实验数据的统计分析和计算机处理

刘志万　编著

中国科学技术大学出版社

实验数据的统计分析和计算机处理

刘志万　编著

中国科学技术大学出版社

1989・合肥

实验数据的统计分析和计算机处理

刘志万　编著

责任编辑：杜凤兰　　封面设计：王瑞荣

*

中国科学技术大学出版社出版

(安徽省合肥市金寨路96号)

中国科学技术大学印刷厂印刷

安徽省新华书店发行

*

开本：850×1168/32　印张：6.125　字数：157千

1989年6月第1版　1989年6月第1次印刷

印数：1－5000册

ISBN 7-312-00113-0/TP·4　定价：2.50元

内 容 简 介

本书介绍了与测量误差及数据分析和处理有关的基本概念和原理，重点叙述数字修约规则、测量数据的统计检验、粗差的统计剔除、最小二乘法、曲线的回归和修匀处理，并罗列了利用电子计算机处理数据的部分程序流程和源程序。

本书概念清晰，浅显易懂，详略得体，切合实用。立意于误差理论与计算机处理方法的交融。给出了数据处理程序框图15幅，源程序20个，数表24张，并辅以题例28道。所列程序均已在IBM－PC微型计算机上运行通过，编译文本为FORTRAN Compiler Version 2.00。

本书可作为高等院校有关专业的教学参考书，也适合于从事计量和测量、自动检测、各学科的实验研究和数学模拟、统计分析等广大科技人员阅读。

前　言

误差理论是人们对测量实践中量值观念的科学概括和数学抽象，是测量技术和仪器学科的重要理论支柱，在现代实验测量中的地位愈显重要。

科学实践提示，一切实验都不可避免地会产生误差。为了提高测量精确度，要求我们竭力削弱各种误差的影响，获取最佳测量条件，并运用误差理论的有关原则来处理数据，以确定被测量值或揭示被测对象对输入参量的响应规律。

对测量结果数据处理方法的研究，已成为现代统计数学的一个重要组成部分。电子计算技术的迅猛发展，导致误差理论和数据处理方法在理论上和实用上都有新的飞跃，使误差理论中极难处理、并早已建立的许多经典分析方法——诸如蒙特卡洛统计试验法、富氏变换等，变得更加方便和准确，能在现实工作中更大地发挥作用。而具有数据处理功能的仪器装置，现已成为自动检测或动态测试系统中必不可少的重要组成部分。

本书以工程和有关学科实验数据的统计估算和图解拟合问题为主线，概括介绍了与数据分析和处理有关的误差理论主要内容和概念，以及数字修约规则、恒定量数据处理、变化量的回归分析和修匀处理等数学方法，并列举了部分计算机处理程序结构框图或辅以相应的源程序。显然，介绍编写程序的方法将比罗列程序本身更为重要。给出的数据处理程序，其结构形式绝非唯一，功能也不尽完美，但可以作为数据处理原理与计算手段相结合的例证，以期既加深对有关误差原理的理解，又能在应用现代计算工具综合实验结果的过程中，深化和开拓新的数据处理内容和计算方法。

不同学科，数据处理的范畴和涉及的数学方法皆各有侧重。本书使用浅显的数学知识，明晰而简洁地表述统计学的基本原理，旨在误差理论和计算机处理方法的交融，用通用的计算机语言，给出了数种工程上适用的在等精度测量条件下的典型数据处理程序，用于日常统计处理或训练基本技能，并为应用系统处理软件打下基础。相信这对应用数理统计方法并有意于自己开发数据处理软件的科技工作者和从事实验研究的专业人员会有所裨益。

作者衷心感谢中国科学技术大学电子技术基础部和科技开发部负责人及有关同志的鼓励、支持和建议。

限于作者个人的学识与实际经验，难免隐伏欠妥以至谬误之处。敬祈批评指正，并谨申谢忱。

刘志万

1988年1月28日

目　　录

统计计算程序一览表

第一章　有效数字与计算法则

在实验测量和数值计算中，因仪器精确度和分辨力的限制，以及测量者感官的缺陷，所读测的数值是近似值．一个数值中小数点后面的位数愈多，并不表明数值就愈精确．首先，频率值21.4Hz与0.0000214MHz相比，二数值蕴含的精确度完全相同；其次，读记测定值的位数时，绝不能把精确度增加到超过测量本身所能给出的程度．确定应该用几位数字来代表被测量值的最佳估计值，在数值处理中至关重要．有鉴于此，有必要在探讨用数理统计方法分析和处理实验数据之前，先研究实验数据的有效位数和修约规则．

