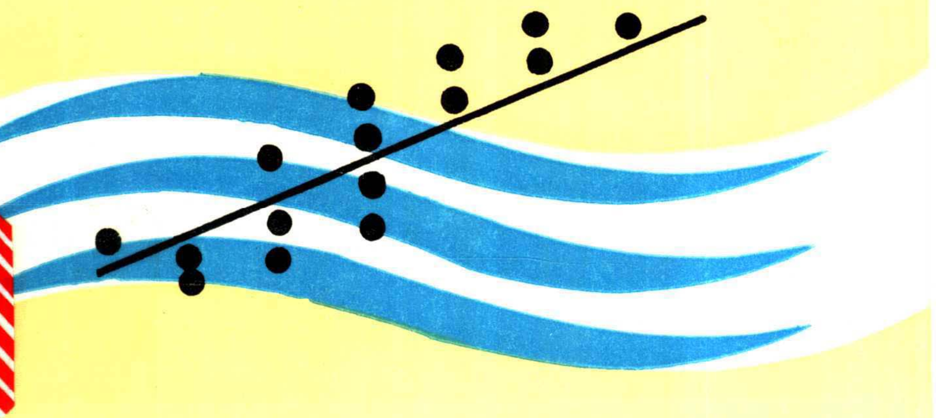
数据分析与测量数据处理

马金铃　高井祥　张书华　编著

SHUJUFENXI YUCELIANGSHUJUCHULI



中国矿业大学出版社

数据分析与测量数据处理

马金玲　高井祥　张书毕　编著

中国矿业大学出版社

(苏)新登字第 010 号

内容简介

本书介绍了测量平差的基础知识及测量数据分析与处理。包括有效数字与运算法则、测量误差理论与权、最小二乘原理与参数平差、粗差及粗差检验、观测数据的统计检验、方差分析、试验设计及其在测量中的应用、回归分析、最小二乘拟合与趋势面分析等内容。

本书可作为测绘专业的大学生、研究生及从事测绘科研和工程技术人员的参考用书。

责任编辑 孙树朴
技术设计 杨 烨
责任校对 杜锦芝

数据分析与测量数据处理
马金玲 高井祥 张书毕 编著

中国矿业大学出版社出版
新华书店经销 中国矿业大学印刷厂印刷
开本 850×1168 毫米 1/32 印张 8.375 字数 210 千字
1994 年 8 月第一版 1994 年 8 月第一次印刷
印数:1—1500 册

ISBN 7-81021-943-X

TP·35 定价:8.50 元

前　言

测量专业本科生、研究生及测量工作人员渴望获得测量数据分析与数据处理方面的知识，为此近几年来给我校研究生开设了《测量数据处理》课，在多年教学的基础上，结合科研成果，参考有关测量平差及数据分析方面的资料，撰写了《数据分析与测量数据处理》一书。

本书在测量、测量误差及有关概率论的基础上，着重介绍了测量上常用的参数平差、测量数据处理、方差分析、回归分析、试验优化设计、最小二乘拟合与趋势面分析等内容。书中融入了近几年的科研成果，配有密切结合测量的实例。全书取材广泛，论述详细，内容新颖。本书以从事测绘科技工作的人员为对象，对其他方面的科技人员进行统计分析及数据处理时，也有重要的参考价值。

全书共分九章，一到三章由马金玲编写，四到六章由张书毕编写，绪论及七到九章由高井祥编写。由高井祥、马金玲主编，负责全书的审阅、修改和定稿工作。

孙树朴同志对本书作了审阅和修改，并给予许多具体指导和帮助。谨此表示感谢。

我们恳切希望读者对本书的错误和不足之处，提出批评和指正。

编著者

1994年8月

目　录

绪　论

数据处理是测绘科学的重要组成内容之一。由于观测成果中不可避免地存在偶然误差、系统误差，甚至存在粗差(即错误)，在实际工作中，为了提高观测成果的质量和可靠性，通常都要进行多余观测。但这又必然导致各观测值之间出现矛盾，必须运用适当的数学方法及计算工具，对这些带有偶然误差的观测值进行处理，求出待估计量的最可靠值，同时用某些数量指标评定出测量成果的质量。上述就是数据处理的主要任务。

早在十九世纪初，高斯(Gauss)和勒让德尔(Legendre)创立最小二乘法以来，测量数据处理理论不断完善。特别是近年来，将最小二乘法和概率统计原理有机地结合起来，使得最小二乘法理论更为周密，内容更加丰富。

随着近代工程数学和电子计算机技术的发展，以及各种测绘新仪器、新设备、新手段的问世，测量数据处理的方法和理论也有了很大的发展，数据处理知识的应用领域越来越广。凡是需要进行实验的工作，都要采用这些知识来分析其实验结果。

进行现代科学实验，测取和积累大量的数据。所获数据的特点是随机性质十分突出，往往使事物或过程的内在规律性被现象的偶然性所掩盖。因此要研究对这种数据的处理方法，充分利用实验提供的信息，找出正确的结论。这一直是数据处理领域的重要研究课题之一。

