



# 用判别函数进行鱼类种群鉴别的计算机程序实现方法

## A COMPUTERIZED APPROACH TO IDENTIFICATION OF FISH POPULATION BASED ON DISCRIMINANT ANALYSIS

陈卫忠 李长松

(东海水产研究所, 上海 200090)

Chen Weizhong and Li Changsong

(East China Sea Fisheries Research Institute, Shanghai 200090)

**关键词** 判别函数, 鱼类种群, 鉴别, 计算机程序

**KEYWORDS** discriminant analysis, fish population, identification, computerized approach

鱼类种群或群体的鉴别通常是通过测量鱼类的量度和计数性状, 利用统计学的一些方法进行显著性检验, 以达到鉴别二个取样是否具有显著性差异, 即是否属于二个不同的种群或群体的目的。常用的方法有: 差异系数比较、均数显著性检验、方差分析和判别函数分析等[四川矿业学院数学教研组, 1978], 其中判别函数分析综合地考虑了鱼类取样的各个性状对其的影响, 是一个比较全面且应用较多的方法。然而在进行判别函数分析时, 需要涉及一些复杂的数学计算, 特别是需要求解一个多元线性方程组, 该方程组的元数与分析用的鱼类取样测量性状项数相同, 当鱼类取样性状项数较多时, 方程组的求解将是一个非常繁琐和庞大的过程。而利用计算机编程进行判别函数分析, 利用了计算机迅速准确的优点, 不失为一种较好的方法。我们在参与东海带鱼种群鉴别的过程中, 利用计算机编程的方法, 取得了较好的效果。

### 1 材料与方法

利用判别函数分析进行鱼类种群鉴别的基本原理参考了有关的文献[广东省水产研究所资源室鱼类组、中山大学数力系数学教研室统计组, 1975; 董一中, 1987], 多元线性方程组的求解及行列式的展开原理参考了《数学手册》[四川矿业学院数学教研组, 1978]一书, 为了便于计算机编程, 我们探索使用了列变换的方法进行高阶行列式的展开, 程序的编制是在386计算机上采取汉语 GWBASIC 语言实现的, 进行程序调试所用的带

收稿年月: 1994-07-21.

鱼取样材料是东海带鱼课题组1993年的种群鉴别取样测定材料。

## 2 结果与分析

编制与调试好的计算机程序见后(为了节省篇幅,940语句以后的行列式展开中没有列出测量性状数  $M=4-8$  的有关语句,即1250-1740和2020-2360语句,但其格式与其前后的有关语句完全相同,输入计算机时请参考前后有关语句补入),整个程序分计算机初始化及数据输入、计算各测量性状平均值与观察值数、计算二个样品间的协方差之和、行列式展开、解线性方程组和计算判别函数  $D$  值和  $F$  值几个部分。为了便于读者理解,下面先将判别函数的概念及有关公式的求解方法介绍如下。

假设有两批鱼类取样  $a$  和  $b$  具有  $k$  项测量性状,  $X_{a,i}$  和  $X_{b,i}$  分别为  $a$ 、 $b$  两批取样中第  $i$  项性状第  $t$  个测量值,而

$$X_{a,i} = \frac{1}{N_a} \sum_{t=1}^{N_a} X_{a,i,t} \quad i = 1, 2, \dots, k \quad (1)$$

及

$$X_{b,i} = \frac{1}{N_b} \sum_{t=1}^{N_b} X_{b,i,t} \quad i = 1, 2, \dots, k \quad (2)$$

分别表示  $a$ 、 $b$  两批取样第  $i$  项性状的平均值,令

$$d_i = X_{a,i} - X_{b,i} \quad i = 1, 2, \dots, k \quad (3)$$

$$W_{ij} = \sum_{t=1}^{N_a} (X_{a,i,t} - X_{a,i})(X_{a,j,t} - X_{a,j}) + \sum_{t=1}^{N_b} (X_{b,i,t} - X_{b,i})(X_{b,j,t} - X_{b,j}) \quad (4)$$

$i, j = 1, 2, \dots, k$

$W_{ij}$  为  $a$ 、 $b$  两批取样的协方差之和,从线性方程组

$$\begin{aligned} L_1 W_{11} + L_2 W_{12} + \dots + L_k W_{1k} \\ L_1 W_{21} + L_2 W_{22} + \dots + L_k W_{2k} \\ \dots \\ L_1 W_{k1} + L_2 W_{k2} + \dots + L_k W_{kk} \end{aligned} \quad (5)$$

