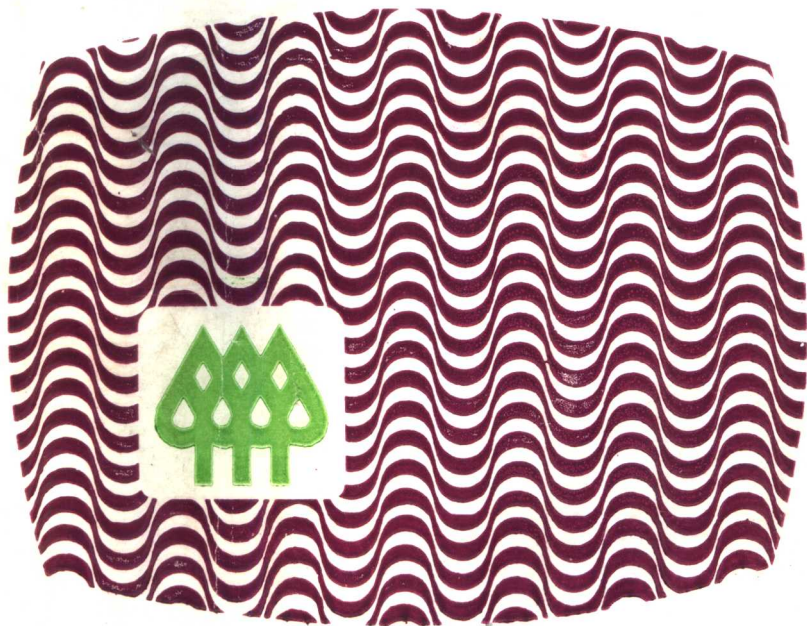


多元统计分析

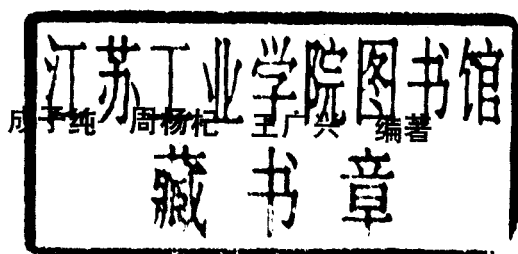
电标程序与应用

成子纯 周扬杞 王广兴编著



中国林业出版社

多元统计分析电算程序与应用



中国林业出版社

多元统计分析电算程序与应用

成子纯 周杨杞 王广兴 编著

中国林业出版社出版发行（北京西城区刘海胡同7号）

中南林学院印刷厂印刷

787×1092毫米32开本 18.25印张 400千字

1990年10月第一版 1990年10月第一次印刷

印数1—3000册 定价：9.50元

（京）第033号 ISBN

7—5038—0770—9/S·0374

前 言

农林系统在处理大量的调查统计资料和科研数据时，数理统计方法应用最多。如果利用电子计算机来完成这项工作，既可以提高工效，又可以提高水平。本书正是将主要的数理统计方法编成电算程序，并结合实例来解决农林系统目前各类数据处理方面的许多问题。书中汇集的程序主要适用于科研资料的研究和分析。在科学研究中常常涉及多个因素的具体问题，要弄清每一个因素的意义，要弄清各因素之间的关系，要阐明多个因素对事物本质的影响，多元统计方法是最有效的。于是我们重点将多元统计方法编成了电算程序，使过去许多复杂的计算分析工作简化和优化。

目前，在农林系统中，由于多种原因，尚未普遍使用电子计算机。为了尽快推广使用电子计算机，首先推广轻便价廉的微型机比较适当。特别是林业调查规划、企业管理、林业科研及木材生产等方面都可以借助微型机把林业计算技术的水平和效率大大提高一步。农业系统数理统计和多元分析等技术手段，近10多年来也发展较快，如果进一步将电算与现代农业技术紧密地配合，直接为生产服务，将会产生更好的社会效益和经济效益。

本书正是根据上述需要，编写了下列内容：统计数据处理，统计特征数的计算，抽样技术，方差分析，回归分析，

判别分析，聚类分析，相关分析，数量化理论I—IV等专业程序。此外，本书所汇编的程序均采用国际通用简单易懂的BASIC语言。它是一种得到广泛应用的程序设计语言，易学易用，也是其它算法语言的基础，对广大基层单位是很适用的。