数据的采集与处理有主动、被动之分。为完成特定的工程项目所采集的数据，大多是被动的；为某项科学试验所采集的数据，大

多是主动的。有许多试验，都是根据试验人员的主观想象和已有经验安排的，主观随意性较强。这样的试验，固然可以得到数据，也能取得一定的效果。但所确定的试验方案是否合理、经济，能否取得最佳效果，就不是每个试验人员都能很好解决的。

随着科学技术的发展，特别是优化理论的发展，人们已越来越注重以较少的投入，获得尽可能大的效益。也就是说，在试验安排前，应进行试验优化设计，必须把试验安排与数据处理问题统筹考虑，使试验设计纳入科学的轨道。

试验设计的方法有多种，对于不同的学科领域，应选用不同的试验设计方法。在测绘科学领域中，由于试验受多种外界环境、仪器、人为等因素的影响，选用能够同时考虑多种因素影响的正交试验设计方法是较为合适的。

试验设计与数据处理是一个问题的两个方面，科学研究以试验获取的原始数据为前提，数据处理则是对试验数据中的信息的提取，是取得正确结论的重要一环。

将试验设计与测量数据处理有机地结合起来，这是测绘科学技术发展的必然结果，许多学者正在从理论上和实用上对现有的试验设计方法和有关数据处理理论进行更深入的探讨，并且已取得一些可喜的成果。这必将推动测绘科学研究事业的发展和进步。

第一章　有效数字与运算法则

第一节　有 效 数 字

在生产和科学实验中，需要记录有关数据供处理分析用。处理数据时，就将涉及到数据的有效数字问题。

一个数值除去数前面的一些零，余下的全部数字称为有效数字。一个有效数字的有效位数指的是最高位数字不为零的位数。例如：

23	有 2 位有效数字
405	有 3 位有效数字
30.06	有 4 位有效数字
29.070	有 5 位有效数字
0.008	有 1 位有效数字
1500	有 4 位有效数字

应当指出，如果将 1500 表示为 15×10^2 时，则有效数字只有 2 位。

在测量和数字计算中，应该用几位数字来代表所测得的数据或计算结果是一件很重要的事情。有人误认为一个数值中小数点后面的位数愈多，这个数愈准确，其实不然。因为小数点位置仅与所用单位的大小有关，不是决定准确与否的标准。例如进行长度测量时，长度记为 68.573m 与记为 68573mm 的准确性完全相同。另外，由于受测量方法、测量仪器的精度与观测者感官的鉴别能力的

限制，在计算结果中无论写多少位数，也不可能把准确程度增加到超过测量所允许的范围。反之，如果写出的数字位数低于测得数据的位数时，将降低数据的精确度。正确的作法是所取位数中，除末位数字是可疑数或不确定外，其余各位数字都应是准确可靠的。一般认为末位数字可能有±1 个单位的误差，或其下一位的误差不超过±5。

对于一个直接测量的值来说，其有效数字的个数是不难确定的，它取决于所用仪器的最小刻度。一般人眼只能估读到最小刻度的十分之一，故在记录测量结果时，只允许末一位是估计得到的不准确数字，其余数字均是可靠的。例如，用一根厘米刻度的钢卷尺测量一段距离时，可估读到毫米。测得的长度读数为 84.516m，则前 4 位是准确可靠的，第 5 位的数字是估读的，不准确的，所以该数有 5 位有效数字。

第二节 舍入法则

一、“四舍五入”法则形成的舍入误差

当确定了有效数字的位数后，进行数据处理时，以往通常采用“四舍五入”法则。但采用这一法则时舍入误差较大，设 X 为舍入误差，则

$$X = b - B$$

式中 B——精确值；

b——近似数。

若某一数值要保留 N 位有效数字，对第 $N+1$ 位按“四舍五入”法则进行处理。则第 $N+1$ 位数及其相应的舍入误差为表 1-1 所示。

表 1-1

第 $N+1$ 位的数	0	1	2	3	4	5	6	7	8	9
舍入误差 X	0	−1	−2	−3	−4	5	4	3	2	1

舍入误差 X 为离散型随机变量，其出现的概率为

$$P(X=0,-1,-2,-3,-4,5,4,3,2,1)=1/10$$

离散性随机变量 X 的数学期望为

$$E(X)=\sum_i P_i X_i$$

因此，按“四舍五入”法则形成的舍入误差 X 的期望值为

$$E(X)=\sum_i P_i X_i=\frac{5}{10}=\frac{1}{2}$$

即舍入误差的期望值等于第 $N+1$ 位单位的二分之一。舍入误差是人为的引入误差，为使舍入误差达到最小，即数学期望为零，只须这样规定：当第 N 位上的数是偶数(包括 0)时，第 $N+1$ 位的 5 予以舍去，此时舍入误差为−5；当第 N 位上的数是奇数时，第 $N+1$ 位上的 5 就进入，此时舍入误差为+5。与其相对应的概率各为二分之一，这样舍入误差就具备了数学期望为零的性质，即舍入误差达到最小。为此，引入下述“四舍五入五单双”法则。

