数学地质专辑

（二）

杨自强 等著

97.2.28

地质出版社

数学地质专辑

（二）

杨自强　等著

地质出版社

内容摘要

本辑共收集了十八篇文章。前三篇是数学方法问题的讨论，其余都是作者结合地质工作实践写出的论文，涉及到金属矿产地质研究和预测、地层、岩石化学、航空像片水系研究、油层压裂、化探和放射性物探资料处理等许多方面。有一篇文章介绍用电子计算机绘制常用地质图形的一些问题，另一篇介绍探井文件数据库的应用。

可供广大地质工作者及石油地质工作者参考。

数学地质专辑

（二）

杨自强 等著

*

地质部书刊编辑室编辑

地 质 出 版 社 出 版

（北京西四）

地 质 印 刷 厂 印 刷

（北京安德路47号）

新华书店北京发行所发行·各地新华书店经售

*

开本：787×1092 1/16·印张：12 3/4·插页：1个·字数：299,000

1981年3月北京第一版·1981年3月北京第一次印刷

印数1—2,580册·定价2.40元

统一书号：15038·新581

目　录

一种新的逐步判别方法

杨 自 强

判别分析是数学地质中广泛使用的一种多元统计分析方法。目前较好的算法是近十年出现的基于Wilks-Λ准则的多类逐步判别算法（以下简称W算法）。该算法最早见于Dixon所编的书[7]，作者曾进行过详细的讨论，并给出了计算程序[1,4]。Enslein 等人1977年所编的书[11]也收集了这一算法。作者与有关同事协作，在油田、岩矿、地热等地质问题中应用此法收到了良好效果。

本文介绍一种新的逐步判别算法，其主要思想是使用多因变量逐步回归方法作多类逐步判别计算。1968年Glahn[5]曾通过典型相关分析讨论了判别分析与回归分析的关系，建立了判别—回归模型，作者[2]把多因变量的逐步回归方法用于这个模型，得到一种新的逐步判别算法（以下简称Y算法），并证明了这个算法与W算法在舍选变量行为方面的等价性，即挑选变量的过程和最终结果完全一样❶。但是新的Y算法比原来的W算法在计算上更节省时间和存贮单元。

限于篇幅，本文仅介绍算法本身，并给出计算程序和例子，有关这里的判别—回归模型与经典的Fisher判别模型[6]的关系，以及两种算法在舍选变量方面的等价性证明都不再给出，有兴趣的读者可以参考作者的另文[2]。

§1. 判别—回归模型

沿用多元回归分析中的惯用符号，记回归方程为

$$\hat{Y}-\bar{Y}=(X-\bar{X})\hat{\beta} \tag{1.1}$$

回归系数的估值为

$$\hat{\beta}=L_{XX}^{-1}L_{XY} \tag{1.2}$$

其中L_{XX}是自变量X的离差矩阵，L_{XX}^{-1}是L_{XX}的逆矩阵，L_{XY}是自变量X与因变量Y的离差矩阵。上面的几个式子对于因变量Y也是多个的情形依然成立。为明确起见，今后假设有p个自变量$x_1,\cdots,x_p$和q个因变量$y_1,\cdots,y_q$，并有n次观测数据($n>p$)，数据矩阵为

$$X=\begin{pmatrix} x_{11} & \cdots\cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots\cdots & x_{np} \end{pmatrix}_{n\times p},\quad Y=\begin{pmatrix} y_{11} & \cdots\cdots & y_{1q} \\ \vdots & & \vdots \\ y_{n1} & \cdots\cdots & y_{nq} \end{pmatrix}_{n\times q}$$

这时均值矩阵为

❶ 尽管选出的变量相同，但两种算法的判别函数并不相同，因此分类效果可以有差别。

$$\bar{X}=\begin{pmatrix}\bar{x}_1 & \cdots\cdots & \bar{x}_p\\ \vdots & & \vdots\\ \bar{x}_1 & \cdots\cdots & \bar{x}_p\end{pmatrix}_{n\times p},\quad \bar{Y}=\begin{pmatrix}\bar{y}_1 & \cdots\cdots & \bar{y}_q\\ \vdots & & \vdots\\ \bar{y}_1 & \cdots\cdots & \bar{y}_q\end{pmatrix}_{n\times q}$$

其中

$$\left.\begin{aligned}\bar{x}_i&=\frac{1}{n}\sum_{k=1}^{n}x_{ki},\quad i=1,\cdots,p\\ \bar{y}_u&=\frac{1}{n}\sum_{k=1}^{n}y_{ku},\quad u=1,\cdots,q\end{aligned}\right\}\tag{1.3}$$

