BAY OF BENGAL PROGRAMME — BOBP/MAG/3

Marine Fishery Resources Management — RAS/81/051

FISHERY STATISTICS ON THE MICROCOMPUTER

— A BASIC version of Hasselblad's NORMSEP program

by D. Pauly
Associate Scientist, ICLARM

N. David
Formerly ICLARM Affiliate Scientist

J. Hertel-Wulff
Fishery Biologist (Associate Professional Officer), Bay of Bengal Programme

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, June 1986.
Mailing Address : C/o. FAO, P.O. Box 1505, Colombo 7, Sri Lanka.
Street Address : NARA Building, Crow Island, Mattukuliya, Colombo 15.
Telex : 2203 A/B FAOR CE   Tel : 522380, 522381, 522382, 522383.

INTRODUCTION

Like most other biologists, fishery biologists often encounter situations where mixtures of distributions need to be separated into their component distributions.

Such situations arise when, for example, the counts of different-sized ova in a fish gonad have to be separated into batches that will be spawned successively, when length-frequency data have to be separated into age groups or when "pulses" of recruitment are to be quantified (Fig. 1).

There exist a number of graphical and mathematical techniques which can be used for such separation (see Everitt and Hand, 1981, for a review). Most of them assume that the distributions to be separated from the mixture are *normal* distributions (Pearson 1984, Harding 1949, Cassie 1954, Tanaka 1962, Hasselblad 1966, Bhattacharya 1967, Yong and Skillman 1971, MacDonald and Pitcher 1978). Of these, only Bhattacharya's method, as adapted in Pauly and Caddy (1984), and the ENORMSEP program of Yong and Skillman, do not require the external input of the expected number of component distributions in the mixture. The NORMSEP program, of which a BASIC version is presented here, also does require such an input. Thus, although NORMSEP is not the most sophisticated tool available to analyse size-frequency distributions, it is useful in situations where the data to be separated do not greatly overlap, or where previous knowledge exists on the number of probable positions of the means of the component distributions (see Mathew 1974).

The version of NORMSEP presented here is a translation of the FORTRAN IV version of NORMSEP in Abrahamson (1971), written by V Hasselblad and modified by P K Tomlinson.

The following changes were made to the original version :

(i) A routine was added which automatically computes, using the (guessed) means entered, a very large range of possible values for the standard deviation corresponding to each mean, by using a very small (0.01) and a very large (0.50) coefficient of variation (C.V. = s.d. $\sqrt{x}$) to generate the range. This follows Abrahamson's suggestion that "anyone desiring to maintain the rigour of Hasselblad's original procedure only needs to place extremely wide bounds on the standard deviations".

(ii) A routine was added which internally computes the number and approximate position of the cut-off points.

(iii) The input routine has been made completely interactive, with prompts provided by the program, along with routines for checking and correcting the data entered from the keyboard.

(iv) A routine was added along with prompts, to load length frequency data from a disk into memory, and use them as an input to NORMSEP. Such files must have been created by ELEFAN (see David *et al* 1982). No routine for checking the correctness of file inputs has been included.

The maximum number of iterations, and of component distributions that can be separated, have been set at 125 and 10 respectively; note that the program cannot be used to estimate the parameters (mean, standard deviation) of a single normal distribution. Users should heed Abrahamson's advice, issued in conjunction with the original publication of NORMSEP, that "under any circumstances, Hasselblad (1966) and Cohen (1966) should be reviewed before proceeding."

AN APPLICATION EXAMPLE

Table 1 : Length-frequency data for testing NORMSEP.

Note number and position of modes (underlined)

Lower class limit	Frequency	Lower class limit	Frequency
0	2	10	5
1	23	11	2
2	22	12	3
3	1	13	6
4	37	14	6
5	28	15	15
6	1	16	6
7	6	17	2
8	17	18	2
9	16	–	–

Table 1 gives a set of frequencies generated (by Everitt and Hand 1981, Table 2.7) from a mixture of components of known characteristics. From the overall appearance of the data, we shall assume that there are four components, with means close to the modes, i.e., 1.5, 4.5, 8.5 and 15.5.

Question : Can our BASIC version of NORMSEP recover the means and standard deviation originally used to generate the data, if these modes are used as initial guesses of the means?

