

单 龙 海 刘 乐 融 编 著

微 电 脑

通 用 多 元 统 计 分 析

BASIC 语言程序汇编

BASIC

691743-S

73.814

74-C57

阅 8 卷

73.814

24-C57

微电脑通用多元统计分析BASIC语言

程 序 汇 编

李龙海

刘乐融

编写

前 言

在国际上,近几十年来,多元统计方法已广泛而深入地渗透到各个领域,得到日益普遍的应用。为使多元统计方法能在更广大的领域被实际工作者所采用,我们将具有通用性的程序,汇编成集,据我们所知,这些程序已被广泛地用于生物学、地学、医学、气象学、天文学、农业和社会学等领域。

目前,国内出版物中已有一些用BCY语言和ALGOL-60语言编成的多元统计分析程序。但考虑到当今国内外推广普及的是微型计算机,而微型计算机所配有的国际上最流行BASIC语言,又是推广普及的对象。(在微型计算机上不配有BCY和ALGOL-60语言)。因此,出版用BASIC语言编成的多元统计分析程序集是必要的。

全部程序均在我国推广APPLE-II-PLUS或IBM-PC机上通过。这些程序不仅包括了国际上一些著名的多元统计程序包,如EMDP中的各种主要算法,而且还收入了一些较新的算法,如岭回归分析、图论聚论、模糊聚类和弗晰模型识别等。为便于使用,我们对其数学方法,仅作简要的概述,程序的使用则作较详细的说明,并附有例题。使用者只需给出具体数据,按例题的步骤进行,就能得到满意的答案。

西安交通大学计算机中心主任王以和副教授在百忙中审阅了本书。在此我们表示深切的谢意。

由于我们水平有限,加之时间仓促,书中的错误和不足在所难免,恳请同志们批评指正。

编 者

内 容 简 介

本书是国内第一部用BASIC语言编写，内容较新颖、完整的多元统计程序集。全部程序均在国内推广的APPLE—II和IBM—PC微型计算机上通过。主要内容不仅包括国际上一些著名的多元统计程序包，如BMDP中的各主要算法，而且还收入了一些较新的算法。书中每个程序都力求简明扼要，准确无误，并配有程序功能、数学方法概述、程序使用说明及应用举例。

本书既可作为生物、医学、地理、气象、地质、天文、农业、社会学等专业从事统计工作的科技人员和技术人员的实用手册，亦是高等院校有关学员、教师的参考读物。

目 录

1	多元线性回归分析	1
2	多元逐步回归及多项式逐步回归分析	8
3	岭回归分析	20
4	多元三角回归分析	34
5	费歇准则二类线性判别分析	42
6	贝叶斯准则多类逐步判别分析	54
7	多类训练迭代法	66
8	Q模式系统聚类分析	74
9	R模式系统聚类分析	87
10	模糊聚类	95
11	图论聚类	100
12	二维网格测点观测值多项式趋势分析	106
13	三维散布测点观测值多项式趋势分析	114
14	调和趋势分析	123
15	主分量分析	132
16	因子分析	150
17	主坐标分析	171
18	对应因子分析	179
19	非线性映射	188
20	弗晰模型识别的隶属度算法和贴近度算法	193
21	一维时间序列分析	206
	附录 基本BASIC语言简介	217
	主要参考文献	220

1. 多元线性回归分析

一、程序功能

本程序根据最小二乘估计的原理, 采用对系数矩阵求逆的方法, 得到回归方程的各回归系数, 各类偏差平方和, 复相关系数, 回归方程和回归系数的 F 检验值, 回归值及剩余值。

二、方法概要

设有自变量 $X_1, X_2, \dots, X_m$, 因变量 Y , 共作 n 次试验。若 Y 与 X_i ($i=1, 2, \dots, m$) 间有线性关系, 回归方程则为