离差矩阵为

$$L_{XX}=(l_{ij})_{p\times p}$$

其中

$$l_{ij}=\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)\qquad i,j=1,\cdots,p\tag{1.4}$$

$$L_{XY}=(l_{iu})_{p\times q}$$

其中

$$l_{iu}=\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)(y_{ku}-\bar{y}_u)\qquad i=1,\cdots,p;u=1,\cdots,q\tag{1.5}$$

所得的回归系数矩阵记为

$$\hat{\beta}=\begin{pmatrix}\hat{\beta}_{11} & \cdots\cdots & \hat{\beta}_{1q}\\ \vdots & & \vdots\\ \hat{\beta}_{p1} & \cdots\cdots & \hat{\beta}_{pq}\end{pmatrix}_{p\times q}$$

估值记为

$$\hat{Y}=\begin{pmatrix}\hat{y}_{11} & \cdots\cdots & \hat{y}_{1q}\\ \vdots & & \vdots\\ \hat{y}_{n1} & \cdots\cdots & \hat{y}_{nq}\end{pmatrix}_{n\times q}$$

假设数据来自G类母体，各类的样本容量分别为$n_1,\cdots,n_G$这时

$$n=\sum_{g=1}^{G}n_g\tag{1.6}$$

判别分类问题本来没有因变量Y，但是为了使用回归模型作判别分类，我们不妨虚构G个因变量的n次观测数据如右（此处暂令$q=G$）：

	y_1	y_2	$\cdots$	y_G	
第1类	1 1 ⋮ 1	0 0 ⋮ 0	… … …	0 0 ⋮ 0	n_1
第2类	0 0 ⋮ 0	1 1 ⋮ 1	… … …	0 0 ⋮ 0	n_2
⋮	⋮	⋮		⋮	
第G类	0 0 ⋮ 0	0 0 ⋮ 0	… … …	1 1 ⋮ 1	n_G

或简记为

$$y_{kg}=\begin{cases}1 & \text{第 }k\text{ 次观测值属于第 }g\text{ 类}\\ 0 & \text{第 }k\text{ 次观测值不属第 }g\text{ 类}\end{cases}\tag{1.7}$$

由于Y有特殊形式，如（1.5）等与Y有关的一些式子在实际计算时当然应当化简，回归系数的估值$\hat{\beta}$仍可按常规办法计算。

所谓判别—回归模型，就是把上述具有特殊形式Y的回归模型用于判别分类。办法是：在估计出$\hat{\beta}$后，对于任一观测点$X_k=(x_{k1},\cdots,x_{kp})$，由（1.1）算出相应的预报量$\hat{Y}_k=(\hat{y}_{k1},\cdots,\hat{y}_{kG})$，如果其中的$\hat{y}_{ku}$最靠近1，即

$$|\hat{y}_{ku}-1|=\min_{1\leqslant g\leqslant G}|\hat{y}_{kg}-1| \tag{1.8}$$

则认为X_k来自第u类母体。在作分类预报时，观测点X_k既可以是原来n个观测点中的一个，也可以是其它的任意一个。

在实际计算时，请注意下述定理：

定理1.1 如果Y有（1.7）的特殊形式，则预报量满足

$$\sum_{g=1}^{G}\hat{y}_{kg}=1,\qquad k=1,\cdots,n \tag{1.9}$$

证 由Y的构造易知

$$\sum_{g=1}^{G}y_{kg}=\sum_{g=1}^{G}\bar{y}_g=1,\qquad k=1,\cdots,n \tag{1.10}$$

$$\sum_{g=1}^{G}(y_{kg}-\bar{y}_g)=0 \tag{1.11}$$

引入记号 $L_{XX}^{-1}=(c_{ij})_{p\times p}$，这时

$$\begin{aligned}
\sum_{g=1}^{G}(\hat{y}_{kg}-\bar{y}_g) &=\sum_{g=1}^{G}\sum_{i=1}^{p}(x_{ki}-\bar{x}_i)\hat{\beta}_{ig} \quad \text{〔见(1.1)〕}\\
&=\sum_{g=1}^{G}\sum_{i=1}^{p}(x_{ki}-\bar{x}_i)\sum_{j=1}^{p}c_{ij}l_{jg} \quad \text{〔见(1.2)〕}\\
&=\sum_{g=1}^{G}\sum_{i=1}^{p}(x_{ki}-\bar{x}_i)\sum_{j=1}^{p}c_{ij}\sum_{k=1}^{n}(x_{kj}-\bar{x}_j)(y_{kg}-\bar{y}_g)\\
&=\sum_{i=1}^{p}(x_{ki}-\bar{x}_i)\sum_{j=1}^{p}c_{ij}\sum_{k=1}^{n}(x_{kj}-\bar{x}_j)\sum_{g=1}^{G}(y_{kg}-\bar{y}_g)\\
&=0 \quad \text{〔见(1.11)〕}
\end{aligned}$$