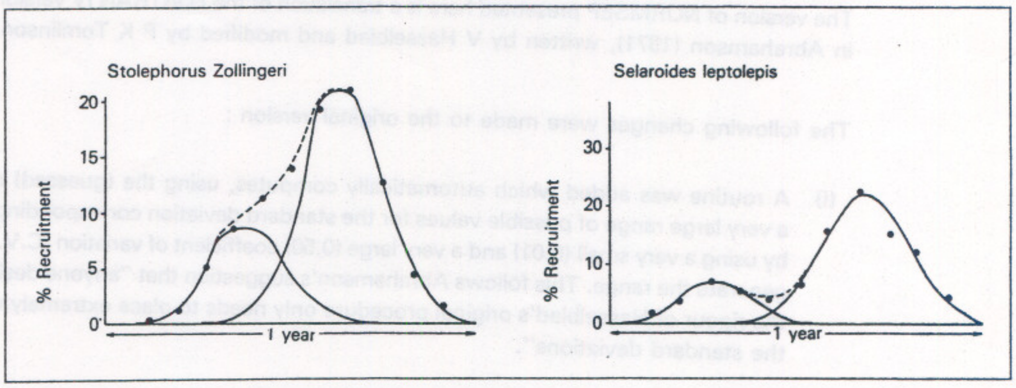

Fig. 1 Recruitment patterns in two stocks of Philippine fishes, derived from length-frequency data projected backwards onto the time axis and used to separate the two pulses of recruitment occuring within a year. The separation of the recruitment patterns into their component distribution was done using NORMSEP; note that the chi-square statistic cannot be used here because recruitment patterns are expressed in % (after Pauly and Navaluna 1984).

The output obtained for a run of our BASIC NORMSEP and the inputs mentioned above are reproduced on page 3. The mean and standard deviations are close to the values used by Everitt and Hand (1981) to generate the data in Table 1.

Component	*Mean*	*Standard deviation*
1	2.0	0.5
2	5.0	0.5
3	9.0	1.0
4	15.0	2.0

The only exception is the standard deviation of the 4th component. These results are exactly the same as those obtained by Everitt and Hand (1981). The interpretation of this and other results obtained by NORMSEP should be undertaken carefully. It is essential, among other things that the user realizes that NORMSEP will always "find" the mean and standard deviation of a distribution if one is assumed to be present, and the preliminary estimate of its mean entered.

RESULTS FOR FILE : TEST

The NORMSEP output, based on data in Table 1 (see text for interpretation), is given below.

Values after 21 iterations

Group	Mean	St dev	Per cent	Size
1	1.94045	.59151	23.7234	47.4468
2	4.92919	.524968	32.9876	65.9751
3	8.96506	.961204	22.6216	45.2432
4	15.0991	1.59989	20.6674	41.3348

Total sample size 200

Actual vs. Predicted Frequencies

Actual	Predicted	Actual	Predicted
2	2.6536	16	15.622
23	22.9724	5	5.78087
22	20.091	2	1.61373
1	4.24912	3	2.8621
37	34.0085	6	6.25408
28	28.1286	6	9.4763
1	2.24172	15	9.83824
6	6.21149	6	6.99787
17	16.146	2	3.40968
		2	1.44271

Degrees of Freedom = 4

Chi Square Value = 8.18143 Prob = .0851538

Log of likelihood = 218.802

The "goodness of fit" statistics (the chi-square) may be used to distinguish various alternative interpretations. However, it is essential that this be done after the basic features of the chi-square test have been assimilated (consult a good statistics textbook!).

RUNNING THE PROGRAM

The corresponding prompts of and input to the program are given below.
Using data from a disk file :

Prompt No.	*Prompt*	*Input*
1	ENTER RUN — ID ?	TEST
2	DO YOU WANT TO READ FILE FROM DISK (Y/N). THE ALTERNATIVE IS TO ENTER DATA FROM KEYBOARD ?	Y
3	ENTER FILE NAME ?	B : TEST DATA
4	ENTER NUMBER OF GROUPS YOU WANT TO SPLIT SAMPLE INTO ?	4
5	ENTER ESTIMATE OF MEAN No. 1 ?	1.5
	ENTER ESTIMATE OF MEAN No. 2 ?	4.5
	ENTER ESTIMATE OF MEAN No. 3 ?	8.5
	ENTER ESTIMATE OF MEAN No. 4 ?	15.5
6	PRINT RESULTS (Y/N) ?	Y
7	PRESS (RETURN)	(RETURN)
8	PRESS (RETURN)	(RETURN)
9	CONTINUE WITH NEXT SAMPLE (Y/N) ?	Y

Using data input from the keyboard :