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_m X_m \quad (1.1)$$

显而易见, 只要确定了 $b_0, b_1, \dots, b_m$ 各回归系数, 方程 (1.1) 也就确定了。

若我们把得到的试验数据用矩阵的形式表示为

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

由最小二乘知道, $b_0, b_1, b_2, \dots, b_m$ 应使得全部观察值 Y_n 与回归值 $\hat{Y}_n$ 的偏差平方和 Q 达到最小, 即使

$$\begin{aligned} Q &= \sum_{n=1}^n (Y_n - \hat{Y}_n)^2 \\ &= \sum_{n=1}^n (Y_n - b_0 - b_1 X_{n1} - b_2 X_{n2} - \dots - b_m X_{nm})^2 = \text{最小} \end{aligned}$$

根据微积分学中的极值原理, $b_0, b_1, b_2, \dots, b_m$ 应是下列方程组的解:

$$\begin{cases} \frac{\partial}{\partial b_0} Q = -2 \sum_{n=1}^n (Y_n - \hat{Y}_n) = 0 \\ \frac{\partial}{\partial b_j} Q = -2 \sum_{n=1}^n (Y_n - \hat{Y}_n) X_{nj} = 0 \quad (j=1, 2, \dots, m) \end{cases} \quad (1.2)$$

或

$$\begin{cases} \sum_{n=1}^n (Y_n - b_0 - b_1 X_{n1} - b_2 X_{n2} - \dots - b_m X_{nm}) = 0 \\ \sum_{n=1}^n (Y_n - b_0 - b_1 X_{n1} - b_2 X_{n2} - \dots - b_m X_{nm}) X_{nj} = 0 \quad (j=1, 2, 3, \dots, m) \end{cases}$$

显然方程组 (1.2) 的系数矩阵是对称矩阵, 若用 A 来表示它, 则 $A = XX'$ 即

$$A = \begin{pmatrix} N & \sum_{m=1}^m X_{m1} & \sum_{m=1}^m X_{m2} & \cdots & \sum_{m=1}^m X_{mn} \\ \sum_{m=1}^m X_{m1} & \sum_{m=1}^m X_{m1}^2 & \sum_{m=1}^m X_{m1}X_{m2} & \cdots & \sum_{m=1}^m X_{m1}X_{mn} \\ \sum_{m=1}^m X_{m2} & \sum_{m=1}^m X_{m1}X_{m2} & \sum_{m=1}^m X_{m2}^2 & \cdots & \sum_{m=1}^m X_{m2}X_{mn} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{m=1}^m X_{mn} & \sum_{m=1}^m X_{m1}X_{mn} & \sum_{m=1}^m X_{m2}X_{mn} & \cdots & \sum_{m=1}^m X_{mn}^2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1n} \\ 1 & X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix} \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1n} \\ 1 & X_{21} & X_{22} & \cdots & X_{2n} \\ 1 & X_{31} & X_{32} & \cdots & X_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix} = X'X$$

方程组 (1.2) 右端常数项矩阵 B 也可用矩阵 X 和 Y 来表示:

$$B = \begin{pmatrix} \sum_{m=1}^m y_m \\ \sum_{m=1}^m X_{m1}y_m \\ \sum_{m=1}^m X_{m2}y_m \\ \vdots \\ \sum_{m=1}^m X_{mn}y_m \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{m1} \\ X_{12} & X_{22} & \cdots & X_{m2} \\ \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{mn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = X'Y$$

这样方程组 (1.2) 的矩阵形式是

$$(X'X)b = X'Y$$

或 $Ab = B$

那么就有: $b = A^{-1}B$

若设

$$A^{-1} = C = \begin{pmatrix} C_{00} & C_{01} & C_{02} & \cdots & C_{0m} \\ C_{10} & C_{11} & C_{12} & \cdots & C_{1m} \\ C_{20} & C_{21} & C_{22} & \cdots & C_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n0} & C_{n1} & C_{n2} & \cdots & C_{nm} \end{pmatrix} = \begin{pmatrix} B_0 \\ B_1 \\ B_2 \\ \vdots \\ B_m \end{pmatrix}$$

就可得到各回归系数

$$b_k = C_{k0}B_0 + C_{k1}B_1 + C_{k2}B_2 + \cdots + C_{km}B_m \quad (k=0, 1, 2, \cdots, m)$$

回归方程及系数的显著性检验分别为:

