)	0.32 (0.11)	0.14 (0.16)
	2	0.34 (0.18)	0.18 (0.18)	0.19 (0.19)	0.14 (0.17)	0.17 (0.18)	0.16 (0.19)	0.35 (0.19)	0.19 (0.17)	0.27 (0.21)	0.12 (0.14)	0.24 (0.18)	0.19 (0.21)
1000	20	0.00 (0.00)	0.28 (0.09)	0.09 (0.12)	0.12 (0.14)	0.03 (0.08)	0.26 (0.06)	0.22 (0.08)	0.19 (0.12)	0.01 (0.06)	0.14 (0.13)	0.01 (0.05)	0.25 (0.09)
	15	0.11 (0.13)	0.32 (0.12)	0.10 (0.15)	0.25 (0.13)	0.25 (0.19)	0.27 (0.09)	0.22 (0.09)	0.29 (0.11)	0.03 (0.08)	0.19 (0.11)	0.19 (0.11)	0.34 (0.16)
	10	0.42 (0.18)	0.34 (0.15)	0.31 (0.18)	0.35 (0.17)	0.48 (0.19)	0.34 (0.14)	0.32 (0.12)	0.37 (0.15)	0.28 (0.16)	0.25 (0.11)	0.30 (0.10)	0.43 (0.15)
	5	0.52 (0.21)	0.31 (0.20)	0.30 (0.15)	0.32 (0.19)	0.37 (0.14)	0.41 (0.21)	0.51 (0.15)	0.31 (0.13)	0.36 (0.16)	0.31 (0.15)	0.40 (0.13)	0.40 (0.25)
	2	0.43 (0.16)	0.14 (0.18)	0.31 (0.21)	0.16 (0.17)	0.28 (0.18)	0.14 (0.17)	0.40 (0.17)	0.23 (0.24)	0.41 (0.19)	0.14 (0.16)	0.28 (0.20)	0.13 (0.22)