解出  $L_1, L_2, \dots, L_k$ , 令

$$D = L_1 d_1 + L_2 d_2 + \dots + L_k d_k \quad (6)$$

判别函数

$$F = N_a * N_b / (N_a + N_b) * (N_a + N_b - k - 1) / k * D \quad (7)$$

求出

$$F = N_a * N_b / (N_a + N_b) * (N_a + N_b - k - 1) / k * D \quad (7)$$

之值,根据显著性水平及自由度 ( $k, N_a + N_b - k - 1$ ) 查  $F$  分布表得  $F$  查表值,若求出的  $F$  值大于  $F$  查表值,则表示该两批取样在该显著性水平下具有显著差异。

在进行判别函数分析时,上述多元线性方程组的求解将是一个最为复杂和繁琐的过程,特别是在鱼类取样具有较多项性状测量值时,方程组的求解一般采用行列式展开法。

$$L_1 = dt_1 / dtt \quad L_2 = dt_2 / dtt \quad \dots \quad L_k = dt_k / dtt \quad (8)$$

而

$$dtt = \begin{vmatrix} W_{11} & W_{12} & \dots & W_{1k} \\ W_{21} & W_{22} & \dots & W_{2k} \\ \dots & \dots & \dots & \dots \\ W_{k1} & W_{k2} & \dots & W_{kk} \end{vmatrix} \quad (9)$$

$$dt_1 = \begin{vmatrix} d_1 & W_{12} & \dots & W_{1k} \\ d_2 & W_{22} & \dots & W_{2k} \\ \dots & \dots & \dots & \dots \\ d_k & W_{k2} & \dots & W_{kk} \end{vmatrix} \quad (10)$$

$$dt_2 = \begin{vmatrix} W_{11} & d_1 & \dots & W_{1k} \\ W_{21} & d_2 & \dots & W_{2k} \\ \dots & \dots & \dots & \dots \\ W_{k1} & d_k & \dots & W_{kk} \end{vmatrix} \quad (11)$$

$$dt_k = \begin{vmatrix} W_{11} & W_{12} & \dots & d_1 \\ W_{21} & W_{22} & \dots & d_2 \\ \dots & \dots & \dots & \dots \\ W_{k1} & W_{k2} & \dots & d_k \end{vmatrix} \quad (12)$$

对于一个  $k$  阶的高阶行列式可按某行(或列)的展开法展开,例如按第一行展开:

$$\begin{vmatrix} W_{11} & W_{12} & \dots & W_{1k} \\ W_{21} & W_{22} & \dots & W_{2k} \\ \dots & \dots & \dots & \dots \\ W_{k1} & W_{k2} & \dots & W_{kk} \end{vmatrix} = (-1)^{1+1}W_{11} \begin{vmatrix} W_{22} & W_{23} & \dots & W_{2k} \\ W_{32} & W_{33} & \dots & W_{3k} \\ \dots & \dots & \dots & \dots \\ W_{k2} & W_{k3} & \dots & W_{kk} \end{vmatrix} + \\ + (-1)^{1+2}W_{12} \begin{vmatrix} W_{21} & W_{23} & \dots & W_{2k} \\ W_{31} & W_{33} & \dots & W_{3k} \\ \dots & \dots & \dots & \dots \\ W_{k1} & W_{k3} & \dots & W_{kk} \end{vmatrix} + \\ + \dots + (-1)^{1+k}W_{1k} \begin{vmatrix} W_{21} & W_{22} & \dots & W_{2(k-1)} \\ W_{31} & W_{32} & \dots & W_{3(k-1)} \\ \dots & \dots & \dots & \dots \\ W_{k1} & W_{k2} & \dots & W_{k(k-1)} \end{vmatrix} \quad (13)$$

因此,一个  $k$  阶行列式等于  $k$  个  $k-1$  阶行列式各自乘上其对应的首行元素之和(其正负符号按一定的规律变化),而每一个  $k-1$  阶行列式又等于  $k-1$  个  $k-2$  阶行列式各自乘上其对应的首行元素之和(其正负符号按同样的规律变化)。以此类推,一直分解到二阶行列式,而二阶行列式

$$\begin{vmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{vmatrix} = W_{11}W_{22} - W_{21}W_{12} \quad (14)$$