$$\sum_{g=1}^{G}\hat{y}_{kg}=\sum_{g=1}^{G}\bar{y}_g=1. \quad \text{证毕。}$$

上述定理表明，G个虚构的因变量是相互有关的，而且它们的估计值也保留了这个特性，因此实际计算时，仅需对前$G-1$个y进行，一旦估计出$\hat{y}_{k1},\cdots,\hat{y}_{k,G-1}$之后，就可由下式确定$\hat{y}_{kG}$：

$$\hat{y}_{kG}=1-\sum_{g=1}^{G-1}\hat{y}_{kg} \tag{1.12}$$

在其后的讨论中，我们总假设因变量的个数为$q=G-1$。

显然，过去用于两类判别的二值回归[8,1,4]是这里的判别—回归模型的特例。

§2. 逐 步 算 法

单因变量的逐步回归方法（例如参考[8]或[4,7,9]）已为众所周知，并且可以用于二值回归，实现两类母体的逐步判别分类[1,4]。不难想象，如果把单因变量的逐步回归方法推广到多因变量的情形，就可以直接用于上述判别一回归模型，并实现多类母体的逐步判别分类，这就是本文要介绍的Y算法。因此本节只需介绍多因变量的逐步回归方法。

多因变量逐步回归方法[3]与单因变量逐步回归方法十分相似，它从p个自变量$x_1,\cdots,x_p$中逐步选出对建立多个因变量的回归方程是"重要"的变量，每步只选一个，直至按某一准则再没有重要变量为止。随着选入回归方程的变量不断增加，较早选入的某个变量，其作用可能为其后进入的变量（或变量组合）所代替，如果确有这样的变量，就应及时从方程中剔除出去，使最终的方程只保留重要变量。计算的核心仍然是使用无回代过程的高斯一约唐消去法，在求解$\hat{\beta}$的过程中同时求L_{XX}^{-1}，并且利用中间结果构造Hotelling的T^2统计量或与之相当的F统计量，用以检验各自变量的重要性。

把L_{XX}，L_{XY}，L_{YY}按如下形式组成新矩阵

$$S=\begin{pmatrix} L_{XX} & L_{XY} \\ L_{YX} & L_{YY}\end{pmatrix} \tag{2.1}$$

其中L_{YX}是L_{XY}的转置矩阵，即$L_{YX}=L'_{XY}$。新矩阵的元素统一编号，记为l_{IJ}（I，$J=1,\cdots p+q$）。利用高斯一约唐消去法，第$t+1$步对以$l_{kk}(k=1,\cdots,p)$为轴的变换公式是

$$l_{IJ}^{(t+1)}=\begin{cases} \dfrac{l_{kJ}^{(t)}}{l_{kk}^{(t)}} & (I=k,\ J\neq k) \\[2ex] l_{IJ}^{(t)}-\dfrac{l_{Ik}^{(t)}l_{kJ}^{(t)}}{l_{kk}^{(t)}} & (I\neq k,\ J\neq k) \\[2ex] \dfrac{1}{l_{kk}^{(t)}} & (I=k,\ J=k) \\[2ex] -\dfrac{l_{Ik}^{(t)}}{l_{kk}^{(t)}} & (I\neq k,\ J=k) \end{cases} \tag{2.2}$$

其中$l_{IJ}^{(0)}=l_{IJ}$。如果对上述矩阵进行了p次变换($k=1,\cdots,p$)，则有

$$\begin{pmatrix} L_{XX}^{(p)} & L_{XY}^{(p)} \\ L_{YX}^{(p)} & L_{YY}^{(p)}\end{pmatrix}=\begin{pmatrix} L_{XX}^{-1} & L_{XX}^{-1}L_{XY} \\ -L_{YX}L_{XX}^{-1} & L_{YY}-L_{YX}L_{XX}^{-1}L_{XY}\end{pmatrix}=\begin{pmatrix} L_{XX}^{-1} & \hat{\beta} \\ -\hat{\beta}' & Q\end{pmatrix}$$