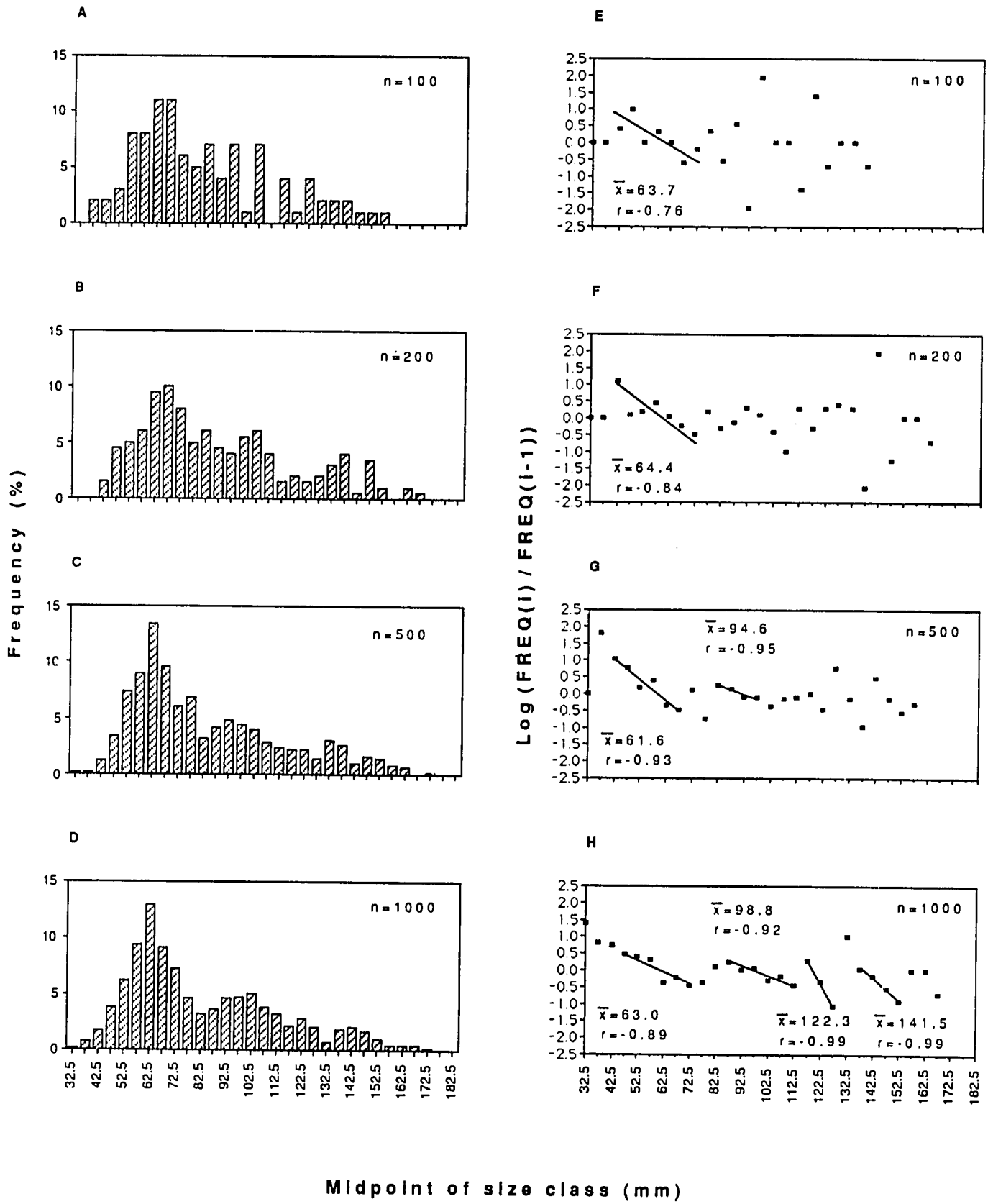
TABLE 4.
Mean proportion of modes successfully identified for each case 2 situation for 4 sample sizes and 5 interval sizes.
n = 50 for each mean and standard deviation.