Prompt No.	Prompt	Input
1	ENTER RUN — ID?	TEST
2	DO YOU WANT TO READ FILE FROM DISK (Y/N). THE ALTERNATIVE IS TO ENTER DATA FROM KEYBOARD?	N
3	ENTER NUMBER OF (LENGTH) CLASSES	19
4	ENTER NUMBER OF GROUPS YOU WANT TO SPLIT SAMPLE INTO?	4
5	ENTER ESTIMATE OF MEAN No. 1?	1.5
	ENTER ESTIMATE OF MEAN No. 2?	4.5
	ENTER ESTIMATE OF MEAN No. 3?	8.5
	ENTER ESTIMATE OF MEAN No. 4?	15.5
6	ENTER LOWER LIMIT OF SMALLEST CLASS?	0
7	ENTER CLASS WIDTH?	1
8	ENTER FREQUENCY : 1?	2
	ENTER FREQUENCY : 2?	23
	... enter frequencies from Table 1 ..	
9	ANY CORRECTIONS (Y/N)?	N
10	REVIEW FREQUENCIES (Y/N)?	N
11	PRINT RESULTS (Y/N)?	N
	... continue as 7 above ...	

The prompts given by the program are shown above. Error messages for keyboard are not shown.

Data files created or extracted by ELEFAN for analysis by NORMSEP must consist of only one sample (i.e., a set of corresponding midlength and frequencies from only one week, month or year etc). For example, a set of monthly samples collected during a year (12 samples) must therefore be subdivided using ELEFAN into 12 sub-files.

Acknowledgement

We are most thankful to Mr. H. Larssen, Danish Institute for Fisheries and Marine Research, Denmark for his assistance with testing the program and his suggestions towards improving the routines.

REFERENCES

Abrahamson, N.J. 1971. Computer progams for fish stock assessment. FAO Fish. Tech. Pap. 101, 148p.

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

Cassie, R. M. 1954. Some uses of probability paper in the analysis of size-frequency distribution. *Aust. J. Mar. Freshw. Res.* 5:513-522.

Cohen, A.C. 1966. Estimation of parameters for a mixture of normal distribution by V.Hasselblad. *Technometrics* 8(3):445-446.

David, N., Palomares L. and Pauly D. (1982). ELEFAN O, a BASIC program for creating and editing files for use with the ELEFAN I, II and III programs. ICLARM. 16. p. (mimeo).

Everitt, B.S. and D.J. Hand. 1981. *Finite mixture distributions.* Chapman and Hall, London. 143 p.

Harding, J.F. 1949. The use of probability paper for the graphical analysis of polymodal frequency distribution. *J. Mar. Biol. Assoc. U.K.* 28:141-153.

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

Lien, D.A. 1981. *The BASIC handbook encyclopedia of the BASIC computer language* (2nd ed.) Compusoft Publishing, San Diego.

Macdonald, P.D.M. and T.J. Pitcher. 1978. Age-groups from size-frequency data : a versatile and efficient method of analyzing distribution mixture. *J.Fish. Res. Board Can.* 36:987-1001.

Mathew, C.P. 1974. An account of some methods of overcoming errors in ageing tropical and sub-tropical fish populations when the hard tissue growth markings are unreliable and the data sparse, p. 158-166. *In* T.B. Bagenal (ed.) *Ageing of Fish.* Unwin Brothers, Old Woking. 234p.

Pauly, D. and J.F. Caddy. 1985. A modification of Bhattacharya's method for the analysis of mixture of normal distributions. *FAO Fish. Circ. No. 781,* Rome, 16p (+ Errata Sheet).

Pauly, D. and N.A. Navaluna. 1984. Monsoon-induced seasonality in the recruitment of Philippine fishes p. 822 — 833. *In* G.D. Sharp and J.Csirke (eds.) *Proceedings of the expert consultation to examine changes in abundance and species composition of neritic fish resources.* San Jose, Costa Rica, 18 — 29 April 1983. F.A.O. Fish Rep. No. 191, Vol 3.

Pearson, K. 1894. Contribution to the mathematical theory of evolution. *Phil. Trans. Roy. Soc.* London. 185:71-110.

Tanaka, S. 1962. A method of analyzing polymodal frequency distributions and its application to the length distribution of the porgy, *Taius tumifrons* (T. and S.). *J. Fish. Res. Board Can.* 19 : 1143-1159.

Yong, M.Y. and R.A. Skillman. 1971. A computer program for analysis of polymodal frequency distributions (ENORMSEP), FORTRAN IV, *U.S. Fish Bull.* 73(3):681.

PROGRAM LISTING

The listing given below (pages 6-13) is the version of BASIC suitable for Apple IIe with CP/M operating system, which is a dialect of Microsoft BASIC. It should therefore be easily converted for use of Microcomputers other than Apple IIe using Microsoft BASIC. The BASIC translation guide by Lien (1981) should help in such conversions, should any problem occur.