回归平方和

$$S_{\text{回}} = \sum (\hat{Y}_k - \bar{Y})^2 \quad (k=1, 2, \cdots, n)$$

剩余平方和

$$S_{\text{剩}} = \sum_k (Y_k - \hat{Y}_k)^2 \quad (k=1, 2, \cdots, n)$$

总计平方和

$$S_{\text{总}} = \sum_k Y_k^2 - \frac{1}{n} (\sum_k Y_k)^2 \quad (k=1, 2, \dots, n)$$

回归方程的 F 检验值

$$F = \frac{S_{\text{回}} / (m+1)}{S_{\text{剩}} / (n-m)}$$

回归系数的 F 检验值

$$F = \frac{b_i^2 / C_{ii}}{S_{\text{剩}} / (n-m)} \quad (i=1, 2, \dots, m)$$

回归系数的 t 检验值

$$t = \frac{b_i}{\sqrt{C_{ii} S_{\text{剩}} / (n-m)}} \quad (i=1, 2, \dots, m)$$

三、程序使用说明

1. 输入格式

N1 样品数 (试验次数)

N2 自变量总数 + 1

上述变量由 INPUT 语句输入

A 二维数组用来存放原始数据

数组 A 由 DATA/READ 语句按行输入

2. 输出格式

N1 样品数

N2 自变量总数 + 1

X 建立回归方程的原始矩阵

XB(i) 各变量平均值 (i=1, 2, ..., N2)

B(i) 回归系数 (i=0, 1, 2, ..., N2-1)

R 复相关系数

SUR 剩余平方和

SUM 总计平方和

SRE 回归平方和

FO 回归方程的 F 检验

Ti 回归系数的 t 检验 (i=1, 2, ..., N2-1)

Fi 回归系数的 F 检验 (i=1, 2, ..., N2-1)

NO 样品号

Y 因变量实测值

Y< 回归值

E Y-Y< 剩余值

3. 附注

本程序要求 $N_1 > N_2$, 最好 $N_1 \gg N_2$.

四、例题

1. 原始数据

在表1.1中已知Y对 X_1 , X_2 , X_3 存在着线性关系, 求它们的回归方程, 并检验之。

2. 计算结果

表1.2为各变量的平均值及回归方程的系数。表1.3为回归系数的F、t显著性检验值。回归方程的计算值及剩余值在表1.4中。

$$R = 0.741174867$$

$$SUR = 5583.49985$$

$$SUM = 12389.6111$$

$$SRE = 6806.1112$$

$$FO = 5.68852031$$

需要说明的是: 在本节以及以后各节, 我们列出的计算结果主要供使用者校核程序之用, 因此, 计算结果的顺序和格式已经有别于计算机的输出结果。

表1.1

建立回归方程的数据矩阵

变量 样品	X_1	X_2	X_3	Y	变量 样品	X_1	X_2	X_3	Y
1	0.4	53	158	64	10	12.6	58	112	51
2	0.4	23	163	60	11	10.9	37	111	76
3	3.1	19	37	71	12	23.1	46	114	96
4	0.6	34	157	61	13	23.1	50	134	77
5	4.7	24	59	54	14	21.6	44	73	93
6	1.7	65	123	77	15	23.1	56	168	95
7	9.4	44	46	81	16	1.9	36	143	54
8	10.1	31	117	93	17	26.8	58	202	168
9	11.6	29	173	93	18	29.9	51	124	99

表1.2

回归方程系数及平均值

平均值	X_1	X_2	X_3	Y
	11.9444444	42.1111111	123	81.2777778
系数	b_0	b_1	b_2	b_3
	43.652198	1.78477968	-0.0833970941	0.161132702

表1.3

回归方程系数 t, F 检验

t 检验值	F 检验值
$t_1 = 3.31931289$	$F_1 = 11.0178381$
$t_2 = 0.199655206$	$F_2 = 0.0398622015$
$t_3 = 1.44299786$	$F_3 = 2.08224282$