SAMPLE SIZE	INTERVAL (mm)	SITUATION 1	2	3	4	5	6	7	8	9	10	11	12
100	20	0.03 (0.06)	0.08 (0.08)	0.09 (0.08)	0.13 (0.12)	0.07 (0.11)	0.12 (0.10)	0.09 (0.07)	0.09 (0.08)	0.04 (0.08)	0.09 (0.07)	0.04 (0.06)	0.06 (0.09)
	15	0.07 (0.10)	0.10 (0.09)	0.12 (0.08)	0.11 (0.12)	0.12 (0.10)	0.08 (0.09)	0.14 (0.08)	0.10 (0.09)	0.05 (0.08)	0.06 (0.07)	0.03 (0.07)	0.02 (0.06)
	10	0.10 (0.11)	0.06 (0.10)	0.04 (0.07)	0.09 (0.11)	0.11 (0.09)	0.05 (0.08)	0.13 (0.06)	0.08 (0.09)	0.06 (0.08)	0.06 (0.08)	0.03 (0.07)	0.03 (0.06)
	5	0.07 (0.08)	0.02 (0.07)	0.04 (0.07)	0.04 (0.08)	0.07 (0.09)	0.04 (0.09)	0.10 (0.07)	0.03 (0.06)	0.04 (0.06)	0.02 (0.05)	0.03 (0.07)	0.02 (0.05)
	2	0.04 (0.07)	0.01 (0.07)	0.03 (0.07)	0.04 (0.06)	0.03 (0.06)	0.03 (0.06)	0.03 (0.06)	0.01 (0.06)	0.03 (0.06)	0.01 (0.03)	0.01 (0.04)	0.00 (0.00)
200	20	0.10 (0.14)	0.19 (0.11)	0.15 (0.06)	0.24 (0.12)	0.17 (0.14)	0.21 (0.10)	0.16 (0.07)	0.15 (0.10)	0.09 (0.08)	0.13 (0.07)	0.09 (0.10)	0.12 (0.11)
	15	0.14 (0.13)	0.18 (0.11)	0.16 (0.08)	0.2 (0.13)	0.27 (0.12)	0.19 (0.10)	0.17 (0.07)	0.17 (0.11)	0.12 (0.11)	0.13 (0.07)	0.07 (0.08)	0.09 (0.09)
	10	0.15 (0.11)	0.14 (0.10)	0.11 (0.12)	0.11 (0.11)	0.24 (0.11)	0.10 (0.11)	0.18 (0.06)	0.14 (0.10)	0.15 (0.09)	0.11 (0.08)	0.06 (0.0[illegible])	0.08 (0.09)
	5	0.14 (0.14)	0.05 (0.09)	0.07 (0.10)	0.08 (0.10)	0.17 (0.09)	0.10 (0.11)	0.15 (0.05)	0.08 (0.10)	0.07 (0.09)	0.02 (0.05)	0.06 (0.09)	0.07 (0.10)
	2	0.05 (0.08)	0.05 (0.09)	0.10 (0.10)	0.07 (0.09)	0.06 (0.08)	0.07 (0.10)	0.06 (0.07)	0.07 (0.08)	0.03 (0.06)	0.02 (0.05)	0.05 (0.07)	0.05 (0.08)
500	20	0.24 (0.15)	0.35 (0.16)	0.23 (0.10)	0.36 (0.12)	0.30 (0.12)	0.27 (0.08)	0.21 (0.09)	0.27 (0.09)	0.19 (0.12)	0.26 (0.11)	0.20 (0.10)	0.20 (0.09)
	15	0.31 (0.14)	0.31 (0.11)	0.24 (0.10)	0.32 (0.11)	0.39 (0.13)	0.27 (0.10)	0.24 (0.09)	0.23 (0.10)	0.27 (0.12)	0.22 (0.08)	0.16 (0.14)	0.14 (0.12)
	10	0.32 (0.13)	0.23 (0.11)	0.31 (0.13)	0.27 (0.13)	0.38 (0.11)	0.20 (0.13)	0.25 (0.11)	0.23 (0.10)	0.24 (0.11)	0.2 (0.08)	0.11 (0.11)	0.10 (0.11)
	5	0.33 (0.15)	0.14 (0.13)	0.13 (0.13)	0.07 (0.09)	0.27 (0.09)	0.14 (0.12)	0.23 (0.09)	0.16 (0.11)	0.23 (0.11)	0.06 (0.08)	0.08 (0.11)	0.09 (0.11)
	2	0.15 (0.13)	0.16 (0.15)	0.15 (0.11)	0.16 (0.14)	0.13 (0.11)	0.19 (0.15)	0.08 (0.10)	0.12 (0.11)	0.14 (0.11)	0.07 (0.08)	0.16 (0.14)	0.13 (0.12)
1000	20	0.32 (0.13)	0.46 (0.16)	0.25 (0.09)	0.42 (0.08)	0.33 (0.11)	0.33 (0.10)	0.29 (0.13)	0.35 (0.10)	0.26 (0.12)	0.33 (0.12)	0.23 (0.09)	0.23 (0.10)
	15	0.35 (0.12)	0.38 (0.14)	0.31 (0.09)	0.38 (0.12)	0.46 (0.12)	0.34 (0.11)	0.36 (0.13)	0.32 (0.12)	0.38 (0.12)	0.28 (0.07)	0.27 (0.13)	0.19 (0.09)
	10	0.37 (0.11)	0.25 (0.12)	0.32 (0.09)	0.38 (0.11)	0.49 (0.12)	0.25 (0.14)	0.35 (0.12)	0.30 (0.11)	0.32 (0.11)	0.26 (0.11)	0.19 (0.15)	0.18 (0.11)
	5	0.42 (0.12)	0.19 (0.15)	0.25 (0.16)	0.17 (0.13)	0.35 (0.10)	0.15 (0.13)	0.36 (0.12)	0.24 (0.11)	0.33 (0.13)	0.07 (0.10)	0.13 (0.11)	0.12 (0.11)
	2	0.24 (0.15)	0.28 (0.16)	0.21 (0.14)	0.16 (0.14)	0.21 (0.14)	0.19 (0.16)	0.19 (0.14)	0.16 (0.12)	0.17 (0.12)	0.10 (0.11)	0.21 (0.12)	0.19 (0.13)

