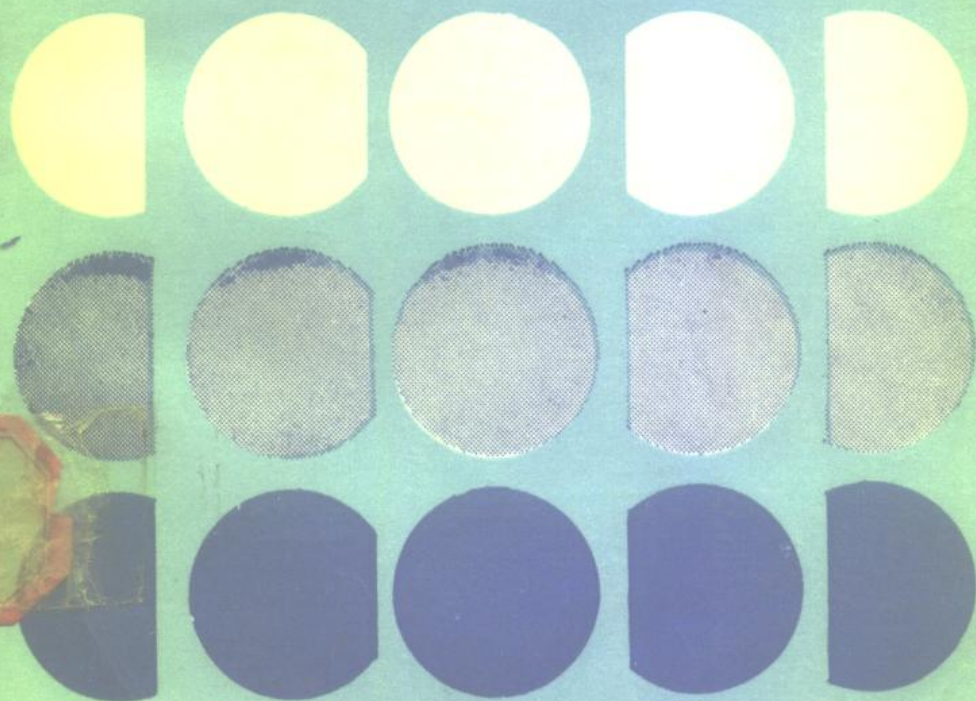


工程数据处理 及其微机电算

肖作善 编



工程数据处理 及其微机电算

肖作善 编

水利电力出版社

内 容 提 要

本书介绍工程技术和科研试验中常用的数据处理及其微型计算机电算方法。内容包括：数据整理、平均值的统计推断与检验、数据的方差分析、数据的回归分析、数据的曲线拟合与平滑、插值法等。

书中简要叙述了基本数学方法，以应用示例的方式介绍了具体计算或分析的步骤；给出了基本BASIC语言程序，配有程序说明、上机例题和上机练习题，可用于APPLE-II（紫金-II）、IBM-PC（长城0520）等微型计算机及其兼容机进行数据处理。

本书内容简明、系统，有利于读者在较短时间内学会数据处理及其微机电算的方法。本书可供电力、能源、化工、冶金、机械、建筑、轻纺等各专业工程技术人员、科研工作人员、大专院校师生使用。

工程数据处理及其微机电算

肖作善 编

*

水利电力出版社出版、发行

（北京三里河路6号）

各地新华书店经售

水利电力出版社印刷厂印刷

*

787×1092毫米 32开本 7.75印张 170千字

1990年3月第一版 1990年3月北京第一次印刷

印数0001—4370册

ISBN 7-120-00975-3/O·14

定价5.40元

前 言

数据是从工程技术现场、试验设备或实验室经量测（或测定）、采集获得的，它是花费了许多物力、人力和宝贵的时间所得到的一种资料。但是，在许多情况下，未经正确分析或处理的数据，并不能反映事物的真实特征或者事物之间的相互关系，甚至会导致作出错误的结论或判断。只有科学地对数据进行整理、分析（称为数据处理或数据分析），才能从数据所提供的信息中，得出分析问题和解决问题的正确结论。可以说处理数据也是科技工作者的一项基本功。

