

罗积玉 邢 瑛 苏显康 罗昌荣著

微机用多元 统计分析软件

四川科学技术出版社

微机用多元统计分析软件

罗积玉 邢 瑛 著
苏显康 罗昌荣

四川科学技术出版社

一九八六年·成都

责任编辑：梅 红

封面设计：魏天禄

微机用多元统计分析软件

罗积玉 邢 瑛 著
苏显康 罗昌荣

出版：四川科学技术出版社
印刷：自贡新华印刷厂
发行：四川省新华书店
开本：850×1168毫米 1/32
印张：17.25 插页：4
字数：428千
印数：1—3,250
版次：1986年6月 第一版
印次：1986年6月第一次印刷
书号：15298·101
定价：3.80元

内 容 简 介

本书包括了23个多元统计分析软件，比较系统、全面地反映了多元统计分析领域。内容包括方差、协方差分析、回归分析、主成分分析、因子分析、聚类分析、判别分析、对应分析、典型相关分析、非线性映射、一维时间序列分析、马尔科夫概型分析。并且还吸收了一些较新的算法。为便于读者使用，每种方法都对数学原理作了简述、对使用方法作了说明，并附有算例。

本书全部软件均在IBM PC—XT(长城100)微型电子计算机上实现，系用微机 BASIC 语言书写，输入、输出说明及程序说明均用汉字说明，面向用户，结构清晰。

本书可以提供给从事于微型计算机工作的同志，从事于多元统计分析应用开发、应用研究的同志，从事于统计分析处理的经济、人口、办公自动化系统、地质、地震、气象、生态、医学、药学、农业、林业、水文等方面专业研究人员使用，也可以作为高等院校有关专业研究生、高年级学生从事计算机应用和多元统计分析应用的参考书。

前 言

随着微型电子计算机的发展和应用普及，在统计领域、计划经济领域、人口分析领域、办公自动化系统、地质、地震、生态、医学、药学、林业、农业、气象、水文等方面都需要对大量统计数据进行统计分析，用多因素、多指标来全面反映事物的本质，而多元统计分析方法是一个处理这种随机现象的科学方法，也是一个成效显著的实用统计分析方法。

多元统计分析是近二十年来随着电子计算机的发展和应用而迅速发展起来的数理统计方法，在应用数学领域，它已经逐渐形成了自己的理论体系，在应用方面，已经引起了我国各个领域的应用工作者、研究工作者越来越广泛的注意，我国的计算机软件人员也逐渐加强对多元统计分析软件的开发。但随着微型计算机的发展和普及，应用的深入，更需要一本便于推广应用，便于普及的微机用多元统计分析软件书籍，鉴于这种情况，我们将近几年来在微型计算机上开发的软件经过整理、付印成册，以适应微型计算机应用发展的需要。

本书全部软件均在IBM PC-XT（长城100）微型电子计算机上实现，系用微机BASIC语言书写。输入、输出说明和程序说明部分均用汉字说明，采用汉字版本印刷，这套软件可以在其它型号（如Eagle、Apple II）微型电子计算机上安装、使用，仅需对输入参数部分作少许修改。

本书包括23个软件，比较系统、全面地反映了多元统计分析领域，而且还收入了一些较新的算法，如岭回归等。全部软件采

用结构化程序设计、面向用户、采用屏格式设计特点，使输入参数方便、输出结果格式清晰，每个程序开始都附有比较完整的汉字说明，清晰、易读。软件已通过大量应用实例的考证，运行可靠。为方便读者使用，对每种方法的数学原理、使用方法作了说明。这本书对从事于微型计算机工作的同志，从事于多元统计分析应用开发、应用研究的同志都会有帮助的。

按照BASIC标准文本，提供原始数据的方式都用 DATA 语句，但为了方便使用，便于原始数据的检查、校对和保存，也可采用文件管理方式存贮、管理数据，关于文件的建立、使用以及和源程序的接口详见附录。