是由一正一负两个元素积相加组成,这样,一个  $k$  阶的行列式的展开就由  $k * (k-1) * \dots * 2 * 1 = k!$  个元素积相加组成,其正负符号按一定的规律变化,每个元素积又由  $k$  个元素相乘组成。例如,第一个元素积为  $W_{11}W_{22} \dots W_{kk}$  组成,符号为正。假如  $k=10$ ,即鱼类取样有10个测量性状,一个10阶的行列式的展开就由  $10! = 10 * 9 * \dots * 2 * 1 = 3628800$  个元素积相加组成,每个元素积又由10个元素相乘组成,其正负符号又有一定的变化规律,因此,整个计算过程将是一个非常繁琐和庞大的过程,靠手工或计算器几乎是无法完成这样的计算的。

为了便于编程计算,我们找出了  $k$  阶行列式展开时各项之间的关系,利用列变换的方法,较好地解决了计算机编程问题,其基本原理如下:

如(13)式所示的  $k$  阶行列式展开后的各项就相当于在展开前把该项所对应的列移到原行列式的第一列,并把该列之前的每一列依次往后移一列,再展开取其第一项,但正负号仍按原来的变化规律取,例如,  $k$  阶行列式展开后的第3项就相当于在展开前,把该行列式的第3列移到第1列,然后将原第1、第2列依次移到第2、第3列后再展开,取其第1项,对于  $k-1, k-2, \dots, 2$  阶行列式,可采用相同的原理展开,因此,在进行  $k$  阶行列式展开时,先依次将  $1, 2, \dots, k$  列移到首列,每移动一列后,把该列之前的各列依次往后移一列,然后展开,取其第1项,该第1项包括一个正负符号项乘上原行列式的首行首列元素( $W_{11}$ )再乘上一个  $k-1$  阶行列式。而一个  $k-1$  阶行列式同样可采用列变换的方法展开,依次类推,将  $k-2, \dots, 2$  阶行列式展开,但要注意,在每一次列变换、展开后,要进行列还原,再进行下一次的列变换和展开。整个  $k$  阶行列式展开后就等于  $k!$  个展开项之

和,具体程序见后第940—2590行程序。

在求解5式所示的多元线性方程组时,采用了与上述方法相似但稍有不同列替换方法,即用  $d_1, d_2, \dots, d_k$  依次替换9式中的第  $1, 2, \dots, k$  列,然后展开求得  $dt_1, dt_2, \dots, dt_k$ ,再根据8式依次求得  $L_1, L_2, \dots, L_k$ ,具体程序见后第580—930行程序。

程序第10—210行为计算机初始化及数据输入程序,它首先要输入需比较的二批取样数据文件名,该二个数据文件都存放在当前目录下,它们是在用 FOXBASE 数据库管理系统生成的鱼类取样测定数据库的基础上,用 COPY TO 命令传送的利用逗号作为分界符的通用格式文件,直接可供 GWBASIC 语言接受。在输入文件名后,再输入取样的测量性状数及二批取样中较大样品的取样数,然后计算机就进入计算状态,最后打印出计算结果 D 和 F 值。

为了节约篇幅,程序仅给出了样品测量性状数为10以内任意的判别函数分析,这在一般的情况下已足够了,如果读者需要计算样品测量性状数10以上的判别函数时,只要在940—2590行的行列式展开程序中按上述行列式展开原理增加一些相应的语句即可。本程序也适用于其它方面二个样本间的判别函数分析。

本程序已应用于东海带鱼群体的鉴别工作,其分析结果被有关专家认为是可信可靠的。

### 3 讨论

关于利用判别函数分析进行鱼类种群鉴定的研究国内已有一些学者做过[广东省水产研究所资源室鱼类组、中山大学数力系数学教研室统计组,1975;韦 晟、周彬彬,1988;张其水、蔡泽平,1983;熊国强等,1992],对利用判别函数分析进行鱼类种群鉴定的基本原理及应用例子,上述有关的文献已有介绍,但对利用判别函数分析进行鱼类种群鉴定的实现方法方面(主要是有关多元线性方程组的求解方法问题)的报导还没有见到过,作者参考了本文材料与方法中已引用的《数学手册》,采用行列式展开的方法求解多元线性方程,同时,为了便于计算机编程,我们采用了列变换形式进行高阶行列展开的方法,较好地解决了多元线性方程的求解和计算机编程问题。所编的程序具有较大的灵活性,能对测量性状数为10以内的任意数的二个样品进行判别函数分析。当然,采用行列式展开的方法求解多元线性方程组可能不是唯一的方法,用列变换的形式进行高阶行列式展开也可能不是最佳方法,我们在这里主要起抛砖引玉的作用,供大家参考,同时,也为有关读者今后应用判别函数分析提供方便。