表1.4

回归方程的计算值及剩余值

样品	实测值	回 归 值	剩 余 值	样品	实测值	回 归 值	剩 余 值
1	64	65.4050308	-1.40503076	10	51	79.3502531	-28.3502531
2	60	68.7126071	-8.71260709	11	76	77.9063339	-1.90633386
3	71	53.5623802	17.4376198	12	96	99.4134702	-3.41347018
4	61	67.1853988	-6.18539879	13	77	102.302536	-25.3025358
5	54	59.5459616	-5.54596161	14	93	90.2966541	2.70334592
6	77	61.0848347	15.9151654	15	95	107.280665	-12.2806651
7	81	64.1717591	16.8282409	16	54	67.0829605	-13.0829603
8	93	77.9456889	15.0543111	17	168	119.196068	48.8039324
9	93	89.8130839	3.18691608	18	99	112.744314	-13.744313

五、程序

```

0LIST
1030 INPUT 'N1/N2=';N1,N2
1010 DIM A(N1,N2),A1(N2,N2),W(N2),
B(N2)
1020 DIM C(N2),P(N2),Q(N2),X1(N2)
1030 PRINT 'X=';PRINT ;
1040 FOR X = 1 TO N1
1050 L1 = 10
1060 FOR Y = 1 TO N2
1070 READ A(X,Y)
1080 PRINT TAB( L1);A(X,Y);
1090 IF L1 = 40 THEN L1 = 0
1100 L1 = L1 + 10

```

```

1110 NEXT Y
1120 PRINT
1130 NEXT X
1140 PRINT
1150 FOR J = 1 TO N2
1160 L1 = 0
1170 FOR I = 1 TO N1
1180 L1 = L1 + A(I,J)
1190 NEXT I
1200 PRINT 'XB(';J;')=';L1 / N1
1210 NEXT J
1220 PRINT
1230 I = 0;J = 0

```

```

1240 A1(1,1) = N1
1250 FOR Y = 1 TO N2 - 1
1260 L = 0
1270 FOR X = 1 TO N1
1280 L = L + A(X,Y)
1290 NEXT X
1300 I = I + 1: J = I
1310 A1(J + 1,1) = L
1320 A1(1,J + 1) = L
1330 NEXT Y
1340 C = C + 1: W = W + 1
1350 FOR Y = 1 TO N2 - 1
1360 L = 0
1370 FOR X = 1 TO N1
1380 L = L + A(X,Y) * A(X,C)
1390 NEXT X
1400 A1(Y + 1,W + 1) = L
1410 NEXT Y
1420 IF C < > N2 - 1 THEN 1340
1430 L = 0: Y1 = 0: C1 = 0
1440 FOR X = 1 TO N1
1450 L = L + A(X,N2)
1460 Y1 = Y1 + A(X,N2) * 2
1470 NEXT X
1480 W(1) = L: C1 = W(1) * 2
1490 FOR Y = 1 TO N2 - 1
1500 L = 0
1510 FOR X = 1 TO N1
1520 L = L + A(X,N2) * A(X,Y)
1530 NEXT X
1540 W(Y + 1) = L
1550 NEXT Y
1560 FOR K = 1 TO N2
1570 W = 0
1580 FOR I = K TO N2
1590 FOR J = K TO N2
1600 IF ABS(A1(I,J)) > ABS(W)
THEN W = A1(I,J): ID = I: JD = J
1610 NEXT J
1620 NEXT I

```

```

1630 IF ABS(W) < = 1E - 12 THEN
PRINT "FAIL"
1640 IF ID = K THEN 1700
1650 FOR J = 1 TO N2
1660 Z = A1(ID,J)
1670 A1(ID,J) = A1(K,J)
1680 A1(K,J) = Z
1690 NEXT J
1700 IF JD = K THEN 1760
1710 FOR I = 1 TO N2
1720 Z = A1(I,JD)
1730 A1(I,JD) = A1(I,K)
1740 A1(I,K) = Z
1750 NEXT I
1760 P(K) = ID: Q(K) = JD
1770 FOR J = 1 TO N2
1780 IF J = K THEN B(J) = 1 / W: C(J) = 1: GOTO 1800
1790 B(J) = -A1(K,J) / W: C(J) = A1(J,K)
1800 A1(J,K) = 0: A1(K,J) = A1(J,K)
1810 NEXT J
1820 FOR I = 1 TO N2
1830 FOR J = 1 TO N2
1840 A1(I,J) = A1(I,J) + C(I) * B(J)
1850 NEXT J
1860 NEXT I
1870 NEXT K
1880 FOR K = N2 TO 1 STEP -1
1890 ID = P(K): JD = Q(K)
1900 IF ID = K THEN 1950
1910 FOR I = 1 TO N2
1920 Z = A1(ID,I): A1(I,ID) = A1(I,K)
1930 A1(I,K) = Z
1940 NEXT I
1950 IF JD = K THEN 2000
1960 FOR J = 1 TO N2
1970 Z = A1(JD,J): A1(JD,J) = A1(K,J)