Our version of NORMSEP does not require that a printer be connected to the computer (although this will be helpful). About 12K of program memory must be available for the program and the data; these memory requirements can be slightly reduced by deleting all REMs. In case a printer is connected, the interface card must be configured for the CP/M system to enable hard copy.

```
10    REM ================================================================
20    REM
30    REM
40    REM              ********    NORMSEP PROGRAM    ********
50    REM
60    REM
70    REM ================================================================
80    REM THIS IS A BASIC LANGUAGE VERSION OF PROGRAM TCPAI
90    REM (FORMALLY NORMSTEP) WHICH IS AVAILABLE IN THE FORTRAN LANGUAGE.
100   REM
110   REM IT WAS TRANSLATED INTO RADIO SHACK'S BASIC II FROM THE ORIGINAL BY
120   REM N. DAVID AND IT'S INPUT ROUTINE MODIFIED AND EXPANDED BY D. PAULY
121   REM (BOTH ICLARM).
130   REM
140   REM THE VERSION FOR APPLE IIe (WITH CP/M-80 OPERATING SYSTEM) WAS PREPARED
150   REM BY J. HERTEL-WULFF, BOBP.
160   REM
170   REM IT USES THE STEEPEST DESCENT METHOD AS NORMAL
180   REM DISTRIBUTION SEPARATOR.
190   REM
200   REM THIS PROGRAM CONSISTS OF A MAIN ROUTINE AND SEVERAL
210   REM SUBROUTINES:
220   REM         CHISQ
230   REM         NORM
240   REM         TRUNC
250   REM         GTRUN
260   REM         VPLOT
270   REM
280   REM APPLE IIe VERSION, FEB. 1986.
290   REM
300   REM =========================================================
310   CLEAR 200
320     DIM Q(120,10),QQ(120),P(20),MU(20),SG(20),F(120),FS(120),YQ(10)
330   DIM NF(120),NS(20),XT$(8),YT$(50),PP(20),MP(20),SP(20),NX(20),NG(120)
340 DIM BU(20),BL(20),AU(20),AL(20),FX(20),XA$(30),XH$(30)
350 HOME:PRINT TAB(25);"NORMSEP":PRINT
360   CO=.39894228#
370   INPUT"ENTER RUN-ID ";M1$
380 PRINT"DO YOU WANT TO READ FILE FROM DISK (Y/N). THE"
390 INPUT"ALTERNATIVE IS TO ENTER DATA FROM KEYBOARD ";S$
400 IF S$ <> "Y" AND S$ <> "N" THEN GOTO 380
410 IF S$="N" THEN GOTO 430
420 GOSUB 4590:IF S$="Y" THEN GOTO 440
430   INPUT "ENTER NUMBER OF (LENGTH) CLASSES";N
440   INPUT"ENTER NUMBER OF GROUPS YOU WANT TO SPLIT SAMPLE INTO";K
450   LC=1:REM NUMBER OF SETS OF CUT-OFF POINTS SET TO 1 SINCE THE CUT-OFF POINTS
 ARE COMPUTED INTERNALLY...
460   FOR I=1 TO K
470 PRINT"ENTER ESTIMATE OF MEAN NO.";I;:INPUT YQ(I)
480   NEXT I
490   FOR I=1 TO K-1
500   IF YQ(I)>=YQ(I+1) THEN HOME:PRINT "ERROR!  ENTER MEAN VALUES FROM LOWEST TO
 HIGHEST":GOTO 460
510   NEXT I:IF S$="Y" THEN GOTO 700
520   IF K>10 THEN K=10:REM MAXIMUM DISTRIBUTIONS IS 10
530   INPUT "ENTER LOWER LIMIT OF SMALLEST CLASS";XI
540   INPUT "ENTER CLASS WIDTH";XL
550   FOR I=1 TO N
560   PRINT"ENTER FREQUENCY:";I;:INPUT F(I)
570   NEXT I