数据处理（或数据分析）中，有时会用到计算工作量很繁重的数学方法，手算相当不易，有的甚至使人望而却步。但是，随着电子计算机的广泛使用，特别是微型计算机的普及，利用编制好的算法程序，可以非常容易地解决数据处理的问题。当然，应用微机于数据处理时，最好基本上已掌握或了解数据处理所依据的原理、方法和步骤。

当前，许多从事工程技术实际工作的同志和大专院校的师生，迫切需要一本实用、简明、便于自学的书，以期能较快地掌握工程数据处理的方法和使用微机处理数据的技能。为了适应这种客观需要，编者将校内教学所用的讲义改编成书，希望能对学习工程数据处理的人们有所帮助。

本书共六章。前四章所讨论的问题，是涉及随机变量的数据处理；后两章讨论的问题，是有确定性关系的普通变量的数据处理。书中内容着眼于实用，在介绍各种数据处理方法的基本原理时，尽可能地将数学公式的来龙去脉交待清

楚，使读者能够理解和掌握。各章节都有应用示例说明所述方法的具体分析、计算步骤，以帮助读者学会实际问题。

书中的算法语言程序是用基本BASIC语言编排的。程序编排的思路与书中介绍的计算方法、步骤等基本上一致，使程序的易读性好。书中程序有简明、准确、方便使用的特点。全部程序均系作者参考有关程序汇编资料重新改排而成，并全部经作者在APPLE-II型微型计算机上实算通过。还抽取部分程序在其它类型计算机上验算通过。

书中上机练习题选自有关参考文献，涉及工程技术的许多专业。通过上机实算，读者可举一反三，触类旁通，为解决各自专业的实际问题打下基础。实际上读者只要把具体问题的数据输入计算机，起动程序后，就可得到结果。

北方工业大学陈凯副教授审阅了本书全稿，提出了许多宝贵的修改意见，作者向他表示深切的谢意。在编写本书的过程中，还得到校内外许多同志的热情帮助和支持，在此一并表示感谢！

限于编者的数学理论水平和计算机数据处理的实际经验，书中的错误与不妥之处在所难免，欢迎批评指正。

肖作善

于武汉水利电力学院

1989年3月定稿

目 录

前 言

第一章	数据整理	1
第一节	将数据按顺序排列	1
第二节	频率分布	4
第三节	数据的数字特征参数计算	11
第四节	离群数据的取舍	16
第二章	平均值的统计推断与检验	32
第一节	估计总体平均值的置信区间	32
第二节	总体平均值的检验	38
第三节	成对比较试验数据差值的推断	47
第四节	成组比较试验平均值的推断	56
第三章	数据的方差分析	74
第一节	单因素试验数据的方差分析	75
第二节	双因素试验数据的方差分析	87
第四章	数据的回归分析	104
第一节	一元线性回归	105
第二节	化非线性回归为线性回归	125
第三节	多元线性回归	133
第四节	非线性相关	152
第五节	逐步回归分析	156
第五章	数据的曲线拟合与平滑	175
第一节	数据的曲线拟合	175
第二节	数据的平滑化	187
第六章	插值法	194

第一节	一元 n 点拉格朗日插值法	195
第二节	一元三点插值法	199
第三节	二元三点插值法	203
第四节	厄米特插值法	209
第五节	样条插值法	213
附录 I	t 分布的双侧分位数表	224
附录 II	F 检验的临界值(F_{α})表	226
参考文献		240

第一章 数 据 整 理

在工程实践或试验中，对某一特性量多次重复地进行量测或测定之后，得到一组数据。这些数据一般不会是同一数值，而是一些分散的数值。如果直接用这些数据来分析研究对象的特性或者要拿它们与其他数据相比较，往往是不太方便的，甚至无法进行。为了使数据中的信息变得更加简明，便于利用，有必要对数据进行整理。