```

```

1980 A1(K,J) = Z
1990 NEXT J
2000 NEXT K
2010 FOR X = 1 TO N2
2020 L = 0
2030 FOR Y = 1 TO N2
2040 L = L + A1(X,Y) * W(Y)
2050 NEXT Y
2060 X1(X) = L
2070 PRINT "B";X-1;"=";L
2080 NEXT X
2090 PRINT : PRINT
2100 L = 0
2110 FOR X = 1 TO N2
2120 L = L + X1(X) * W(X)
2130 NEXT X
2140 S = 0;S1 = 0
2150 S = Y1 - L;S1 = Y1 - (1 / N1)
    * C1
2160 PRINT "P=";(1 - S / S1) ^ (1
    / 2)
2170 PRINT
2180 PRINT "SURPLUS=";S
2190 PRINT "SUM=";S1
2200 PRINT "SREG=";S1 - S
2210 PRINT

```

```

2220 FD = ((S1 - S) / (N2 - 1)) / (
    S / (N1 - N2))
2230 PRINT "FD=";FD
2240 PRINT
2250 FOR Y = 2 TO N2
2260 P = X1(Y) ^ 2 / A1(Y,Y)
2270 T = P ^ (1 / 2) / (S / (N1 - N
    2)) ^ (1 / 2)
2280 F = P / (S / (N1 - N2))
2290 PRINT "T";Y-1;"=";T;
2300 PRINT TAB(20);"F";Y-1;"="
    ;F
2310 NEXT Y
2320 PRINT : PRINT
2330 PRINT "NO.";TAB(20);"Y";TA
    B(40);"Y";TAB(20);"E"
2340 FOR I = 1 TO N1
2350 P = 0
2360 FOR J = 1 TO N2 - 1
2370 P = P + X1(J+1) * A1(J,J)
2380 NEXT J
2390 T = P + X1(1)
2400 PRINT I;TAB(20);A1(I,N2);TA
    B(40);T;TAB(20);A1(I,N2) - T
2410 NEXT I
2420 END

```

2. 多元逐步回归与多项式 逐步回归分析

一、程序功能

本程序可直接用来进行多元线性回归分析, 多元线性逐步回归分析以及多项式逐步回归分析。并给出因变量的估计值, 剩余值和各分析值。

二、方法概要

对于因变量的每一观测值 Y_i , 对应有一组自变量 X_{ij} , 其中 $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$ 。在这种情况下可以建立以下回归模式

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1m} \\ 1 & X_{21} & \cdots & X_{2m} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{nm} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

或写成矩阵形式

$$Y = XA + \varepsilon \quad (2.1)$$

根据最小二乘准则, 可得正规方程如下

$$(X'X) \hat{A} = X'Y$$

其中 $\hat{A}$ 为系数向量 A 的最小二乘估计, 当 $(X'X)$ 有逆, 则可得

$$\hat{A} = (X'X)^{-1} X'Y \quad (2.2)$$

此时, 我们称

$$\hat{Y} = X\hat{A} \quad (2.3)$$

为回归值, 并称

$$\hat{\varepsilon} = Y - \hat{Y} \quad (2.4)$$

为剩余。

如果我们记 $(X'X \ X'Y)$ 为 W 。在求逆时若要在第 $L+1$ 步引入或删除第 r 号变量, 我们可以应用统一的变换公式

$$W_{ij}^{(L+1)} = \begin{cases} W_{ij}^{(L)} / W_{rr}^{(L)} & (i=r, j \neq r) \\ W_{ij}^{(L)} - W_{ri}^{(L)} W_{rj}^{(L)} / W_{rr}^{(L)} & (i \neq r, j \neq r) \\ 1 / W_{rr}^{(L)} & (i=r, j=r) \\ -W_{ir}^{(L)} / W_{rr}^{(L)} & (i \neq r, j=r) \end{cases} \quad (2.5)$$