Figure captions

Fig. 1. Examples of 4 samples (n = 100, 200, 500, and 1000) for case 1, situation 1 (Table 1), with the transformed data and the correctly estimated mean lengths-at-age. A critical level of P = 0.05 was used and the interval size for grouping was 5 mm. The actual mean lengths-at-age of the original simulated "population" from which these samples were drawn were 63, 97, 124, and 144 mm.

Figure 1


A
n=100
B
n=200
C
n=500
D
n=1000
E
n=100
x̄=63.7
r=-0.76
F
n=200
x̄=64.4
r=-0.84
G
n=500
x̄=94.6
r=-0.95
x̄=61.6
r=-0.93
H
n=1000
x̄=98.8
r=-0.92
x̄=63.0
r=-0.89
x̄=122.3
r=-0.99
x̄=141.5
r=-0.99
Frequency (%)
Log (FREQ(I) / FREQ(I-1))
Midpoint of size class (mm)

Appendix

The program **BHATTREG.BAS** is written in **BASIC**. All series of 3 to 32 consecutive frequencies on a length frequency distribution are analysed. The significance of the correlation coefficients can be tested at the P = 0.05 or the P = 0.01 critical levels. Regressions of the log of the ratios of successive frequencies on the midpoints of the corresponding size classes are carried out. Negative correlation coefficients are tested at the specified critical levels. If the correlation coefficient is significantly different from 0, the regression coefficient and constant is used to estimate the mean and the standard deviation of the normal distribution.

Program inputs are the number of size classes (NC), the midpoint of the first size class (MP), the class width or interval size (CW), and the frequencies. Data can be input while running the program or read from a file. If data is in a file, it should be in the following format (Example: Figure 1 B, n = 200):

26, 32.5, 5
3, 9, 10, 12, 19, 20, 16, 10, 12, 9, 8, 11, 12, 8, 3, 4, 3, 4, 6, 8, 1, 7, 2, 0, 2, 1

In other words, the first three numbers in the file must be NC, MP, CW, followed by the frequencies.

The user has a choice of P = 0.01 or P=0.05, and the data can be saved.

Output

The output consists of the file name, the selected critical level, and five columns corresponding to all the significant regressions:

Reg	The number of points in the regression.
Index	The index for the first point in the regression (1 to NC).
Mean	The estimated mean.
SD	The standard deviation.
r	The correlation coefficient.

The following are outputs for the above example:

TEST1

P = 0.05

Reg	Index	Mean	SD	r
3	5	65.90	4.14	-0.9995
3	12	100.94	2.70	-0.9990
4	4	66.76	3.77	-0.9925
4	11	99.71	3.05	-0.9842
7	1	64.38	4.84	-0.8397
9	1	70.02	6.25	-0.7192
10	1	72.33	6.87	-0.6937
14	1	77.50	7.95	-0.6567
15	1	81.74	9.42	-0.5172
16	1	82.50	9.69	-0.5320

TEST1

P = 0.01

Reg	Index	Mean	SD	r
4	4	66.76	3.77	-0.9925

Note: Different series of points will give different estimates of mean length at age for the same mode. Users must use their own judgement in deciding which is the best estimate. For this example, the actual mean lengths at age for the original simulated "population" from which this sample was drawn were: 63, 97, 124, and 144 mm. As we can see from these results, a significant correlation does not necessarily correspond to a correctly identified mean length at age.