第一节 将数据按顺序排列

一、数学方法的简述

在整理数据时，有时需要将数据按从大到小（或从小到大）的顺序排列。数据按顺序排列后，很容易看出一组数据中的最大值和最小值。

一组数据（常称为样本或子样）中最大值 $x_{\max}$ 与最小值 $x_{\min}$ 之差，称为极差 R

$$R = x_{\max} - x_{\min}$$

极差是描述数据分散程度的一个最简单的离散特性参数。

要将一组数据按值从大到小顺序排列，在设计计算机BASIC语言程序时，用数组 $A(N)$ 存放 $x_1, x_2, \dots, x_n$ 共 n 个数据的值。先把 $A(1)$ 和 $A(2)$ 比，如 $A(1) \geq A(2)$ ，则 $A(1)$ 不动。若 $A(1) < A(2)$ ，就把 $A(1)$ 和 $A(2)$ 中的数值对调；然后再把 $A(1)$ 与 $A(3)$ 相比较，步骤同

上。这样一直到比完 $A(N)$ 之后, $A(1)$ 中的数就是最大值。

再求次大值。将 $A(2)$ 用上述同样的方法和其它变量 $A(3) \sim A(N)$ 一一相比处理, 最后所得 $A(2)$ 中的数就是 N 个数中的次大值。

按同样的方法继续做下去, 一直做到 $A(N-1)$ 和 $A(N)$ 相比后为止; 这时 $A(N-1)$ 中的数为次小值, 剩下的 $A(N)$ 中的数就是整个数组中的最小值。

二、BASIC语言程序

(一) 程序说明

1. 变量

N ——数据个数;

$A(N)$ ——数据值;

$M1$ ——数据中的最大值;

$M2$ ——数据中的最小值;

R ——极差。

2. 程序使用

(1) 在程序第250句置数语句, 键入原始数据;

(2) 键入RUN, 起动程序;

(3) 人机对话:

屏幕显示

用键盘回答

"NUMBER OF DATA"

数据个数 N

(4) 计算机自动显示(或打印)结果;

(5) 程序中第100句 $A(I) \geq A(J)$ 改为 $A(I) < A(J)$ 后, 数据将按由小到大的顺序排列。

(二) 程序

```
10 PRINT "NUMBER OF DATA"
```

```

20 INPUT N
30 PRINT
40 DIM A(N)
50 FOR I=1 TO N
60 READ A(I)
70 NEXT I
80 FOR I=1 TO N-1
90 FOR J=I+1 TO N
100 IF A(I) >= A(J) THEN 140
110 LET A1=A(I)
120 LET A(I)=A(J)
130 LET A(J)=A1
140 NEXT J
150 NEXT I
160 FOR I=1 TO N-1
170 PRINT A(I); ", ";
180 NEXT I
190 PRINT A(N); ". "
200 PRINT "M1="; A(1)
210 PRINT "M2="; A(N)
220 LET R=A(1)-A(N)
230 PRINT "R="; R
240 END
250 DATA.....

```

(三) 例题

【例 1-1】 将下面50个数据按由大到小的顺序排列:36, 49, 43, 41, 37, 40, 32, 42, 47, 39, 41, 36, 48, 34,

42, 42, 45, 35, 42, 39, 44, 42, 39, 42, 42, 45, 30,
34, 42, 37, 36, 37, 34, 37, 37, 44, 32, 48, 40, 45,
39, 49, 39, 53, 36, 48, 40, 39, 38, 40.

【解】 程序中本题的置数语句为

```
250 DATA 36, 49, 43, 41, 37, 40, 32, 42,  
47, 39, 41, 36, 48, 34, 42, 42, 45, 35, 42,  
39, 44, 42, 39, 42, 42, 45, 30, 34, 42, 37,  
36, 37, 34, 37, 37, 44, 32, 48, 40, 45, 39,  
49, 39, 53, 36, 48, 40, 39, 38, 40
```

计算机运行结果如下:

RUN

NUMBER OF DATA

? 50

53, 49, 49, 48, 48, 48, 47, 45, 45, 45, 44,
44, 43, 42, 42, 42, 42, 42, 42, 42, 42, 41,
41, 40, 40, 40, 40, 39, 39, 39, 39, 39, 39,
38, 37, 37, 37, 37, 37, 36, 36, 36, 36, 35,
34, 34, 34, 32, 32, 30

$M_1 = 53$

$M_2 = 30$

$R = 23$

第二节 频 率 分 布

整理数据时, 有时要用频率分布的形式把一组数据的分布情况显示出来。

一、数学方法简述

对某一特性量测得了 n 个数据，这也就是说从整体中抽取了容量为 N 的一组子样。将这些数据按值的大小分成若干组，设有 f_i 个数据落在第 i 组内，则 f_i 称为第 i 组的频数， f_i/n 称为频率（%）。这样可算出这些数据的频率分布表，由频率分布表可以了解数据的一些分布特征，如数据的密集位置、数据的离散情况等。

在一些情况下，需要知道小于或等于某一数值的频数和频率，这就要把小于设定数值的各组的频数和频率分别相加，这样得到的频数叫累计频数，所得到的频率叫累计频率（%）。

在数据分组时，可考虑具体情况从自行指定的最小值开始，按自行选定的增量（或称组距）分组。为了使每个数据都能确切地放入某组内，分组时通常把每一组的上下限多取一位数。例如对整数65，当组距为5时，可设定组限为64.5~69.5，以使65确切落入这一组中。这就是说，对于整数，其分组的上下限应取一位小数。

二、BASIC语言程序

（一）程序说明

1. 变量

N ——数据个数；

V ——按具体情况自行选定的最小数值；

D ——按具体情况自行选定的组距；

$A(N)$ ——数据值；

M_1 ——组频数；

M_2 ——组频率；

N_1 、 N_2 ——累计频数；

N3——累计频率。

2. 使用说明

(1) 在程序第370句的置数语句中,用键盘将数据 $x_1, x_2, \dots, x_n$ 依次输入计算机,如果数据很多,可以分几句输入;

(2) 键入RUN起动程序;

(3) 人机对话:

屏幕显示	用键盘输入
"NUMBER OF DATA"	数据数目N
"PRIMARY VALUE"	最小数值V
"INCREMENT"	增量或组距D

(4) 计算机运算并显示(或打印)计算结果。

(二) 程序

```
10 PRINT "NUMBER OF DATA";
20 INPUT N
30 PRINT "PRIMARY VALUE";
40 INPUT V
50 PRINT "INCREMENT";
60 INPUT D
70 PRINT
80 DIM A(N)
90 LET N1=0
100 FOR I=0 TO N
110 LET M1=0
120 PRINT V+I*D; " "; " "; V+(I+1)*
    D; TAB(20);
130 FOR J=1 TO N
```

```

140 READ A(J)
150 IF A(J) < V + I * D THEN 180
160 IF A(J) >= V + (I + 1) * D THEN 180
170 LET M1 = M1 + 1
180 NEXT J
190 LET M2 = INT(M1 / N * 10000 + 0.5) / 100
200 PRINT "(", M2, "%)", TAB(30);
210 RESTORE
220 LET N2 = 0
230 FOR J = 1 TO N
240 READ A(J)
250 IF A(J) > V + (I + 1) * D THEN 270
260 LET N2 = N2 + 1
270 NEXT J
280 IF N1 = N2 THEN 330
290 LET N1 = N2
300 LET N3 = INT(N1 / N * 10000 + 0.5) / 100
310 PRINT N3, "% "
320 IF N1 = N THEN 360
330 PRINT
340 RESTORE
350 NEXT I
360 END
370 DATA.....