式中 W_{ij} 表示矩阵 W 第 i 行第 j 列的元素, (L) 和 $(L+1)$ 分别表示在第 (L) 步和第 $(L+1)$ 步时矩阵 W 的元素. 利用公式 (2.5) 时, 前 L 步的次序可以任意颠倒而不影响经 L 步变换后的结果, 而且变换的中间结果给出了将部分变量引入回归方程时的回归系数.

为了确定引入或删除某一变量的标准, 可引用“方差贡献”的概念. 可以证明: 在第 L 步引入或删除第 j 号变量时它的方差贡献为

$$V_{i+}^{(L)} = V_{i-}^{(L)} = (W_{ij}^{(L)})^2 / W_{ii}^{(L)}$$

式中 $W_{ij}^{(L)}$ 表示第 L 步时矩阵 W 中第 i 行第 j 列 (最后一列) 的元素.

然后利用方差贡献计算统计量 $F^{(L)}$, 对于未引入的自变量

$$F_{i+}^{(L)} = (N-P-2) \frac{V_{i+}^{(L)}}{W_{jj}^{(L)} - V_{i+}^{(L)}} \quad (\text{引入}) \quad (2.6)$$

对已引入的自变量

$$F_{i-}^{(L)} = (N-P-1) \frac{V_{i-}^{(L)}}{W_{jj}^{(L)}} \quad (\text{剔除}) \quad (2.7)$$

式中 N 为样品, P 为回归方程中已引入的变量数.

我们规定一个统一的临界值 F_α , 若

$$\max F_{i+}^{(L)} > F_\alpha \quad (2.8)$$

则引进 X_i , 若

$$\min F_{i-}^{(L)} < F_\alpha \quad (2.9)$$

则剔除变量 X_i .

计算过程中引入或删除一个变量都称为一步, 在每步计算中, 总是先检验是否有已选变量需剔除, 仅当不需剔除时才考虑引入新变量. 在既无已选变量可剔除, 又无未选变量可引进时, 全部选变量工作才结束.

当引入 P 个自变量后, 可按下列公式计算复相关系数 R 与剩余标准差 d_r ,

$$R^2 = 1 - W_{jj}^{(L)} \quad (2.10)$$

$$d_r = \sigma_y \cdot \sqrt{W_{jj}^{(L)} / (N-P-1)} \quad (2.11)$$

式中 σ_y 为因变量 Y 的标准差.

三、程序使用说明

1. 输入格式

M 样品数

N	回归方程右端的项数
NN	变量总数 (即因变量数 + 1)
B	因变量个数
T2	各线性因子共同的最高幂指数
F	F检验的临界值
X1(I)	回归方程的交互积项中, 第一个变量的下标序号 (仅当 $T2 \geq 2$ 时)
X2(I)	回归方程的交互积项中, 第二个变量的下标序号 (仅当 $T2 \geq 2$ 时)