580 HOME:PRINT TAB(10);"LOOK FOR ERRORS"
590   FOR I=1 TO N
600   PRINT "FREQUENCY NO.";I;F(I)
610   IF I=14 THEN INPUT "PRESS <RETURN>":P$
620   NEXT I
630   FOR I=1 TO N
640   INPUT "ANY CORRECTIONS (Y/N)";P$
650   IF P$="N" THEN GOTO 680
660   INPUT"ENTER FREQUENCY NO. TO BE CORRECTED";J:PRINT"FREQ NO: ";J;:INPUT F(J)
670 NEXT I
```

```
680  INPUT "REVIEW FREQUENCIES (Y/N)";P$
690  IF P$="Y" THEN GOTO 580
700  SF=0
710  FOR I=1 TO N
720  SF=SF+F(I)
730  NF(I)=FIX(F(I)+.5)
740  NEXT I
750  REM
760  REM
770 HOME:PRINT TAB(20);"NORMSEP RUNNING. WAIT"
780 REM  ======================
790  FOR LY=1 TO LC
800 REM ======================
810 REM  GET CUT-OFF POINTS FX'S
820  FOR J=1 TO K-1
830     FX(J)=(YQ(J)+YQ(J+1))/2
840  NEXT J
850  FX(K)=(N-1)*XL+XI
860  REM
870  REM GET LOWER AND UPPER BOUNDS FOR MEANS, BL AND BU'S
880  BL(1)=YQ(1)-.5*(YQ(2)-YQ(1))
890  FOR J=1 TO K-1
900     BU(J)=YQ(J)+.5*(YQ(J+1)-YQ(J))
910     BL(J+1)=BU(J)
920  NEXT J
930  BU(K)=(N-1)*XL+XI
940  REM GET LOWER AND UPPER BOUNDS FOR STANDARD DEVIATION, AL & AU'S
950  FOR J=1 TO K
960     AL(J)=.01*YQ(J)
970     AU(J)=.5*YQ(J)
980  NEXT J
990  REM
1000  FOR J=1 TO K
1010     NX(J)=FIX(((FX(J)-XI)/XL)+.5)
1020      BL(J)=(BL(J)-XI)/XL
1030      BU(J)=(BU(J)-XI)/XL
1040      AL(J)=AL(J)/XL
1050      AU(J)=AU(J)/XL
1060  NEXT J
1070  REM
1080  NX(K)=N
1090  REM
1100  FOR I=1 TO N
1110      FS(I)=F(I)
1120  NEXT I
1130 NB=1
1140 REM COMPUTE INITIAL ESTIMATES USING HALD'S PROCEDURE
1150 AS=0
1160 REM
1170 FOR J=1 TO K
1180    NE=NX(J)
1190    XB=NE
1200     A=0
1210     B=0
1220     C=0
1230 FOR I=NB TO NE
1240    CI=XB-I+.5
1250    CF=FS(I)
1260     A=A+CF
1270     B=B+CF*CI
1280     C=C+CF*CI*CI
1290 NEXT I
1300 REM
1310 ZZ=A*C/(2*B*B)
1320  X=ZZ:GOSUB 4350 :Z=TR: REM CALL TRUNC(ZZ)
1330 REM
1340 X=Z:GOSUB 3800 :G=GT: REM CALL GTRUN(Z)
1350 REM
1360 SG(J)=B*G/A
1370 IF SG(J) > AU(J) THEN SG(J)=AU(J)
```

```
1380 IF SG(J) < AL(J) THEN SG(J)=AL(J)
1390  AV=-Z*SG(J)
1400  IF AV > BU(J) THEN AV=BU(J)
1410  IF AV < BL(J) THEN AV=BL(J)
1420  IF ZZ >= .55 THEN GOTO 1490
1430  AV=B/A
1440  IF AV > BU(J) THEN AV=BU(J)
1450  IF AV < BL(J) THEN AV=BL(J)
1460  SG(J)=SQR ((C-B*B/A)/A)
1470  IF SG(J) > AU(J) THEN SG(J) = AU(J)
1480  IF SG(J) < AL(J) THEN SG(J) = AL(J)
1490  REM
1500  X=AV/SG(J):GOSUB 4080 :PK=NR:REM CALL NORM (AV/SG(J))
1510  REM
1520  P(J)=A/PK
1530  AS=AS+P(J)
1540  FOR KK=1 TO N
1550      II=KK+NE
1560      IF II>N THEN GOTO 1670
1570      UM=(AV+KK)/SG(J)
1580      BM=(AV+KK-1)/SG(J)
1590       X=UM:GOSUB 4080 :KP=NR:REM CALL NORM (ULIM)
1600       X=BM:GOSUB 4080 :KP=KP-NR:REM CALL NORM (BM)
1610      KP=A*KP/PK
1620      REM SUBTRACT ESTIMATED JTH SAMPLE FROM REMAINING ONES
1630      FS(II)=FS(II)-KP
1640      IF FS(II) <0 THEN FS(II)=0