在软件开发和编写过程中，得到了国家计委计算中心姜均露同志和张大洋同志的关心和支持，在此表示感谢。

由于水平有限，缺点错误难免，愿望读者指正。

编 者

1984年10月

目 录

前言

§ 1 单因素方差分析·····	1
§ 2 多因素方差分析·····	9
§ 3 协方差分析·····	22
§ 4 多元线性回归分析·····	36
§ 5 多元逐步回归分析·····	59
§ 6 非规则测点的多项式趋势分析·····	87
§ 7 岭回归分析·····	109
§ 8 多元三角回归分析·····	137
§ 9 主成分分析·····	155
§ 10 因子分析·····	175
§ 11 系统聚类分析·····	211
§ 12 动态聚类分析·····	243
§ 13 模糊聚类分析·····	256
§ 14 图论聚类分析·····	273
§ 15 两组判别分析·····	287
§ 16 多组判别分析·····	309
§ 17 逐步判别分析·····	333
§ 18 训练迭代法·····	363
§ 19 对应分析·····	379
§ 20 典型相关分析·····	397

§ 21 非线性映射.....	430
§ 22 一维时间序列分析.....	453
§ 23 马尔科夫概型分析.....	478
附录1.....	516
附录2.....	535
主要参考文献	543

§ 1 单因素方差分析

一、程序应用范围

方差分析是分析试验数据的一种方法。一个事物中有许多因素互相制约、互相矛盾、互相依存。方差分析就是要找出各个因素及各因素之间交互作用的影响；充分利用试验结果，对所关心的事物（因素影响）作出合理的推断。

方差分析首先应用于农业试验，它在农业、工业、生物学和医学、经济等许多方面，均具有广泛的应用。

二、单因素方差分析功能

对一组实验观察数据（假定满足相互独立地来自等方差的正态的条件）进行重复试验单因素方差分析。

三、方法原理

在单因素方差分析中，一般假设对 K 个不同水平分别进行

$n_1, n_2, \dots, n_k$ 次重复试验，得到 $N = \sum_{i=1}^k n_i$ 个实验观察数据

X_{ij} ($i=1, 2, \dots, k; j=1, 2, \dots, n_i$)。假设这些数据可作线性分解：

$$X_{ij} = u + \alpha_i + \varepsilon_{ij} \quad (i=1, 2, \dots, k) \quad (1.1)$$

其中 u 表示总体均值， α_i 表示各个不同水平对实验观测数据的影响， ε_{ij} 表示各种随机影响引起的随机误差，且假定它们相互独立并服从分布 $N(0, \sigma^2)$ 。

对实验数据进行方差分析就是将总离差平方和 S_t 分解为各个不同水平效应的离差平方和 S_b 及随机误差引起的离差平方和 S_w ，比较它们的大小，以检验统计。假设

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0 \quad (1.2)$$

易知 $S_t = S_b + S_w$ ，其中

$$S_t = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2, \quad S_b = \sum_{i=1}^k n_i (\bar{x}_{i.} - \bar{x}_{..})^2,$$

$$S_w = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2,$$

$$\bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad \bar{x}_{..} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} = \frac{1}{N} \sum_{i=1}^k n_i \bar{x}_{i.}$$

对(1.1)式线性模型的各项参数从最小平方和估计

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - u - \alpha_i)^2 = \min \quad \text{出发。}$$

设

$$\sum_{i=1}^k n_i \alpha_i = 0,$$

可得：

$$u = \bar{x}_{..}, \quad \alpha_i = \bar{x}_{i.} - \bar{x}_{..} \quad (1.3)$$

在(1.1)式线性模型假定下，统计量 S_i/σ^2 ， S_b/σ^2 ， S_w/σ^2 分别服从具有自由度 $N-1$ ， $K-1$ ， $N-K$ 的 X^2 分布。统计量