在农林系统尚有许多同志不能熟练运用各种数理统计方法，本书为这些同志提供了方便，他们只要按照本书程序所给出的步骤去操作按键，就能解决各类计算问题，而且我们尽力将农林专业计算的各类要求包括进去，供更多的用户使用。

在编著本书的过程中，我们参考应用了部分专业程序的成果。我们自己编写的大部分程序，都有其功能，数学方法，程序清单，使用方法和举例的详细说明，力求实用可靠。由于作者的资料和水平有限，错误和遗漏之处在所难免，恳请有关同志批评指正。

编 者

1990年9月于株洲

目 录

第一章 数据处理	(1)
§ 1.1 数据排序.....	(1)
§ 1.2 频数分布与直方图.....	(5)
§ 1.3 累计频率分布.....	(9)
§ 1.4 样本均值、标准差和变异系数.....	(12)
第二章 抽样技术	(15)
§ 2.1 简单随机抽样.....	(15)
§ 2.2 分层抽样.....	(21)
§ 2.3 成数抽样.....	(30)
§ 2.4 3 P 抽样.....	(41)
第三章 区间估计与假设检验	(48)
§ 3.1 总体均值的区间估计.....	(48)
§ 3.2 总体方差的区间估计.....	(52)
§ 3.3 总体平均数和频率的假设检验.....	(56)
§ 3.4 两样本间的方差齐性和均值显著性检验.....	(62)
第四章 方差分析	(71)
§ 4.1 单因素方差分析.....	(71)
§ 4.2 双因素方差分析 (不考虑交互作用)	(85)

§ 4.3	双因素方差分析 (考虑交互作用)	(100)
第五章	线性代数与规划	(118)
§ 5.1	高斯—约旦法解线性方程组	(118)
§ 5.2	高斯—约旦消去法求逆矩阵	(124)
§ 5.3	求实对称矩阵特征值和特征向量	(130)
§ 5.4	解线性规划问题 (二阶段法)	(139)
第六章	回归分析	(153)
§ 6.1	多元线性逐步回归	(153)
§ 6.2	扩展变量的多元线性逐步回归	(174)
§ 6.3	多元线性模型	(192)
第七章	判别分析	(222)
§ 7.1	费歇准则下的两类判别分析	(222)
§ 7.2	贝叶斯准则下的多类逐步判别	(238)
§ 7.3	多类训练迭代法	(263)
第八章	聚类分析	(276)
§ 8.1	系统聚类分析 (R型和Q型)	(276)
§ 8.2	动态聚类	(322)
§ 8.3	图论聚类	(333)
§ 8.4	模糊聚类	(343)
§ 8.5	有序样品的聚类	(356)
第九章	相关分析	(371)
§ 9.1	典型相关分析	(371)

§ 9.2	主成分分析.....	(394)
§ 9.3	因子分析.....	(414)
§ 9.4	对应因子分析.....	(432)
第十章	数量化理论.....	(481)
§ 10.1	数量化理论 I	(481)
§ 10.2	数量化理论 II	(503)
§ 10.3	数量化理论 III	(535)
§ 10.4	数量化理论 IV	(556)
§ 10.5	参考文献.....	(557)

第一章 数据处理

§ 1.1 数据排序

一、功 能

要对事物在数量上有客观的认识,唯一可靠的办法就是通过调查、实验或实践获得数据,然后进行分析研究。数据是科学实验的分析依据,一般来说数据可分为连续性的变量数据和不连续性的计数数据。但不管是连续性的或不连续性的数据,人们为了对它进行全面分析,常常需要将杂乱无序的调查试验数据进行系统整理。由于需要不同,整理方法也不同。例如对一批随机取样的调查数据,为了了解它的变化幅度和顺序,常需要把一批杂乱的数据按大小顺序重新排列,从这些排列的数据中可以一目了然地知道最大值、最小值、数据的变化顺序和数据分布规律等,如果用手工来排列,则相当烦琐,用电子计算机就可以将大量数据很快地按要求顺序排列并打印出来。

二、数学方法

用堆法分类法,从第一个数开始,将第 i 个数与第 $2i$ 个数、第 $2i+1$ 个数组成一组。将 n 个数数据分入各组,然后从最后一组开始,逐组把组内数据按从小到大的顺序重新排列,