1650      IF KP < .05 THEN GOTO 1670
1660  NEXT KK
1670  REM CONTINUE
1680  NB=NE+1
1690  MU(J)=NX(J)-AV
1700  IF MU(J) > BU(J) THEN MU(J)=BU(J)
1710  IF MU(J) < BL(J) THEN MU(J)=BL(J)
1720  NEXT J
1730  FOR J=1 TO K
1740   P(J)=P(J)/AS
1750  NEXT J
1760  L1=0
1770  REM COMPUTE LIKELIHOOD FUNCTION L1
1780  FOR I=1 TO N
1790     QQ(I)=0
1800     FOR J=1 TO K
1810         Q(I,J)=CO*EXP(-(I-MU(J)-.5)^2/(2!*SG(J)*SG(J)))/SG(J)
1820        QQ(I)=QQ(I)+Q(I,J)*P(J)
1830  NEXT J
1840  L1=L1+F(I)*LOG(QQ(I))/LOG(10)
1850  NEXT I
1860  REM START OF METHOD OF STEEPEST DESCENT
1870  FOR IT=1 TO 125
1880      L3=IT
1890      PS=0
1900      FOR J=1 TO K
1910          XD=0
1920          SX=0
1930          S3=0
1940          FOR I=1 TO N
1950              TE=F(I)*Q(I,J)/QQ(I)
1960              XD=XD+TE
1970              CI=I-.5
1980              CF=CI-MU(J)
1990              SX=SX+TE*CI
2000              S3=S3+TE*CF*CF
2010  NEXT I
2020  MP(J)=SX/XD
2030  IF MP(J) < BL(J) THEN MP(J)=BL(J)
2040  IF MP(J) > BU(J) THEN MP(J)=BU(J)
2050  SP(J)=SQR(S3/XD)
2060  IF SP(J) > AU(J) THEN SP(J)=AU(J)
2070  IF SP(J) < AL(J) THEN SP(J)=AL(J)
```

```
2080  PP(J)=XD*P(J)
2090  PS=PS+PP(J)
2100  NEXT J
2110  FOR J=1 TO K
2120      PP(J)=PP(J)/PS
2130  NEXT J
2140  LL=0
2150  FOR I=1 TO N
2160      QQ(I)=0
2170      FOR J=1 TO K
2180          Q(I,J)=CO*EXP(-(I-MP(J)-.5)^2/(2!*SP(J)*SP(J)))/SP(J)
2190          QQ(I)=QQ(I)+Q(I,J)*PP(J)
2200  NEXT J
2210  LL=LL+F(I)*LOG(QQ(I))/LOG(10)
2220  NEXT I
2230  REM TEST FOR IMPROVED LIKELIHOOD FUNCTION LL
2240  IF LL<=L1 THEN GOTO 2320
2250  FOR J=1TO K
2260      P(J)=PP(J)
2270      MU(J)=MP(J)
2280      SG(J)=SP(J)
2290  NEXT J
2300  L1=LL
2310  NEXT IT
2320  REM
2330  L3=L3-1
2340  FOR I=1 TO N
2350      QQ(I)=0
2360      FOR J=1 TO K
2370          X=(I-MU(J))/SG(J):GOSUB 4080 :Q(I,J)=NR
2380          X=(I-MU(J)-1!)/SG(J):GOSUB 4080 :Q(I,J)=Q(I,J)-NR
2390         QQ(I)=QQ(I)+Q(I,J)*P(J)
2400  NEXT J
2410  FS(I)=QQ(I)*SF
2420  NEXT I
2430  X=-MU(1)/SG(1):GOSUB 4080 :QQ(1)=QQ(1)+NR*P(1)
2440  X=(MU(K)-N)/SG(K):GOSUB 4080 :QQ(N)=QQ(N)+NR*P(K)
2450  FS(1)=QQ(1)*SF
2460  FS(N)=QQ(N)*SF
2470  NN=MU(1)+1!
2480  PR=0
2490  REM COLLAPSE CHI-SQUARE TABLE
2500  FOR I=1 TO NN
2510     II=I
2520     PR=PR+FS(I)
2530 IF PR-5! < 0 THEN GOTO 2540 ELSE GOTO 2550
2540  NEXT I
2550  NL=II
2560  NN=N-MU(K)+1
2570  PS=0
2580  FOR I=1 TO NN
2590     NC=N-I+1
2600     PS=PS+FS(NC)
2610     IF PS-5! < 0 THEN GOTO 2620 ELSE GOTO 2630
2620  NEXT I
2630  NT=NC
2640  AF=0
2650  FOR I=1 TO NL
2660  AF=AF+F(I)
2670  NEXT I
2680  CH=(AF-PR)^2/PR
2690  NH=NT-1
2700  L2=NL+1
2710  FOR I=L2 TO NH
2720     CH=CH+(F(I)-FS(I))^2/FS(I)
2730  NEXT I
2740  AF = 0
2750  FOR I=NT TO N
2760     AF=AF+F(I)
2770  NEXT I
```

```