这里的Q是剩余矩阵

$$\begin{aligned} Q&=(Y-\hat{Y})'(Y-\hat{Y})=L_{YY}-L_{YX}\hat{\beta} \\ &=L_{YY}-L_{YX}L_{XX}^{-1}L_{XY} \end{aligned} \tag{2.3}$$

把$\hat{\beta}$的第i行元素所成的列向量记为

$$\hat{\beta}_i=(\hat{\beta}_{i1},\cdots,\hat{\beta}_{iq})'=(l_{i,p+1}^{(p)},\cdots,l_{i,p+q}^{(p)})'=L_i^{(p)} \tag{2.4}$$

它是变量x_i的回归系数的估计值。对于假设

$$H_o:\beta_i=0$$

有

$$T_i^2=(n-p-1)\frac{1}{c_{ii}}\hat{\beta}'_i Q^{-1}\hat{\beta}_i \tag{2.5}$$

其中c_{ii}是L_{XX}^{-1}对角线上的第 i 个元素。量T_i^2遵从T^2分布，与F分布有如下关系

$$\frac{n-p-q}{(n-p-1)q}T_i^2\sim F_i(q,\ n-p-q) \tag{2.6}$$

或

$$F_i(q,n-p-q)=\frac{n-p-q}{q}\ \frac{1}{l_{ii}^{(p)}}L_i^{(p)\prime}[L_{YY}^{(p)}]^{-1}L_i^{(p)} \tag{2.7}$$

如果已计算了 t 步，并选进了m个变量（包括 $t=0$，$m=0$），则第 $t+1$步的计算步骤如下：

（1）对于已选的m个变量（如果有的话），计算

$$F_i(q,n-m-q)=\frac{n-m-q}{q}\ \frac{1}{l_{ii}^{(t)}}L_i^{(t)\prime}[L_{YY}^{(t)}]^{-1}L_i^{(t)} \tag{2.8}$$

并从中选出最小者（记为$F_{\min}$）。对于未选入的$p-m$个变量，计算

$$F_i(q,n-(m+1)-q)=\frac{n-(m+1)-q}{q}\ \frac{1}{l_{ii}^{(t+1)}}L_i^{(t+1)\prime}[L_{YY}^{(t+1)}]^{-1}L_i^{(t+1)}$$

$$=\frac{n-(m+1)-q}{q}\ \frac{L_i^{(t)\prime}[L_{YY}^{(t)}]^{-1}L_i^{(t)}}{l_{ii}^{(t)}-L_i^{(t)\prime}[L_{YY}^{(t)}]^{-1}L_i^{(t)}} \text{❶} \tag{2.9}$$

并从中选出最大者（记为$F_{\max}$）。

（2）如果$F_{\min}\leqslant F_\alpha(q,n-m-q)$❷，则剔除与$F_{\min}$相应的变量$x$（记为$x_k$），其后计算见（3）；否则考察是否有$F_{\max}>F_\alpha(q,n-(m+1)-q)$，如果满足，则引入与$F_{\max}$相应的$x$（为方便仍记为$x_k$），其后计算也是（3）。在既不能剔除已选量，又不能引入未选量时，逐步计算便告完成。

（3）无论引入或剔除x_k都使用（2.2）式计算。其后重复（1）—（3）进行下一步计算。

上述多因变量逐步回归的计算公式与数据Y是否有（1.7）的特殊形式无关。如果q较大，可以使用下述方法计算两个F检验式（2.8）和（2.9）中的主要计算部分

$$a_i=L'_iL_{YY}^{-1}L_i$$

为方便，上式已略去了上标。由L_{YY}的正定性，得分解式

$$L_{YY}=LL' \tag{2.10}$$

其中L是下三角矩阵，于是

$$a_i=L'_i(LL')^{-1}L_i=(L^{-1}L_i)'(L^{-1}L_i)$$

记

$$L^{-1}L_i=V_i\ \text{或}\ LV_i=L_i \tag{2.11}$$

---

❶ 推演中需使用（2.2）式和矩阵论中的如下公式：

$$(A-rs')^{-1}=A^{-1}+\frac{A^{-1}rs'A^{-1}}{1-s'A^{-1}r}$$

其中A为方阵，A^{-1}存在，r,s为列向量。

❷ $F_\alpha(q,n-m-q)$是自由度为q和$n-m-q$的F分布在显著水平α时的临界值，当n足够大时，它随m的变化很小，为方便，实际计算时不再考虑m的变化，各步都用同一个临界值F_α。

得

$$a_i=V'_i V_i$$

这样一来，在每一步的主要计算量将是按(2.10)分解一个矩阵$L_{YY}^{(t)}$，并对p个不同的$L_i^{(t)}$求解p个系数矩阵为下三角形的方程组（2.11）。这种算法比直接算法来得节省。