上述变量均由INPUT语句输入

X	二维数组用来存放样品数
Y	一维数组用来存放变量数

X, Y由语句DATA/READ按列输入。

2. 控制变量

Y	Y = 1打印或显示中间结果, 否则不打印或显示
---	--------------------------

3. 输出格式

M	样品数
N	回归方程右端的项数
NN	变量总数
B	因变量个数
T2	各线性因子共同的最高幂指数
X	原始数据
Y	控制变量
F	F检验的临界值
X1(I)	回归方程的交互积项中, 第一个变量的下标序号 (仅当 $T2 \geq 2$ 时)
X2(I)	回归方程的交互积项中, 第二个变量的下标序号 (仅当 $T2 \geq 2$ 时)
XB(i)	各变量的平均值 ($i = 1, 2, \dots, N$)
SG(i)	各变量的标准差 ($i = 1, 2, \dots, N$)
A	相关系数矩阵
MAX	每次矩阵变换后偏回归平方和最大的变量的偏回归平方和的值 ($Y = 1$ 时)
K	引入或删除变量号 ($Y = 1$ 时)
F1	引入变量的F值 ($Y = 1$ 时)
F2	删除变量的F值 ($Y = 1$ 时)
Y2	回归方程的剩余标准差 ($Y = 1$ 时)
R	回归方程的复相关系数
A - N	A矩阵的变换矩阵 ($Y = 1$ 时)
B(i)	回归系数 (i 回归方程右端的项号 $i = 2, 3, \dots, N$)
B(0)	回归方程的常数项
S	回归方程的最终剩余标准差
FF	回归方程的最终F计算值

NO.	样品号
Y	自变量实测值
Y _A	自变量回归值
E	Y - Y _A 的差值

四、例题

1. 原始数据

样品的原始数据见表2.1。分别作线性逐步回归方程，多项式逐步回归方程及多元线性回归方程。

2. 计算结果

(1) 形式为

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$$

的线性逐步回归分析结果。

此时，各输入变量的取值为

$$\begin{array}{llll} M = 13 & N = 5 & NN = 5 & B = 4 \\ T2 = 1 & Y = 1 & F = 3.28 & \end{array}$$

表2.2各变量的均值与标准差，表2.3相关系数矩阵仅给出上三角部分，表2.4为各回归值及剩余值。

$$R = 0.989220042 \quad S = 2.41225434 \quad FF = 228.168612$$

最终回归方程

$$\hat{Y} = 52.5854275 + 1.46761888X_1 + 0.6619495X_2$$

(2) 形式为

$$\begin{aligned} Y = & b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_1^2 + b_6 X_2^2 + b_7 X_3^2 + b_8 X_4^2 \\ & + b_9 X_1 X_3 + b_{10} X_1 X_4 + b_{11} X_2 X_3 + b_{12} X_2 X_4 \end{aligned}$$

的多项式逐步回归分析结果

此时，各输入变量的取值为

$$\begin{array}{lllll} M = 13 & N = 13 & NN = 5 & E = 4 & T2 = 2 \\ X1(1) = 1 & X1(2) = 1 & X1(3) = 2 & X1(4) = 2 & X2(1) = 3 \\ X2(2) = 4 & X2(3) = 3 & X2(4) = 4 & Y = 1 & F = 0.8 \end{array}$$

表2.7为各变量的均值与标准差，表2.6为相关系数矩阵，表2.5各回归值与剩余值

$$R = 0.993616326 \quad S = 2.07770692 \quad FF = 155.151204$$

最终回归方程

$$\hat{Y} = 39.2593605 + 1.59745248X_1 + 1.18494101X_2 - 6.03371855E$$

$$- 03X_3 + 4.53641892E - 03X_2X_3$$

(3) 形式为

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

的多元线性回归分析结果。

此时，各变量的取值除 $F = 0$ ，其它均如(1)所示。

表2.8为回归方程的系数

$$R = 0.991080431$$

$$S = 2.45439519$$

$$FF = 110.615275$$

表2.9为回归值及剩余值。

表2.1

用于建立回归方程的原始数据

变量 样品	X_1	X_2	X_3	X_4	Y	变量 样品	X_1	X_2	X_3	X_4	Y
1	7	26	6	60	78.5	8	1	31	22	44	72.5
2	1	29	15	52	74.3	9	2	54	18	22	93.09
3	11	56	8	20	104.19	10	21	47	4	26	115.9
4	11	31	8	47	87.59	11	1	40	23	34	83.8
5	7	52	6	33	95.9	12	11	66	9	12	118.3
9	11	55	9	22	109.19	13	10	68	8	12	109.4
7	8	71	17	6	102.69						

表2.2

变量均值及标准差

	X_1	X_2	X_3	X_4	Y
平均值	7.46153847	48.1538462	11.7692308	30	95.4115385
标准差	20.377212	53.904474	22.1880079	57.9827562	52.0922526