例 1.1　方程组（A），（B）及其对应解分别为

$$(A)\ \begin{cases} x-y=1 \\ x-1.0001y=0 \end{cases} \qquad \text{解为} \begin{cases} x=10001 \\ y=10000 \end{cases}$$

$$(B)\ \begin{cases} x-y=1 \\ x-0.9999y=0 \end{cases} \qquad \text{解为} \begin{cases} x=-9999 \\ y=-10000 \end{cases}$$

两个方程组的系数仅差万分之二，而结果却差异极大．由此说明研究数值的取值和有效位数以及遵循近似值计算规则的重要性．

§1.1　有效数字

测量结果通常表示为一定的数值．由于存在误差，测量结果的可信数字位数总是有限的，与测量精确度有关．一般而言，数值的倒数第二位是真正有意义的，而最后一位数则是最佳推测值．有效数字是指测量误差的绝对值不超过末位数字的0.5时，从

它左边第一个不为零的数字算起，到包括零的最末一位数为止。换言之，有效数字是指一串数中的任一数字，它只表示一个数值的大小，而不是用来指明小数点的位置。例如读数0.02080V，2的前面两个0不是有效数字，中间及末尾的0都是有效数字，该数字的有效位数为4。再如例1.2一组频率测定值中，

例 1.2 一组频率测定值

（1）128Hz　　（2）0.128kHz　　（3）12.08Hz

（4）12.80Hz　　（5）12.800kHz

测定值（1）的有效位数为3，它表示128比127或129更接近于所测量值的大小，亦即被测量值介于127.5与128.5之间。测定值（2）的有效位数亦为3，有效数字前的0仅供指示小数点的位置之用，（1）与（2）的量值完全相等，不同之处仅在于表示量值的单位。测定值（3），（4）的有效位数皆为4，前者量值介于12.075与12.085之间，后者为12.795与12.805之间。测定值（5）的有效位数为5，亦可记为1.2800×10^4Hz。

数据较好的表示方法是指数形式，指数部分前面的数字为有效数字。如上例，可以将各数值分别写成：

1.28×10^2Hz，　1.28×10^2Hz，1.208×10^1Hz，

1.280×10^1Hz，1.2800×10^4Hz

理论计算出的数值，可以认为有效位数无限制；而由实验得到的测定值，其有效位数则决定于实验方法和所用仪器设备的精确度。

§1.2　数字修约规则

一个数据，超过有效位数的数字应按数字修约（舍入）规则取舍。

古典的四舍五入规则会使修约后的数据量值偏高，无法消除。为此，现用的修约规则规定：以被保留数字的末位为基准，它后面的尾数小于5舍，大于5入；恰为5者，则将末位凑成偶数。

也就是说，当末位为奇数时进1，为偶数时舍去该尾数。以上规则简称为"四舍六入五凑偶"规则。凑偶可以使尾数被取舍的概率接近相等，且处理值作除法运算时，被除尽的可能性较奇数多，以利于减少计算误差。

例 1.3 一实测数列，取其三位有效数字，单纯截尾、四舍五入和按修约规则处理的数据如下表：

表 1.1 三种数据处理方法比较

测定值	截尾	四舍五入	修约凑偶
12.365	12.3	12.4	12.4
12.550	12.5	12.6	12.6
12.140	12.1	12.1	12.1
12.193	12.1	12.2	12.2
12.854	12.8	12.9	12.9
12.201	12.2	12.2	12.2
12.750	12.7	12.8	12.8
12.250	12.2	12.3	12.2
12.450	12.4	12.5	12.4
12.635	12.6	12.6	12.6
12.653	12.6	12.7	12.7
12.850	12.8	12.9	12.8
Σ：149.891	149.3	150.2	149.9

很清楚，从各数列之和可以看到，截尾值偏低，修约凑偶值最接近原测定值，传统的四舍五入方法累积的方法误差比按修约规则处理的误差大。同时亦可见到，每个数值经过修约处理后，数据的差值均不大于0.5（以保留数字的末位为基准），此0.5即为修约误差。

对处理数据不应多次修约，以免造成更大的误差。例如一个

五位实测值0.81149，取四位有效数字时，按修约规则应为0.8115，再取其三位，又为0.812；而如果由测定数据直接取三位有效数字，则应为0.811.