2780  CH=CH+(AF-PS)^2/PS
2790  ND=NT-NL-3*K+1
2800  REM ===========================================================
2810  REM
2820  REM                           *****   ALARM   *****
2830  REM
2840   FOR I=1 TO 5
2850   BEEP 15,10:BEEP 15,10:BEEP 15,10: FOR J=1 TO 200:NEXT J
2860   NEXT I
2870  REM
2880  REM ===========================================================
2890 HOME: INPUT"PRINT RESULTS (Y/N)";P$
2900  PRINT "RESULTS FOR RUN:";M1$
2910  IF P$="Y" THEN LPRINT" ":LPRINT"RESULTS FOR RUN: ";M1$:LPRINT"------------
-----"
2920 PRINT"VALUES AFTER";L3;" ITERATIONS":
2930  IF P$="Y" THEN LPRINT"VALUES AFTER";L3;" ITERATIONS"
2940  PRINT"GROUP        MEAN        ST DEV      PERCENT     SIZE"
2950 IF P$="Y" THEN LPRINT"GROUP        MEAN        ST DEV      PERCENT     SIZE"
2960  FOR J=1 TO K
2970     SO=SG(J)*XL
2980  XV=MU(J)*XL+XI
2990  PZ=P(J)*SF
3000  PS=P(J)*100!
3010  PRINT J;TAB(11);XV;TAB(21);SO;TAB(32);PS;TAB(44);PZ
3020  IF P$="Y" THEN LPRINT J;TAB(11);XV;TAB(21);SO;TAB(32);PS;TAB(44);PZ
3030  NEXT J
3040  NS%=FIX(SF+.5)
3050  PRINT"TOTAL SAMPLE SIZE";NS%
3060  IF P$="Y" THEN LPRINT" ":LPRINT "TOTAL SAMPLE SIZE";NS%
3070  INPUT"PRESS <RETURN>";S$
3080  PRINT"ACTUAL VS. PREDICTED FREQUENCIES"
3090  IF P$="Y" THEN LPRINT" ":LPRINT"ACTUAL VS. PREDICTED FREQUENCIES"
3100  PRINT"ACTUAL";TAB(12);"PREDICTED"
3110  IF P$="Y" THEN LPRINT"ACTUAL";TAB(12);"PREDICTED"
3120  NZ%=(N-1)/10+1
3130  FOR I=1 TO NZ%
3140    Q1%=I*10 -9
3150    Q2%=N
3160    IF Q2% > I*10 THEN Q2%=I*10
3170    FOR J=Q1% TO Q2%
3180        PRINT F(J);TAB(12);FS(J)
3190        IF P$="Y" THEN LPRINT F(J);TAB(12);FS(J)
3200  NEXT J
3210  REM
3220  NEXT I
3230  INPUT"PRESS <RETURN>";S$
3240  PRINT"DEGREES OF FREEDOM =";ND
3250  IF P$="Y" THEN LPRINT " ":LPRINT"DEGREES OF FREEDOM =";ND
3260  X1=CH:N1=ND:GOSUB 3390 :S=CH#
3270  PRINT"CHI SQUARE VALUE =";CH;" PROB. =";S
3280  IF P$="Y" THEN LPRINT"CHI SQUARE VALUE =";CH;" PROB. =";S
3290  PRINT"LOG OF LIKELIHOOD =";L1
3300  IF P$="Y" THEN LPRINT"LOG OF LIKELIHOOD =";L1
3310  REM================
3320  NEXT LY
3330  REM================
3340  INPUT"CONTINUE WITH NEXT SAMPLE (Y/N)";S$
3350  IF S$="Y" THEN GOTO 310
3360 RUN "ELEFANO.BAS"
3370  REM
3380  REM ........................................
3390  REM ========================
3400  REM ROUTINE CHISQ(X,N)
3410  REM
3420  REM EXACT FORMULA - SHARE DISTRIBUTION NO.528
3430  REM
3440  REM INPUTS ......  X1
3450  REM        ......  N1
3460  REM OUTPUT ......  CH#
```

```
3470  REM
3480  SU=0
3490  IF X1>0 AND N1>0 THEN GOTO 3520
3500  CH#=1!
3510  RETURN
3520  REM
3530  IF FIX(N1/2)*2 <> N1 THEN GOTO 3650
3540  IF N1 <= 2 THEN GOTO 3620
3550  REM
3560  TE#=1
3570  FOR IO=1 TO FIX(N1/2-1)
3580     CI#=2*IO
3590     TE#=TE#*X1/CI#
3600      SU=SU+TE#