A listing of the program is provided. Copies of the program **BHATTREG.BAS** and **SIMULPOP.BAS** can be optained by sending a 3.5 inch or 5.25 inch diskette to the author.

```
1000 REM                          Program   BHATTREG.BAS
1010 REM
1020 REM                          K. ERZINI  4-28-89
1030 REM
1040 REM  Modified Bhattacharya Analysis
1050 REM  References: Bhattacharya, C.G. 1967. Biometrics
                       23:115-135.
1060 REM              Pauly, D. and J.F. Caddy. 1985. FAO
1070 REM              Fish. Circular No. 781, 16p.
1080 REM
1090 OPTION BASE 1
1100 DIM L(1000), F(1000), LF(1000), LFLF(1000), R01(100),
         R05(100)
1110 REM L( ) : midpoints of classes
1120 REM F( ) : frequencies
1130 REM LF() : log(freq i+1) - log(freq)
1140 REM N( ) : number of partitions
1150 REM   Critical values P = 0.05  (Rohlf, F.J. and R.R.
           Sokal. 1981)
1160 FOR I = 3 TO 32
1170     READ R05(I)
1180     R05(I) = R05(I) * -1
1190 NEXT I
1200 '
1210 REM   Critical values P = 0.01
1220 FOR I = 3 TO 32
1230     READ R01(I)
1240     R01(I) = R01(I) * -1
1250 NEXT I
1260 CLS : PRINT : PRINT : PRINT
1270 INPUT "Input new data   Y/N "; D$
1280 IF D$ = "" THEN 1270
1290 IF D$ = "Y" OR D$ = "y" THEN 1300 ELSE 1670
1300 '
1310 PRINT : PRINT : PRINT
1320 INPUT "Number of Classes        "; NC
1330 INPUT "MidPoint of first class  "; MP
1340 INPUT "Class Width              "; CW
1350 '
1360 FOR I = 1 TO NC
1370     PRINT "Frequency  "; : PRINT I; : INPUT " "; F(I)
1380 NEXT I
1390 PRINT : PRINT
1400 L(1) = MP
1410 FOR I = 2 TO NC
1420     L(I) = L(I - 1) + CW
1430 NEXT I
1440 FOR I = 1 TO NC
1450     IF F(I) = 0 THEN 1460 ELSE 1470
1460     LF(I) = -9999: GOTO 1480
1470     LF(I) = LOG(F(I))
1480 NEXT I
1490 FOR I = 1 TO NC - 1
1500     IF LF(I) = -9999 THEN 1510 ELSE 1520
```

```
1510      LFLF(I) = -9999: GOTO 1550
1520      IF LF(I + 1) = -9999 THEN 1530 ELSE 1540
1530      LFLF(I) = -9999: GOTO 1550
1540      LFLF(I) = LF(I + 1) - LF(I)
1550 NEXT I
1560 LFLF(NC) = 0
1570 PRINT : PRINT : PRINT
1580 IF FLAG = 1 THEN 1770
1590 INPUT "File name to save data  "; F$
1600 OPEN F$ FOR OUTPUT AS #1
1610      PRINT #1, NC, MP, CW
1620      FOR I = 1 TO NC
1630          PRINT #1, F(I)
1640      NEXT I
1650 CLOSE #1
1660 GOTO 1770
1670 INPUT "Data file name "; F$
1680 OPEN F$ FOR INPUT AS #2
1690      INPUT #2, NC, MP, CW
1700      FOR I = 1 TO NC
1710          INPUT #2, F(I)
1720      NEXT I
1730 CLOSE #2
1740 FLAG = 1
1750 GOTO 1400
1760 REM
1770 REM ----------------Calculations----------------------
1780 REM
1790 LPRINT F$: LPRINT
1800 PRINT "Select probability level:"
1810 INPUT "1. P = 0.05  2. P = 0.01"; ALPHA
1820 IF ALPHA = 1 THEN 1830 ELSE 1840