```

(三) 例题

【例 1-2】 现有下列40个数据:63, 68, 71, 74, 78, 76, 84, 85, 81, 66, 89, 75, 79, 76, 70, 73, 90, 84,

82, 85, 92, 94, 67, 68, 84, 82, 79, 79, 76, 77, 85, 86, 86, 85, 71, 73, 75, 75, 74, 71。试以59.5为最小合格值, 以5为组距, 计算各组的频率和累计频率。

【解】 在程序第370句置数语句中, 输入数据。本题N为40, V为59.5, D为5。计算机计算结果打印如下:

LIST 370

370 DATA 63,68,71,74, 78, 76, 84, 85, 81, 66,
89,75,79,76,70,73,90,84,82,85, 92, 94, 67,
68,84,82,79,79,76,77, 85, 86,86,85, 71,73,
75, 75,74,71

RUN

NUMBER OF DATA ? 40

PRIMARY VALUE ? 59.5

INCREMENT ? 5

59.5; 64.5	(2.5%)	2.5%
64.5; 69.5	(10%)	12.5%
69.5; 74.5	(20%)	32.5%
74.5; 79.5	(27.5%)	60%
79.5; 84.5	(15%)	75%
84.5; 89.5	(17.5%)	92.5%
89.5; 94.5	(7.5%)	100%

【例 1-3】 现有下列100个数据: 36, 49, 43, 41, 37, 40, 32, 42, 47, 39, 41, 36, 48, 34, 42, 42, 45, 35, 42, 39, 44, 42, 39, 42, 42, 30, 34, 42, 37, 36, 37, 34, 37, 37, 44, 45, 32, 48, 40, 45, 39, 49, 39, 53, 36, 48, 40, 39, 38, 40, 36, 45, 50, 43, 38, 43, 41, 48, 39, 45, 37, 37, 39, 45, 31, 41, 44, 44, 42, 47,

35, 36, 39, 40, 38, 35, 42, 43, 44, 48, 42, 40, 41, 37, 46, 36, 37, 27, 37, 38, 42, 34, 43, 42, 41, 44, 41, 48, 55, 37。计算各数值的频率和累计频率。

【解】 本题可设定26.5为最小值，以1为组距。这样，数据27即确切落入26.5~27.5之中，并为组内中值。其它数据类推。

在程序置数语句第370句和第380句中，各输入50个数据；本题的数据数目N为100，最小数值V为26.51，组距为1。

计算机计算结果打印如下：

LIST 370, 380

370 DATA 36, 49, 43, 41, 37, 40, 32,
42, 47, 39, 41, 36, 48, 34, 42, 42,
45, 35, 42, 39, 44, 42, 39, 42, 42,
30, 34, 42, 37, 36, 37, 34, 37, 37,
44, 45, 32, 48, 40, 45, 39, 49, 39,
53, 36, 48, 40, 39, 38, 40

380 DATA 36, 45, 50, 43, 38, 43, 41,
48, 39, 45, 37, 37, 39, 45, 31, 41,
44, 44, 42, 47, 35, 36, 39, 40, 38,
35, 42, 43, 44, 48, 42, 40, 41, 37,
46, 36, 37, 27, 37, 38, 42, 34, 43,
42, 41, 41, 44, 48, 55, 37

RUN

NUMBER OF DATA ? 100

PRIMARY VALUE ? 26.5

INCREMENT ? 1

26.5; 27.5	(1%)	1%
27.5; 28.5	(0%)	
28.5; 29.5	(0%)	
29.5; 30.5	(1%)	2%
30.5; 31.5	(1%)	3%
31.5; 32.5	(2%)	5%
32.5; 33.5	(0%)	
33.5; 34.5	(4%)	9%
34.5; 35.5	(3%)	12%
35.5; 36.5	(7%)	19%
36.5; 37.5	(11%)	30%
37.5; 38.5	(4%)	34%
38.5; 39.5	(9%)	43%
39.5; 40.5	(6%)	49%
40.5; 41.5	(7%)	56%
41.5; 42.5	(13%)	69%
42.5; 43.5	(5%)	74%
43.5; 44.5	(6%)	80%
44.5; 45.5	(6%)	86%
45.5; 46.5	(1%)	87%
46.5; 47.5	(2%)	89%
47.5; 48.5	(6%)	95%
48.5; 49.5	(2%)	97%
49.5; 50.5	(1%)	98%
50.5; 51.5	(0%)	
51.5; 52.5	(0%)	
52.5; 53.5	(1%)	99%