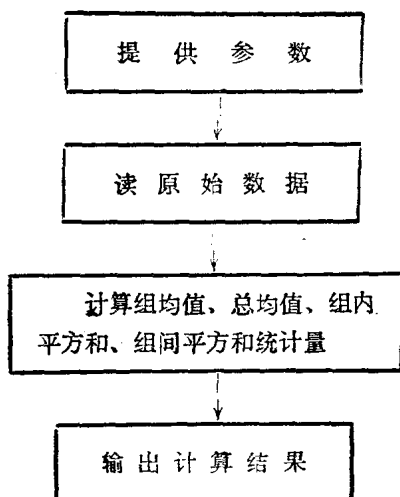
$$F = \frac{S_b/(K-1)}{S_w/(N-K)}$$

服从自由度为 $K-1$ 和 $N-K$ 的 F 分布。由此可对假设 H_0 进行显著性检验。

在这里首先计算组内均值 $\bar{x}_{i.}$ ，总均值 $\bar{x}_{..}$ ，然后计算组内平

方和 S_w ，组间平方和 S_b ，最后计算组内平方和的均方 m_{sb} ，组间平方和的均方 m_{sw} 及统计量 F 。

四、程序结构图



五、程序使用说明

1. 输入格式

程序运行时，通过屏幕显示提示提供下列参数：

水平数 K

各水平的重复实验次数 n_i 。

然后按水平1的各个观测数据，水平2的各个观测数据，……
水平 K 的各个观测数据的次序读原始数据。

2. 标识符说明见源程序说明部分

3. 输出格式

(1) 组均值

(2) 总均值

(3) 以表格形式输出: 组内平方和、自由度、均方, 组间平方和、自由度、均方; 总平方和、自由度及统计量F。

六、计算实例

1. 原始数据

水平数, $K = 3$

第1组观测重复次数, $NI(1) = 4$

第2组观测重复次数, $NI(2) = 5$

第3组观测重复次数, $NI(3) = 3$

表1.1 原始数据

水平1	0.365	0.255	0.195	0.215	
水平2	0.270	0.275	0.240	0.265	0.185
水平3	0.305	0.250	0.225		

组均值 = 0.258 0.247 0.260

总均值 = .25375

2. 计算结果

表1.2 方差分析表

方差源	平方和	自由度	均方	统计量
组间差	SB = 0.000401	NB = 2	MSB = 0.000201	F = 0.06906
组内差	SW = 0.026155	NW = 9	MSW = 0.002906	
总离差	ST = 0.026556	NT = 11		

```

10 REM *****
20 REM *
30 REM * ***单因素方差分析***
40 REM *
50 REM * 程序名: SANVOA
60 REM *
70 REM * 1. 主要功能:

```

```

80 REM *
90 REM *      对一组实验观测数据(假定满足相互独立地来*
100 REM *      自等方差的正态总体的条件进行重复试验单因素方*
    *      差分析。
110 REM *
120 REM *      2. 主要标识符说明:
130 REM *
140 REM *      X(N) -- 按各水平次序依次存放实验观测数据。*
150 REM *      K      -- 水平数。
160 REM *      NI(K) -- 按各水平次序依次存放重复试验次数。*
170 REM *      MI(K) -- 按各水平次序依次存放各组均值。
180 REM *      SB, SW, ST -- 分别存放不同水平之间的影响及*
190 REM *      各组内随机因素影响, 总离差平方*
    *      方和。
200 REM *      NB, NW, NT -- 分别存放对应离差平方和的自由*
205 REM *      度。
210 REM *      MSB, MSW -- 分别存放不同水平之间的影响及*
220 REM *      各组内随机因素影响的均方差。*
230 REM *      f -- 统计量。
240 REM *
250 REM *      3. 参数提供方法:
260 REM *
270 REM *      程序开始运行时, 请从屏幕提供下面参数:
280 REM *      水平数, 每个水平重复试验次数。
290 REM *
300 REM *      4. 原始数据排列说明:
310 REM *
320 REM *      按水平依次存放原始数据, 即
330 REM *      X11, X12, ..., X1n1
340 REM *      X21, X22, ..., X2n2
350 REM *      .....
360 REM *      Xk1, Xk2, ..., Xknk

```

```

370 REM *
380 REM * 5. 主要输出结果:
390 REM *
400 REM * 组均值, 总均值, 方差分析表.
410 REM *****
420 WIDTH "lpt1: ", 132
430 CLS
440 LOCATE 3, 14: INPUT "水平数=", K
450 DIM NI(K), MI(K)
460 FOR I=1 TO K
470 PRINT" 第", I, : INPUT "个水平重复试验次数=", NI(I)
480 N=N+NI(I)
490 NEXT I
500 DIM X(N)
510 FOR I=1 TO N
520 READ X(I)
530 NEXT I
540 REM *****
550 F=0: N=0 : SW=0 : MX=0
560 FOR I=1 TO K
570 MI(I)=0
580 NEXT I
590 FOR I=1 TO K
600 FOR J=1 TO NI(I)
610 MI(I)=MI(I)+X(N+J)
620 SW=SW+X(N+J)^2
630 NEXT J
640 MX=MX+MI(I)
650 N=N+NI(I)
660 MI(I)=MI(I)/NI(I)
670 F=F+NI(I)*MI(I)^2
680 NEXT I