### 参 考 文 献

- [1] 广东省水产研究所资源室鱼类组、中山大学数力系数学教研室统计组,1975.应用判别函数和方差分析对蓝圆鲈分群问题的探讨.数学学报,18(3):185—191.
- [2] 韦 晟、周彬彬,1988.渤、黄海蓝点马鲛种群鉴别的研究.动物学报,34(1):71—81.
- [3] 四川矿业学院数学教研组编,1978.数学手册.7—12,科学出版社(京).
- [4] 张其水、蔡泽平,1983.台湾海峡和北部湾二长棘鲷种群鉴别研究.海洋与湖沼,14(6):511—520
- [5] 董一中(编著),1987.生物统计法.427—449,湖南科学技术出版社(长沙).
- [6] 熊国强等,1992.中国沿海日本鳗鲡幼苗群体鉴别研究.动物学报,38(3):254—265.

## 附录:用判别函数进行鱼类种群鉴别的计算机程序

```

10 ' 判别函数显著性检验程序
20 ' 计算机初始化及数据输入
30 INPUT "请输入第一批数据文件名",DBN1¥
40 INPUT "请输入第二批数据文件名",DBN2¥
50 INPUT "请输入样品性状数",M
60 INPUT "请输入最大样品数",N
70 DIM DD(M),K1(M),K2(M),K3(M),K4(M),
    K5(M),K6(M),K7(M),K8(M),K9(M)
80 DIM X(N,M,2),XX(M,2),TX(M),N(M,2),
    D(M),W(M,M),E(M,M,2),DT(M),K(M),
    L(M)
90 FOR K=1 TO 2
100 IF K=2 THEN OPEN DBN2¥ FOR INPUT
    AS #1:GOTO 120
110 OPEN DBN1¥ FOR INPUT AS #1
120 N1=0
130 FOR I=1 TO N
140 IF EOF(1) THEN 200
150 N1=N1+1
160 FOR J=1 TO M
170 INPUT #1,X(I,J,K)
180 NEXT J
190 NEXT I
200 CLOSE #1
210 ' 计算各测量性状平均值及观察值数
220 FOR J=1 TO M
230 TX(J)=0:T=0
240 FOR I=1 TO N1
250 IF X(I,J,K)=0 THEN T=T+1:GOTO 270
260 TX(J)=TX(J)+X(I,J,K)
270 NEXT I
280 N(J,K)=N1-T
290 IF N(J,K)=0 THEN 310
300 XX(J,K)=TX(J)/N(J,K)
310 NEXT J
320 IF K=2 THEN NB=N1:GOTO 340
330 NA=N1
340 NEXT K
350 ' 计算二个样品间的协方差之和
360 FOR I=1 TO M
370 FOR J=1 TO M
380 E(I,J,1)=0:E(I,J,2)=0
390 FOR T=1 TO NA
400 X1=(X(T,I,1)-XX(I,1))*(X(T,J,1)-XX
    (J,1))
410 IF X(T,I,1)=0 THEN X1=0
420 E(I,J,1)=E(I,J,1)+X1
430 NEXT T
440 FOR T=1 TO NB
450 X2=(X(T,I,2)-XX(I,2))*(X(T,J,2)-XX
    (J,2))
460 IF X(T,I,2)=0 THEN X2=0
470 E(I,J,2)=E(I,J,2)+X2
480 NEXT T
490 NEXT J
500 NEXT I
510 FOR I=1 TO M
520 D(I)=XX(I,1)-XX(I,2)
530 FOR J=1 TO M
540 W(I,J)=E(I,J,1)+E(I,J,2)
550 NEXT J
560 NEXT I
570 ' 解线性方程组
580 D=0
590 GOSUB 940
600 DTT=DT
610 FOR S=1 TO M
620 FOR R=1 TO M
630 K(R)=W(R,S)
640 W(R,S)=D(R)
650 NEXT R
660 GOSUB 940
670 DT(S)=DT
680 FOR R=1 TO M
690 W(R,S)=K(R)
700 NEXT R
710 L(S)=DT(S)/DTT
720 D=D+D(S)*L(S)
730 NEXT S
740 ' 求算判别函数 D 和 F 值