3610  NEXT IO
3620  CH#=EXP(-X1/2!)*(1!+SU)
3630  REM
3640  RETURN
3650  REM
3660  TE#=1!/X1
3670  REM
3680 REM
3690  FOR IO=1 TO FIX((N1-1)/2)
3700  CI#=2!*IO-1!
3710  TE#=TE#*X1/CI
3720  SU=SU+TE#
3730  NEXT IO
3740  XS=SQR(X1)
3750  X=XS:GOSUB 4080 :CH#=2!-2!*NR+XS*EXP(-X1/2!)*.797885*SU: REM" CALL NORM (X
S)
3760  RETURN
3770  REM
3780  REM          END OF ROUTINE CHISQ
3790  REM
3800  REM =======================
3810  REM ROUTINE GTRUN (X)
3820  REM
3830  REM G(X) - HALD - MAX-LIKE EST. - ANNALS MATH STAT.
3840  REM           VOL 24, P.557
3850  REM
3860  REM       INPUT .....    X
3870  REM       OUTPUT ....    GT
3880  REM
3890  SP#=1.12837917#
3900  P# = .327591
3910  A1#= .2258368460000001#
3920  A2#=-.252128668#
3930  A3#=1.25969513#
3940  A4#=-1.28782245#
3950  A5#=.9406460700000001#
3960  S2#=.7071067810000001#
3970  REM
3980  XN#=1!/(1!+P#*ABS(X*S2#))
3990  PH#=SP#*EXP(-X*X/2!)
4000  PS#=1!-XN#*(A1#+XN#*(A2#+XN#*(A3#+XN#*(A4#+XN#*A5#))))*PH#
4010  TE#=.5*PS#
4020  IF X>0 THEN TE#=-TE#
4030  GT=1!/(PH#/(2.82842712#*(TE#+.5))-X)
4040  RETURN
4050  REM
4060  REM END OF ROUTINE GTRUN
4070  REM
4080  REM ===================
4090  REM ROUTINE NORM(X)
4100  REM
4110  REM APPROX. FOR DIG. COMP. - HASTINGS, P. 169
4120  REM
4130  REM     INPUT .....   X
4140  REM     OUTPUT ....   NR
4150  REM
```

```
4160  SP#=1.12837917#
4170  P# =.327591
4180  A1#=.225836846000001#
4190  A2#=-.252128668#
4200  A3#=1.25969513#
4210  A4#=-1.28782245#
4220  A5#=.940646070000001#
4230  S2#=.707106781000001#
4240  XN#=1!/(1!+P#*ABS(X*S2#))
4250  PH#=SP#*EXP(-X*X/2!)
4260  PS#=1!-XN#*(A1#+XN#*(A2#+XN#*(A3#+XN#*(A4#+XN#*A5#))))*PH#
4270  TE#=.5*PS#
4280  IF X < 0 THEN TE#=-TE#
4290  NR = .5+TE#
4300  REM
4310  RETURN
4320  REM
4330  REM     END OF ROUTINE NORM
4340  REM
4350  REM ===================
4360  REM     ROUTINE TRUNC
4370  REM
4380  REM     INPUT  .....  X
4390  REM     OUTPUT .....  TR
4400  REM
4410  A0#=34.758138#
4420  A1#=-69.1654
4430  A2#=38.338991#
4440  A3#=-5.38194
4450  B1#=5.019750360000001#
4460  B2#=-7.0780578#
4470  B3#=1.6336409#
4480  REM
4490 Z#=SQR(-2!*LOG(X-.5))
4500  TR=(((A3#*Z#+A2#)*Z#+A1#)*Z#+A0#)/(((B3#*Z#+B2#)*Z#+B1#)*Z#+1!)
4510  REM
4520  RETURN
4530  REM
4540  REM
4550  REM     END OF ROUTINE TRUNC
4560  REM
4570  REM
4580  REM ===========================================================
4590  REM
4600  REM              ********    READ DATA FROM FILE    ********
4610  REM
4620  REM ===========================================================
4630  LINE INPUT"ENTER FILENAME? ";FT$
4640  OPEN "I",1,FT$
4650  REM READING CATCH AT LENGTH DATA FROM DISK
4660  INPUT#1,N1,N,XL,XI,XH
4670  FOR I=1 TO N
4680  INPUT#1,SN,L,F(I),M,D
4690  NEXT I
4700  IF S1>=9999 THEN CLOSE #1
4710  RETURN
```