如此可将最小的数据调到第一个位置。

同理，再将第二个位置上的数作同理调换，依次将最小数逐个调到最前面，如此最多进行 $n - 2$ 次，就可以将数据排列好。

三、程序说明

N 为数据个数；

$A(N - 1)$ 为数据值。

第110语句，如果改为 $A(J) \leq A(K)$ 时，则数据是由大到小的顺序排列。

四、程序

```
10 PRINT "NUMBER OF DATA"  
20 INPUT N  
30 PRINT  
40 DIM A(N - 1)  
50 FOR I = 0 TO N - 1  
60 READ A(I)  
70 NEXT I  
80 FOR I = 0 TO N - 2  
90 FOR J = N - 1 TO I + 1 STEP  
100 LET K = INT [(J + I - 1) / 2]  
110 IF A(J) >= A(K) THEN 150  
120 LET A = A (J)  
130 LET A(J) = A(K)
```

```

140 LET A(K) = A
150 NEXT J
160 NEXT I
170 FOR I=0 TO N-2
    10 PRINT A(I); ", ",
180 NEXT I
200 PRINT A(N-1); ", "
210 END
220 DATA...

```

五、举 例

例如对某水稻试验区140行（行长4尺）水稻试验产量（单位：克）的结果如表1—1。

表1—1 140行水稻产量表

177	215	19	97	123	159	245	119.5	119.7
131.1	149	152	167	104	161	214	125	175
219.2	118	192	176	175	95	136	199	116
165	214.4	95	158	83	137.8	80	138	151
187	126	196.6	134	206	137	98	97.5	129
143	179	174	159.9	165	136	108	101	141
148	168.5	163	176	102	194.1	145	173	75
130	149	150.5	161	155.4	111	158.6	131	189
91	142	140	154	152	163	123	205	149.5
155	131	209	183	97.3	119	181	149.1	187.7
131.5	215	111	186	118	150.2	155	197.2	116
254.4	239	160	172	179	151	198.5	124	179.3
135.5	184	168	169					

172.5 181 188 211 197.3 175.4 122 151.9
 171 166 175.5 143 190 213 192.5 231 163.3
 159.1 158 159 177.5 147 194 227 141.5
 169.1 124 159.2

将这批数据用置数语句输入电子计算机，则

用220 DATA (再将表1—1数据输入)

RUN

NUMBER OF DATA? 140

其结果按顺序排于下表1—2中。

表1—2 输出结果

75	80	83	91	95	95	97	97.3	97.5	98	101
102	104	108	111	111	116	116	118	118	119	
119.5	119.7	122	123	123	124	124	125	126		
129	130	131	131	131.1	131.5	134	135.5	136		
136	137	137.8	138	140	141	141.5	142	143		
143	145	147	148	149	149	149.1	149.5	150.2		
150.5	151	151	151.9	152	152	154	155	155		
155.4	158	158	158.6	159	159	159.1	159.2			
159.9	160	161	161	163	163	163.3	165	165		
166	167	168	168.5	169	169.1	171	172	172.5		
173	174	175	175	175.4	175.5	176	176	177		
177.5	179	179	179.3	181	181	183	184.4	186		
187	187.7	188	189	190	192	192.5	194.1	194		
196.6	197	197.2	197.3	198.5	199	205	206			
209	211	213	214	214.4	215	215	219.2	227		
231	239	245	254.4							

§ 1.2 频数分布与直方图

一、功 能

如果一个大样本或一大批调查数据均是一大堆杂乱无章的变量,则很难获得明晰的印象,因此需要把一大堆变量分组,使各个变量按照数值大小归并到应属的组内,各组变量个数多少不等,较能集中表现其数据内的变化规律,这种分组后的分组数列便于观察分析,便于图示,便于计算其代表值。

二、数学方法

将数据分组,一般分四个步骤:

1. 从所有变量中找出最大变量及最小变量,再求其差异。即求全距 = 最大值 - 最小值。

2. 根据全距分为若干组,每组距离相等,称为组距。组距小则组数多,组距大则组数少,一般分 8—15 组为宜,或者组距的大小不超过标准差的 $1/4$,如果组数嫌多,则可使组距为 $1/3$ 标准差以至 $1/2$ 标准差。