```

```

750 F=NA * NB/(NA+NB) * (NA+NB-M-
    1)/M * D
760 LPRINT DBN1 ¥;"--";DBN2 ¥
770 LPRINT "判别函数(D)=";D
780 LPRINT "F=";F
790 LPRINT "观察值数(na+nb)=";NA;"+";NB
800 FOR I=1 TO M
810 LPRINT "d(";I;")=";D(I)
820 NEXT I
830 FOR I=1 TO M
840 LPRINT "L(";I;")=";L(I)
850 NEXT I
860 FOR I=1 TO M
870 DL(I)=0
880 FOR J=1 TO M
890 DL(I)=DL(I)+W(I,J) * L(J)
900 NEXT J
910 LPRINT "DL(";I;")=";DL(I)
920 NEXT I
930 END
940 '行列式展开
950 DT=0
960 FOR I1=1 TO M
970 FOR J=1 TO M
980 K1(J)=W(J,I1)
990 FOR JJ=I1 TO 1 STEP -1
1000 W(J,JJ)=W(J,JJ-1)
1010 NEXT JJ
1020 W(J,1)=K1(J)
1030 NEXT J
1040 D1=(-1)^(1+I1) * W(1,1)
1050 FOR I2=2 TO M
1060 FOR J=2 TO M
1070 K2(J)=W(J,I2)
1080 FOR JJ=I2 TO 2 STEP -1
1090 W(J,JJ)=W(J,JJ-1)
1100 NEXT JJ
1110 W(J,2)=K2(J)
1120 NEXT J
1130 IF M=2 THEN D0=D1 * W(2,2):GOTO
    1860
1140 D2=(-1)^ I2 * W(2,2)
1150 FOR I3=3 TO M
1160 FOR J=3 TO M
1170 K3(J)=W(J,I3)
1180 FOR JJ=I3 TO 3 STEP -1
1190 W(J,JJ)=W(J,JJ-1)
1200 NEXT JJ
1210 W(J,3)=K3(J)
1220 NEXT J
1230 IF M=3 THEN D0=D1 * D2 * W(3,3):
    GOTO 1860
1240 D3=(-1)^(I3-1) * W(3,3)
    .....
1250 FOR I9=9 TO M
1260 FOR J=9 TO M
1270 K9(J)=W(J,I9)
1280 FOR JJ=I9 TO 9 STEP -1
1290 W(J,JJ)=W(J,JJ-1)
1300 NEXT JJ
1310 W(J,9)=K9(J)
1320 NEXT J
1330 IF M=9 THEN D0=D1 * D2 * D3 * D4 * D5
    * D6 * D7 * D8 * W(9,9):GOTO 1860
1340 D9=(-1)^(I9-7) * W(9,9)
1350 D0=D1 * D2 * D3 * D4 * D5 * D6 * D7 * D8 *
    D9 * W(10,10)
1360 DT=DT+D0
1370 IF M=9 THEN 2020
1380 IF M=8 THEN 2090
1390 IF M=7 THEN 2160
1400 IF M=6 THEN 2230
1410 IF M=5 THEN 2300
1420 IF M=4 THEN 2370
1430 IF M=3 THEN 2440
1440 IF M=2 THEN 2510
1450 FOR J=9 TO M
1460 FOR JJ=9 TO I9-1
1470 W(J,JJ)=W(J,JJ+1)
1480 NEXT JJ
1490 W(J,I9)=K9(J)
1500 NEXT J
1510 NEXT I9
    .....
1520 FOR J=3 TO M
1530 FOR JJ=3 TO I3-1

```

```
2390 W(J,JJ)=W(J,JJ+1)
2400 NEXT JJ
2410 W(J,I3)=K3(J)
2420 NEXT J
2430 NEXT I3
2440 FOR J=2 TO M
2450 FOR JJ=2 TO I2-1
2460 W(J,JJ)=W(J,JJ+1)
2470 NEXT JJ
2480 W(J,I2)=K2(J)

2490 NEXT J
2500 NEXT I2
2510 FOR J=1 TO M
2520 FOR JJ=1 TO I1-1
2530 W(J,JJ)=W(J,JJ+1)
2540 NEXT JJ
2550 W(J,I1)=K1(J)
2560 NEXT J
2570 NEXT I1
2580 RETURN
```