顺便指出，逐步算法可以使用"向后"方式，即把全部变量引入回归后，再着手考虑剔除不显著的变量。此外尚可证明，如果各步计算都用同一个临界值F_α，则第$t+1$，$t+2$步引入回归方程的变量不可能在第$t+3$步中被剔除，因此每步实际上只需解$p-2$个形如（2.11）的三角方程组。

§3. 程序和数值例子

（一）补充说明

本程序使用多因变量逐步回归技术作多类逐步判别分类。特点是仅对矩阵

$$S=\begin{pmatrix} L_{XX} & L_{XY} \\ L_{YX} & L_{YY} \end{pmatrix}$$

的下三角部分存贮和计算，从而使这两者都得到节省。此外使用者还可以根据需要，选用"向后"逐步计算方式。程序使用TQ-16机BCY语言的过程形式给出。为节省篇幅，输出仅用注解方式指出，而没有给出具体的印刷语句。

为阅读本程序方便，现补充几点说明：

1．均值和L_{XX}使用递推算法。

$$\bar{x}_i^{(k)}=\bar{x}_i^{(k-1)}+\frac{(x_{ki}-\bar{x}_i^{(k-1)})}{k}$$

$$l_{ij}^{(k)}=l_{ij}^{(k-1)}+(x_{ki}-\bar{x}_i^{(k-1)})(x_{kj}-x_j^{(k)})$$

$$(k=1,\cdots,n;\ i=1,\cdots,p;\ j=1,\cdots i)$$

初值$\bar{x}_i^{(0)}=0$，$l_{ij}^{(0)}=0$。终值$\bar{x}_i^{(n)}=\bar{x}_i, l_{ij}^{(n)}=l_{ij}$。

2.为了对下三角部分运算方便，将消去计算所用的式子（2.2）稍作修改，使S在计算过程中始终对称。

$$l_{IJ}^{(t+1)}=\begin{cases} \pm\dfrac{l_{kJ}^{(t)}}{l_{kk}^{(t)}} & (I=k,J\neq k) \\ l_{IJ}^{(t)}-\dfrac{l_{Ik}^{(t)}l_{kJ}^{(t)}}{l_{kk}^{(t)}} & (I\neq k,J\neq k) \\ -\dfrac{1}{l_{kk}^{(t)}} & (I=k,J=k) \\ \pm\dfrac{l_{Ik}^{(t)}}{l_{kk}^{(t)}} & (I\neq k,J=k) \end{cases}$$

其中首末两式的符号在引入x_k时取正，剔除时取负。于是如果$x_1,\cdots,x_p$都引入回归，S最终为

$$\begin{pmatrix} -L_{XX}^{-1} & L_{XX}^{-1}L_{XY} \\ L_{YX}L_{XX}^{-1} & L_{YY}-L_{YX}L_{XX}^{-1}L_{XY} \end{pmatrix}=\begin{pmatrix} -L_{XX}^{-1} & \hat{\beta} \\ \hat{\beta}' & Q \end{pmatrix}$$

在计算过程中，如果 L_{XX} 对角线上的某个元素变为负值，表示相应的变量已被引入回归方程。

3．按惯例，逐步计算前先将 S 标准化为相关矩阵。

（二）参数表

下面八个参数是使用者给出的：

p　变量 x 的个数；

n　用于分析的观测次数；

nn　总的观测次数；通常 $nn=n$，当 $nn>n$ 时，表示后 $nn-n$ 次观测的分类未知，需要预报；

G　分类数；

n_g　场 $n_g[1:G]$ 依次存放各类观测的次数 $n_1,\cdots,n_G$，其中 $n=n_1+\cdots+n_G$；

X　场 $X[1:nn,1:p]$ 存放各次观测数据，要求各类数据依次分类集中；

F_{a12}　F 统计量的临界值 F_a，用于舍选变量；

$back$　若 $back=0$，则采用向前逐步算法，否则用向后逐步算法。

下面七个参数放计算结果：

m　最终选入回归的变量个数；

B　场 $B[1:G-1,0:p]$ 放回归系数。$B[u,0]=\bar{y}_u-\sum_i \bar{x}_i\,\hat{\beta}_{iu}$；

F　场 $F[1:p]$ 放各变量的 F 检验值〔见（2.8），（2.9）〕；

X_2　度量判别能力的统计量。记Wilks统计量为 $\Delta^{(m)}=|W^{(m)}|/|L_{XX}^{(m)}|$，其中 W 为组内离差阵，则 $X_2=-((n-1)-(m+G)/2)\ln\Delta^{(m)}$，它近似服从 $\chi^2(m(G-1))$ 分布。