同时，数值也宜在算术运算后修约.如上例，粗测值为0.81149，取其4倍率的三位有效数字是3.25；若先取三位有效字为0.811，其4倍率的三位有效数则是3.24. 显然前一数值比较准确.

当未注明测量数据的测量误差时，通常认为数据的最后一位数字有数值为0.5的误差，称此为0.5误差原则. 例如测定值为87.45V，则可信范围为87.4—87.5V. 该误差原则已在前例中被引用过.

当测量误差已知时，测量结果的有效数字位数应当与该误差的位数相一致. 例如用一只0.5级电压表的100V量程测量，指示值为87.45V，由于电压表在该量程的最大绝对误差为

$$\Delta V_m = \pm 0.5\% \cdot V_m = \pm 0.5\text{V},$$

根据0.5误差原则，有效数字的末位应该在小数点前一位，此时测量数据的末位应是整数，故电压表的示值处理值应为87V，0.45V作为有效位数后的尾数被舍去. 如果示值为87.50V，考虑到修约规则，处理值则应凑偶为88V.

由上可见，测量结果的有效数字反映了测量数据的准确程度. 上例处理值87V，末位是个位，表明其绝对误差为±0.5V. 现实测量中常盲目追求多位读数，而不考虑仪表精度等级的弊病是常有发生而应竭力避免的.

对实验测量数据进行中间运算时，分别有加减、乘除和函数运算的修约规则可资遵循. 由于计算机的使用，可以不必预先对数列作有效位数的处理，因为计算机的有效位数总是多于常规仪表读测值的有效位数. 可以视运算工具的能力，尽可能多取数位，以保证最后结果不损失精度，只需最终运算值的位数与测量精度相对应即可. 同时也应注意，由于计算机的有效位数有限，在运算程序设计中，尽量不要使两个相差很大的数直接相加或相减，

以免导致终值数位过多而被丢舍所引起的计算误差。

在实际的数据修约处理过程中，也会遇到数值位数和整数位各不相同的数列，仍可按修约规则处理之。

例 1.4 一组整数部分位数不相同、数位最多为 6，且包含负值的实测数列，按修约规则处理，使有效位数为 4，测定值及处理值如表 1.2 所示：

§1.3 修约处理方法和程序

若借助于计算机进行有效数字处理，可以使用库函数 INT(X) 对数据按有效位数作截尾运算，然后按“四舍六入五凑偶”的修约规则处理尾数。当以小数点后的数值标定有效位数，且数据值 $x_i>0$ 时，如例 1.3 之数列，数据处理程序流程图如图1.1 所示。

流程图以图形的形式说明程序的逻辑顺序，清楚地描述计算机如何处理信息。流程图比源程序本身更容易统观全貌，标示各语句的功能和相互关系，且不涉及任何特定的程序语言，对各种机型和语言文本都有较强的适应能力。用粗略的流程图可以按顺序检查所用逻辑的正确性。本书流程图的内容陈述偏细，以程序语句和逻辑框图相结合的形式，表达出程序功能结构的细节，易于直接导出源程序，键入计算机运行处理。

按数列的小数部分确定有效位数 J，如题例 1.3 数列，此时 J＝1，利用取整函数 INT(X) 按有效位数截取数值 B(I)，然后作 Z(I)—100*B(I) 运算，求出有效位数后面包含有两位整数的尾数 U(I)，再按修约规则用算术条件转移语句进行 U(I)－50≷0 的判比。当尾数 U(I) 小于 50 时舍去；大于 50 时，有效位数的末位加 1，即 B(I)＝B(I)＋1；恰为 50 者，先判别有效数值 B(I)的奇偶，即当实型求余函数 AMOD(B(I),2)≠0，为奇数时，将末位加 1 凑成偶数。随后，将小数点在有效位数后面的