```

```

690  MX = MX/N
700  SW = SW - F
710  SB = F - N * MX ^ 2
720  ST = SW + SB
730  NW = N - K
740  NB = K - 1
750  NT = N - 1
760  MSW = SW/NW
770  MSB = SB/NB
780  F = MSB/MSW
790  REM *****
800  REM *           输  出  计  算  结  果           *
810  REM *****
820  LPRINT TAB(40); " < 单因素方差分析计算结果 > "
830  LPRINT " ":LPRINT "
840  LPRINT TAB(36); "水平数, K = ", K
850  LPRINT " "
860  FOR I=1 TO K
870  LPRINT TAB(36); "第", I, "组观测重复次数, NI( ", I, " )
      = ", NI(I)
880  NEXT I
890  LPRINT " ", LPRINT "
900  LPRINT TAB(45); "原始数据 ", LPRINT " ", LPRINT " "
910  N = 0
920  FOR J=1 TO K
930  LPRINT TAB(30);
940  FOR I=1 TO NI(J)
950  LPRINT USING "###.#", X(I+N);
960  NEXT I
970  N = N + NI(J); LPRINT " "
980  NEXT J
990  LPRINT " ", LPRINT " "

```

```

1000 LPRINT TAB(30); "组均值 = ",
1010 FOR I=1 TO K
1020 LPRINT USING "###.### ", MI(I),
1030 NEXT I
1040 LPRINT " "; LPRINT " "
1050 LPRINT TAB(30); "总均值 = ", MX
1060 LPRINT " "; LPRINT " ";
1070 LPRINT TAB(45); "方差分析表 "
1080 LPRINT TAB(30); , FOR I=1 TO 10:LPRINT"---
----- ", , NEXT I
1090 LPRINT TAB(30); "1", "方差源 "; TAB(41); "平方
和 "; TAB(52); "自由度 ";
1100 LPRINT TAB(62); "均方 "; TAB(76); "统计量"; TAB
(89); "1"
1110 LPRINT TAB(30); , FOR I=1 TO 10; LPRINT"---
--- ", , NEXT I
1120 LPRINT TAB(30); "1 "; "组间差 SB= ", , LPRINT
USING "###.#####"; SB,
1130 LPRINT TAB(52); "NB=", , LPRINT USING "##
# "; NB,
1140 LPRINT TAB(60); "MSB=", , LPRINT USING "##.
##### "; MSB,
1150 LPRINT TAB(76); "F=", , LPRINT USING "##.
##### "; F,
1160 LPRINT TAB(89); "1 "
1170 LPRINT TAB(30); "1 "; "组内差 SW=", , LPRINT
USING "###.#####"; SW,
1180 LPRINT TAB(52); "NW=", , LPRINT USING "##
# "; NW,
1190 LPRINT TAB(60); "MSW=", , LPRINT USING "##
.#####"; MSW,
1200 LPRINT TAB(89); "1"

```

```

1210 LPRINT TAB (30); "1 "; "总离差 ST='"; LPRINT
    USING "###.#####" , ST;
1220 LPRINT TAB (52); "NT= "; LPRINT USING "##
    # "; NT;
1230 LPRINT TAB (89); "1"
1240 LPRINT TAB (30); , FOR I=1 TO 10; LPRINT " - -
    - - - "; , NEXT I
1250 LPRINT " "; LPRINT " "
1260 LPRINT TAB (40); "*** 计算结束 ***"
1270 END

```

§ 2 多因素方差分析

一、程序应用范围

同单因素方差分析部分。

二、功能

利用正交变换分解和计算平方和的方法, 对实验观测数据 (假定满足相互独立地来自等方差的正态总体的条件) 进行无重复试验多因素方差分析。

三、方法原理

对于含有 r (≥ 2) 个因素无重复试验的方差分析, 假定每个

因素分别有 $n_1, n_2, \dots, n_r$ 个水平, 则共有 $N = \prod_{j=1}^r n_j$ ($2 \cdot 1$)

个观测, 每个观测可表为 $X_{i_1, i_2, \dots, i_r}$, 其中 $i_j = 1, 2, \dots, n_j, j = 1,$