一般可采用公式:

$$G = 1.87 (N - 1)^{1/4}$$

算出应分的组数。

3. 确定好各组的组限。

4. 根据组限算出各组组频数。

三、程序说明

N 为数据个数:

V为最小值;

V_1 为最大值;

G为组数;

R为组距;

B (G-1) 为组频数。

四、程 序

```
10 PRINT "NUMBERS OF DATA"  
20 INPUT "N=" ; N  
30 PRINT  
40 READ V  
50  $V_1 = V$   
60 FOR I=2 TO N  
    0 READ A  
    80 IF  $A > V$  THEN 100  
    0  $V = A$   
    100 IF  $A < V_1$  THEN 120  
    110  $V_1 = A$   
120 NEXT I  
130 RESTORE  
140  $G = \text{INT} (1.87 * \text{EXP} (2 * \text{LOG} (N-1)/5) + 0.5)$   
150  $R = \text{INT} [(V_1 - V)/G * 100 + 10] / 100$   
160  $V = (V + V_1 - R * G) / 2$   
170 DIM B (G-1)  
180 FOR I=0 TO G-1
```

```

190 B (I) = 0
200 NEXT I
210 FOR I=1 TO N
220 READ A
230 FOR J=0 TO G-1
240 IF A<=V+J * R THEN 280
250 IF A>V+ (J+1) * R THEN 280
260 B (J) =B (J) +1
270 GOTO 290
280 NEXT J
290 NEXT I
300 FOR I=0 TO G-1
310 PRINT "(-", V+R* (I+1) ; ")",
      TAB (10) ; "I",
320 N1=INT [B (I) /N*50+0.5]
330 FOR J=1 TO N1
340 PRINT "※",
350 NEXT J
360 PRINT "(", B (I) ; ")"
370 NEXT I
380 END

```

五、舉 例

現以某杉木标准地上对每株杉木的胸高直径进行检尺，
110株杉木其原始调查记录如下所示。

18.2 15.2 17.0 19.0 16.6 15.0 20.1 18.1

19.1	17.1	17.7	18.1	17.0	16.7	18.3	20.1
19.6	11.6	16.1	18.2	17.5	14.3	17.7	19.2
18.2	13.1	17.3	17.7	17.8	18.1	24.1	15.5
16.7	18.9	18.5	18.3	17.1	20.0	19.1	18.5
25.7	19.2	15.8	17.6	17.3	17.1	16.7	17.4
18.5	12.4	17.7	19.3	19.2	17.4	19.1	17.2
18.1	16.5	18.2	12.6	17.3	19.6	16.3	16.7
17.5	17.6	17.7	16.4	17.3	11.7	18.8	19.1
18.4	18.2	19.1	19.1	20.1	15.8	16.6	10.9
18.2	17.2	18.2	20.1	19.2	17.2	18.6	17.1
17.1	24.6	15.6	16.3	18.6	17.3	18.4	16.4
17.3	19.1	19.1	26.7	15.8	16.1	18.3	16.4
18.3	16.5	17.1	25.6	15.1	27.9		

将上述数据用置数语句输入计算机，则运行结果：

RUN

NUMBERS OF DATA

N=110

(-11.85) I* (3)

(-13.36) I* (3)

(-14.87) I* (1)

(-16.38) I* (12)

(-17.89) I* (40)

(-19.4) I* (38)

(-20.91) I* (7)

(-22.42) I* (0)

(-23.93) I* (0)

$(-25.44) I \times (2)$

$(-26.95) I \times (3)$

$(-28.46) I \times (1)$

上图第一个括号内的数(如-11.85)为该组的组中值,右边括号内的数表示频率。

§1.3 累计频率分布

一、功能

计算大样本数据的累计频率是数据处理的一个重要方法,本程序以最小值开始,按指定的组距计算出各组累积频数,并打印出累计频率分布图。

二、数学方法

1. 给定最小值和组距:一般根据专业知识和具体情况确定,如当计算直径分布的累计频率时,最小值就是起测直径,组距就是径阶的级距(通常为2cm)。
2. 按指定的最小增量计算各组累积频数和累积频率。
3. 打印分布图和累积频率的百分数。

三、程序说明

N数据个数;

V最小观察值;

D组距;

N_1 、 N_2 均为累计频数;