表 1.2 任意数列的修约

测定值	修约处理值		注记
	小数形式	指数形式	
523.749	523.7	0.5237×10^{3}	正值 舍
617.995	618.0	0.6180×10^{3}	正值 入
−0.00739504	−0.007395	-0.7395×10^{-2}	负值 舍
−3.21589	−3.216	-0.3216×10^{1}	负值 入
2960.51	2961.0	0.2961×10^{4}	正值 0.51>0.5，入
−19.02530	−19.03	-0.1903×10^{2}	负值 0.530>0.5，入
−0.0015885	−0.001588	-0.1588×10^{-2}	负值 0.5，偶舍
0.0035165	0.003516	0.3516×10^{-2}	正值 0.5，偶舍
189.950	190.0	0.1900×10^{3}	正值 0.50，奇凑偶
−23.075	−23.08	-0.2308×10^{2}	负值 0.5，奇凑偶
0.0	0.0	0.0000×10^{0}	零
−0.0	0.0	0.0000×10^{0}	零

运算值 B(I) 按其原来的小数位复原为 R(I)，并打印出有效位数 J 和按修约规则处理后的数列 R(I)。

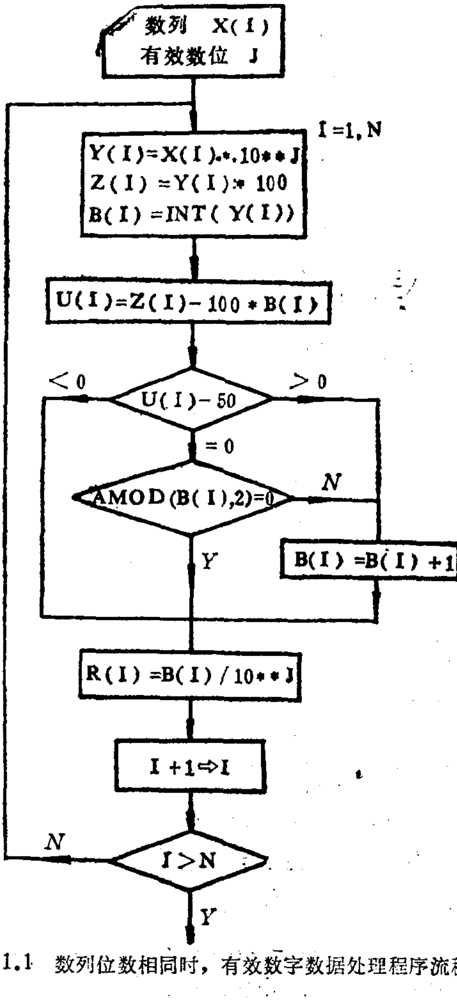

图 1.1 数列位数相同时，有效数字数据处理程序流程图

按以上所述修约规则运算的 FORTRAN 源程序段如下：

```
      DO 80 I=1, N
20    Y(I)=X(I)*10**J
```

```
      Z(I)=Y(I)*100
      B(I)=INT(Y(I))
      U(I)=Z(I)-100*B(I)
25    IF(U(I)-50) 30, 40, 50
40    IF(AMOD(B(I),2.).EQ.0.) GOTO 30
50    B(I)=B(I)+1
30    R(I)=B(I)/10**J
80    CONTINUE
```

计算机中，用数值型格式符作截位打印输出时，大多按四舍五入原则处理截位值的后位数字。

当一组待处理数据的整数部分位数不相同，而要求有效位数K相同，如类似例1.4的数列，在数据处理之初，应先将非零数值作进位或退位运算，使数据的左方第一位有效数字在小数点后。因为左方第一位非零数字前的零仅表示量值的单位，不是有效数字。然后，令数列按小数点对齐，例如将表1.2中的测定值变换为0.523749，0.617995等，以便正确地读取有效位数K。为了在作以上处理后能恢复原数据的位数，当然应在对位运算时分别记录数列各数据的移位值M(I)。

对数列作对位运算，使小数点前导于左方第一位有效数字，满足$0.1\leqslant x_i<1.0$，并记录移位值M(I)的处理程序框图如图1.2所示。先使数据的小数点移位值M(I)初始化，再判别该数值是否介于0.1和1.0之间。若数值大于或等于1.0，作X(I)=X(I)/10的小数点进位运算；若数值小于0.1，则作X(I)=10*X(I)的小数点退位运算。每运算一次，一个数只作一位进位或退位，移位值M(I)加1或减1。若一次判别后仍不能满足$0.1\leqslant X(I)<1.0$，则继续循环移位，直至满足上式关系，转出移位运算，I+1⇒I，处理下一个数据。

一个非负数列作对位原理性运算的FORTRAN源程序段如下：