DM　场 $DM[1:G+1,1:G+1]$ 存放分类矩阵（见数值例子）。

E　场 $E[1:nn]$ 放各次观测的计算分类号码。

$fail$　如果计算顺利，$fail=0$；否则 $fail<0$，约定 $fail=-1$ 表示有某个 x_i 为常数，$fail=-2$ 表示在目前的 F_a 值，找不到重要的变量。

（三）程序

```
过程 SRDA(X,p,n,nn,G,ng,Fa12,back,
          m,B,F,X2,DM,E,fail);
 值 p,n,nn,G,Fa12,back; 简变m,X2,fail;
 场X, ng,B,F,DM,E;
 始简变d,pp2,q,G1,ij,k,T12,F12, Ti,
      Wilks, step,Tmax,Tmin,imax,imin;
  场Mx,Vx,Enter[1:p+G-1],Mgx[1:G,1:p],
    S[1:(p+G-1)*(p+G)/2], L[1:G*(G-1)/2];
  过程PS; {本过程产生S矩阵之下三角部分}
    始简变d,nu,k0,ij,uv;
      0⇒Mx; 0⇒Vx; 0⇒S;
      0⇒k0; pp2⇒uv; {其中pp2=p(p+1)/2}
```

```
对于u=1到G步长1执行
  始n_g[u]⇒n_u;
    对于k=k0+1到k0+n_u步长1执行
      始0⇒ij;
        对于i=1到p步长1执行
          始X[k,i]-Mx[i]⇒d;
            Mx[i]+d/k⇒Mx[i];
            对于j=1到 i 步长1执行
              S[ij+j]+d*(X[k,j]-Mx[j])⇒S[ij+j];
            ij+i⇒ij
          终
      终;
    k0+n_u⇒k0;
    对于i=1到p步长1执行
      始k0*Mx[i]⇒d;
        (d-V_x[i])/n_u⇒M_gx[u,i];
        d⇒V_x[i]
      终
  终;
对于u=1到q步长1执行
  始n_g[u]⇒n_u;  n_u/n⇒Mx[p+u];
    对于i=1到p步长1执行
      n_u*(M_gx[u,i]-Mx[i])⇒S[uv+i];
    对于v=1到u-1步长1执行
      -n_u*n_g[v]/n⇒S[uv+p+v];
    uv+p+u⇒uv;  n_u-n_u*n_u/n⇒S[uv]
  终;{此处可印S（离差矩阵）}
0⇒ij;{把S标准化为相关阵}
对于i=1到p+q步长1执行
  始若S[ij+i]/n<2_10-9
      则始-1⇒fail;转Lend终
      否∈SQRT(S[ij+i]/n)⇒V_x[i]⇒d;
    n*d⇒d;
    对于j=1到 i 步长1执行
      S[ij+j]/(d*V_x[j])⇒S[ij+j];
    ij+i⇒ij
  终;{此处可印M_x（均值），V_x（标准差），
      M_gx（分类均值），S（相关矩阵）}
终;
```

```
过程PLLT; {本过程分解L_YY=LL′}
  始简变c,d,u_i,ij,u_j,j_k;
    pp2⇒u_j;  0⇒j_k;
    对于 j=1 到 q 步长 1 执行
      始u_j+p⇒u_j⇒u_i;  j_k⇒ij;
        对于 i=j到 q 步长 1 执行
          始S[u_i+ij+j]⇒C;
            对于 k=1 到 j-1 步长 1 执行
            C-L[ij+k]*L[j_k+k]⇒C;
            若j<i 则否 1/ЄSQRT(C)⇒d;
            C*d⇒L[ij+j];
            u_i+p⇒u_i;  ij+i⇒ij
          终;
        j_k+j⇒j_k
      终
  终;
过程PT_i(i); 值 i ; {本过程计算T_i=L′_i L_YY^-1 L_i}
  始简变 d,ij,uv; 场H[1:q];
    pp2⇒uv;  0⇒T_i⇒ij;
    对于i_1=1到 q 步长 1 执行
      始 0⇒d;
        对于 j=1 到 i_1-1 步长 1 执行
          d+H[j]*L[ij+j]⇒d;
        (S[uv+ij+i]-d)/L[ij+i_1]⇒d⇒H[i_1];
        T_i+d*d⇒T_i,  ij+i_1⇒ij;  uv+p⇒uv
      终
  终;
过程Pivot(k); 值 k ; {以l_kk为轴的扫描}
  始场A[1:p+q]; 简变d,S_kk,k_0,kk;
    (k-1)*k/2⇒k_0;  k_0+k⇒kk;  1/S[kk]⇒S_kk;
    对于j=1到p+q步长 1 执行
      若j<k则始S[k_0+j]⇒A[j];  0⇒S[k_0+j]终
            否始S[kk]⇒A[j];  0⇒S[kk];
                kk+j⇒kk终;
    Enter[k]⇒A[k];  0⇒kk;
    对于i=1到p+q步长 1 执行
      始A[i]*S_kk⇒d;
        对于j=1到 i 步长 1 执行
          S[kk+j]-d*A[j]⇒S[kk+j];