1830 LPRINT " P = 0.05 ": GOTO 1850
1840 LPRINT " P = 0.01 "
1850 LPRINT
1860 LPRINT "Reg Index      Mean            SD            r "
1870 LPRINT "-----------------------------------------------"
1880 LPRINT : LPRINT
1890     FOR N = 3 TO 32
1900     DF = N - 1
1910     IF N > NC THEN 2520
1920     FOR J = 1 TO NC - N
1930         SUMX = 0
1940         SUMY = 0
1950         SUMXSQ = 0
1960         SUMYSQ = 0
1970         SUMXY = 0
1980         SSXY = 0
1990         SSXX = 0
2000         SSYY = 0
2010         R = 0
2020         B1 = 0
2030         B0 = 0
2040         MEAN = 0
```

```
2050          SD = 0
2060          FOR K = J TO J + DF
2070              IF LF(K) = -9999 THEN 2360
2080              SUMX = SUMX + L(K)
2090              SUMY = SUMY + LFLF(K)
2100              SUMXSQ = SUMXSQ + L(K) ^ 2
2110              SUMYSQ = SUMYSQ + LFLF(K) ^ 2
2120              SUMXY = SUMXY + (L(K) * LFLF(K))
2130          NEXT K
2140          REM  calculate correlation coefficient
2150          SSXY = SUMXY - ((SUMX * SUMY) / N)
2160          SSXX = SUMXSQ - ((SUMX ^ 2) / N)
2170          SSYY = SUMYSQ - ((SUMY ^ 2) / N)
2180          IF SSYY = 0 THEN 2360
2190          IF SSXY = 0 THEN 2360
2200          R = SSXY / (SQR(SSXX * SSYY))
2210          IF R < 0 THEN 2220 ELSE 2360
2220          IF ALPHA = 1 THEN 2230 ELSE 2240
2230          IF R <= R05(N) THEN 2250 ELSE 2360
2240          IF R <= R01(N) THEN 2250 ELSE 2360
2250          REM  fit regression    y = B0 + B1*x
2260          B1 = SSXY / SSXX   ' regression coefficient
2270          B0 = (SUMY / N) - (B1 * (SUMX / N)   ' constant
2280          REM  calculate mean of component
2290          MEAN = (.5 * CW) - (B0 / B1)
2300          IF ABS(1 / B1) <= (CW ^ 2) / 12 THEN 2320
2310          SD = SQR((ABS(1 / B1)) - ((CW ^ 2) / 12))
2320          REM  print results
2330          LPRINT N; J, : LPRINT USING "###.##"; MEAN,
2340          LPRINT , : LPRINT USING "###.##"; SD,
2350          LPRINT , : LPRINT USING "##.####"; R
2360      NEXT J
2370      NEXT N
2380 REM    Critical values for correlation coefficients
2390 REM    P = 0.05
2400 DATA   0.997, 0.950, 0.878, 0.811, 0.754, 0.707
2410 DATA   0.666, 0.632, 0.602, 0.576, 0.553, 0.532
2420 DATA   0.514, 0.497, 0.482, 0.468, 0.456, 0.444
2430 DATA   0.433, 0.423, 0.413, 0.404, 0.396, 0.388
2440 DATA   0.381, 0.374, 0.367, 0.361, 0.355, 0.349
2450 REM    P = 0.01
2460 DATA   1.000, 0.990, 0.959, 0.917, 0.874, 0.834
2470 DATA   0.798, 0.765, 0.735, 0.708, 0.684, 0.661
2480 DATA   0.641, 0.623, 0.606, 0.590, 0.575, 0.561
2490 DATA   0.549, 0.535, 0.526, 0.515, 0.505, 0.496
2500 DATA   0.487, 0.478, 0.470, 0.463, 0.456, 0.449
2510 '
2520 END
```