```

```
          kk+i⇒kk
        终
    终;
  过程 PB; {计算回归系数B}
    始简变 d,uv;
      pp2⇒uv; 0⇒B;
      对于 u=1 到 q 步长 1 执行
        始0⇒d;
          对于 i=1到 p 步长 1 执行
            若 Enter[i]<0 则否
              始 S[uv+i]*(Vx[p+u]/Vx[i])⇒B[u,i];
                d+B[u,i]*Mx[i]⇒d终;
          Mx[p+u]-d⇒B[u,0]; uv+p+u⇒uv
        终{此处可印B (回归系数 估 值β̂′)}
    终;
  过程PC; {作判别分类并给出分类矩阵}
    始简变d,nu,uP,k0,yG,ymin,vmin;
      场 Y[1:G1];
      0⇒DM; 0⇒k0;
      对于u=1到G步长1执行
        始u⇒uP; ng[u]⇒DM[u,G1];
          若 u<G 则 ng[u]⇒nu 否 ng[u]+nn-n⇒nu;
          对于k=k0+1到k0+nu步长 1 执行
            始1₁₀18⇒ymin; 1⇒yG;
              对于v=1到G步长1 执行
                始若v=G则yG⇒d否
                    始 B[v,0]⇒d;
                      对于 i⇒1到 p 步长 1 执行
                        若Enter[i]<0则否
                          d+X[k,i]*B[v,i]⇒d;
                      yG-d⇒yG
                    终;
                  d⇒y[v]; ∈ABS(d-1)⇒d;
                  若d<ymin则始d⇒ymin; v⇒vmin终否
                终;
              Vmin⇒E[k];
              若k≤n则DM[u,vmin]+1⇒DM[u,vmin]
                    否0⇒uP;
                {此处可印k (观测序号), uP (原分类号),
```

```
                    vmin（计算分类号），ymin（|ŷg－1|的最小值）Y（各个y的预报值
                     ŷ）等分类明细表}
                终；
              对于v＝1到G1步长1执行
                DM[G1,v]＋DM[u,v]⇒DM[G1,v]；
              k0＋nu⇒k0
            终        {此处可印DM（分类矩阵）}
        终；          {至此说明部分完，其后开始计算}
      G－1⇒q；   G＋1⇒G1；   p*(p＋1)/2⇒pp2；
      PS；     {计算离差阵}
      若back＝0则Fa12⇒F12否0⇒F12；
      0⇒m⇒k⇒step；   1⇒Wilks；   －1⇒Enter；
  Lstep:step＋1⇒step；  0⇒Tmax；  1₁₀12⇒Tmin；
      PLLT；   {分解LYY＝LL′}
      0⇒ij；
      对于i＝1到p步长1执行
        若i＝k 则ij＋i⇒ij否
          始ij＋i⇒ij；   ∈ABS(S[ij])⇒d；
            若d＜0.00001则0⇒F[i]否
              始PTi(i)；      {计算Ti＝L′iL⁻¹YYLi}
                若Enter[i]＜0
                  则始Ti/(d－Ti)⇒Ti；
                      Ti*(n－q－(m＋1))/q⇒F[i]；
                      若Tmax＜Ti则始Ti⇒Tmax；  i⇒imax终否
                    终
                  否始Ti/d⇒Ti；
                      Ti*(n－q－m)/q⇒F[i]；
                      若Ti＜Tmin则始Ti⇒Tmin；  i⇒imin终否
                    终
              终
          终；
      若Tmin*(n－q－m)/q＜F12
        则始m－1⇒m；  imin⇒k；  Tmin⇒T12终
        否若Tmax*(n－q－(m＋1))/q＜(F12＋0.00001)
            则转L3否始m＋1⇒m；  imax⇒k；  Tmax⇒T12终；
      Wilks/(1＋T12)⇒Wilks；
      －(n－1－(m＋G)/2)*∈LN(Wilks)⇒X2；
      Pivot(k)；{以lkk为轴的扫描}
      －Enter[k]⇒Enter[k]；
```

{此处可印step（当步序号），Enter[k]*k（当步引入或剔除变量的序号，负表示剔除），m（已引入变量个数），F[k](F检验值)，Wilks，X_2（对Wilks 量的近似检验值）；如果需要，还可以调用过程 PB 和 PC，计算并印出当步的回归系数和分类明细表、分类矩阵等}

若m＝p则否转Lstep；

L_3:若F_{a12}＝F_{12}则否始F_{a12}⇒F_{12}；转Lstep终；

若m＝0则始－2⇒fail；转Lend终否0⇒fail；

PB；PC；{计算回归系数和分类明细表、分类矩阵}

Lend:

终；

（四）数值例子

原始数据及计算分类明细表 **表 1**

($F_a = 4.0$)

序号	x_1	x_2	x_3	x_4	原分类	计算分类	$\lVert\hat{y}_g-1\rVert$的最小值	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$
1	6	－24.0	17	－40	1	1	0.11	0.89	0.14	－0.04
2	－4	－16.0	13	54	1	1	0.33	0.67	－0.26	0.60
3	0	－23.0	5	－35	1	1	0.19	0.81	0.19	0.01
4	－90	－21.4	7	－15	1	1	0.23	0.77	0.09	0.14
5	－5	－18.5	15	18	1	1	0.32	0.68	－0.04	0.36
6	10	－18.0	14	50	1	1	0.11	0.89	－0.43	0.54
7	－8	－14.0	16	56	1	3	0.36	0.42	－0.06	0.64
8	－11	－18.5	25	－36	2	2	0.29	0.22	0.71	0.08
9	5	－11.5	19	0	2	2	0.00	－0.38	1.00	0.38
10	－10	－19.0	21	－42	2	2	0.26	0.23	0.74	0.04
11	20	－14.0	8	－20	2	2	0.01	－0.23	0.99	0.24
12	1	－13.0	26	－10	2	2	0.03	－0.27	0.97	0.31
13	－40	－15.0	22	－1	2	2	0.38	0.06	0.62	0.32
14	90	－13.0	17	50	3	3	0.38	0.24	0.13	0.62
15	0	－14.0	20	35	3	3	0.47	0.24	0.23	0.53
16	－90	－12.0	15	50	3	3	0.36	0.11	0.25	0.64
17	13	－15.4	18	42	3	3	0.46	0.48	－0.03	0.54
18	15	－16.6	20	20	3	1	0.55	0.45	0.14	0.41
19	－2	－18.5	15	－10	3	1	0.56	0.44	0.35	0.21
20	1	－15.2	16	21	3	3	0.57	0.28	0.29	0.43
21	25	－19.0	8	40	待 定	1	0.07	0.93	－0.40	0.47

分　类　矩　阵　　表 2

(F$_\alpha$ = 4.0)

原分类 \ 计算分类	1	2	3	合　计
1	6	0	1	7
2	0	6	0	6
3	2	0	5	7
合　计	8	6	6	20

均　值　与　标　准　差　　表 3

		x_1	x_2	x_3	x_4	y_1	y_2
分类均值	1	−13.00000	−19.27143	12.42857	12.57143	1	0
	2	−5.83333	−15.16667	20.16667	−18.16667	0	1
	3	3.85714	−14.95714	17.28571	29.71429	0	0
总 均 值		−4.95000	−16.53000	16.45000	9.35000	0.35000	0.30000
标准差$\sqrt{\frac{1}{n}l_{ii}}$		36.53556	3.44167	5.32424	33.66642	0.47697	0.45826

相　关　矩　阵　　表 4

	x_1	x_2	x_3	x_4	y_1	y_2
x_1	1.000000					
x_2	0.143002	1.000000				
x_3	0.154621	0.363370	1.000000			
x_4	0.126610	0.574836	−0.013989	1.000000		
y_1	−0.161680	−0.584501	−0.554243	0.070215	1.000000	
y_2	−0.015828	0.259325	0.456991	−0.535070	−0.480384	1.000000

各　步　的　主　要　结　果　　表 5

(F$_\alpha$ = 4.0)

步　数	1	2	3	4
舍选的变量	$+x_3$	$+x_4$	$+x_2$	$-x_3$
F 检验值	4.6677	4.6194	31.3711	3.4699
巳选变量个数	1	2	3	2
Wilks-Λ	0.64552	0.40922	0.07896	0.05398
χ^2	7.44	14.74	40.62	48.17