二、“四舍五入五单双”法则

如果需要 N 位有效数字，那么就舍弃第 $N+1$ 位数及其右边的全部数字。$N+1$ 位后的数字按三种情况选择其舍入法则。

1. 若舍去的数值大于所保留末位的 0.5，则末位加 1。例如：
   (1)有一数值 3.6758，要求 3 位有效数字，则为 3.68。
   (2)有一数值 5.4968，要求 3 位有效数字，则为 5.50。
2. 若舍去的数值小于所保留末位的 0.5，则末位不变。例如：
   有一数值 2.7642，要求三位有效数字，则为 2.76。
3. 若舍去的数值等于所保留末位的 0.5，则将末位凑成偶数。例如：

(1)有一数值7.435,要求三位有效数字,则为7.44。

(2)有一数值8.245,要求三位有效数字,则为8.24。

这条法则不仅使舍入误差达到最小(即使舍入误差的期望值等于零),且可以使末位成为偶数,利于以后的计算。

如果第 $N+1$ 位以下的数本身就是通过舍入来的,则须根据舍入前的数据情况应用以上法则。如有一数值9.35是从9.348舍入得到的,要求2位有效数字,则为9.3,而不是9.4。

整数的修约也应遵守上述规则进行处理。例如24368,只取3位有效数字时,则应为 244×10^2。

第三节　计算法则

在测量过程中,应当先确定测量记录数据的有效数字的位数。在有效数字的运算过程中,必须遵循"先进舍、后运算"的原则。具体计算法则将在后面介绍。

一、有效数字的相对误差和有效数字位数之间的关系

设有效数字为 Y,其舍入误差为 X,则该有效数字相对误差为

$$\frac{X}{Y}$$

例如有效数字

$$4,\quad 3.9,\quad 0.399$$

它们的最大相对误差为

$$\frac{0.5}{4}=\frac{1}{8},\quad \frac{0.05}{3.9}=\frac{1}{78},\quad \frac{0.0005}{0.399}=\frac{1}{798}$$

由上可以看出,有效数字相对误差的大小和它的小数位无关,而仅与有效数字的有效位数有关。有效数字的有效位数越少,则其相对误差越大。有效数位少一位,相对误差约增大10倍左右。

二、有效数字的计算法则

1. 加减计算法则

在加减计算中，各数所保留的小数点后的位数，应与所给各数中小数点后位数最少的相同。例如

$$
\begin{array}{r}
13.\overset{?}{4} \\
3.1\overset{?}{5} \\
0.11\overset{?}{3} \\
+) \\
\hline
16.\overset{???}{663}
\end{array}
$$

“?”表示该数带有若干舍入误差。因为有效数中允许末一位为可疑数字，依“先进舍，后运算”的原则，最后运算结果为 16.7，即

$$13.4 + 3.2 + 0.1 = 16.7$$

2. 乘除计算法则

利用上面的分析方法可以得出乘除计算法则。在乘除计算中，应以有效数字最少的或相对误差最大的数字为标准，而与小数点的位置无关；所得积或商的精确度也不大于相乘、除各数值中精确度最小的数值的精确度。

例如进行(853.12×0.41)÷5.014 运算时，各数中以 0.41 的有效数字的位数最小。进行运算时，应“先进舍，后运算”，可得结果为

$$[(85 \times 10^2) \times 0.41] \div 5.0 = 70$$

3. 开方计算法则

在开方计算中，其结果应与原有效数字的有效位数相同。设有

$$Z = Y^{\frac{1}{2}}$$

取自然对数得

$$\ln Z = \frac{1}{2}\ln Y$$

微分得

$$\frac{1}{Z}dZ = \frac{1}{2} \cdot \frac{1}{Y}dY$$

以误差代微分得

$$\frac{\Delta Z}{Z} = \frac{1}{2} \cdot \frac{\Delta Y}{Y}$$

由上式可看出，计算结果 Z 的相对误差为有效数字 Y 的相对误差的一半。因此可以认为计算结果 Z 的有效数字的位数与原有数字 Y 的有效位数相同。

4. 对数计算法则

在对数计算中，所取对数的位数应与真数有效数字的位数相同。若需查对数表时，按真数有效数字的位数选择对数表。设有

$$Z = \lg Y$$

求微分

$$dZ = \frac{1}{\ln 10} \cdot \frac{dY}{Y}$$

以误差代微分

$$\Delta Z = \frac{1}{\ln 10} \cdot \frac{\Delta Y}{Y}$$

即

$$\Delta \lg Y = \frac{1}{\ln 10} \cdot \frac{\Delta Y}{Y}$$

在前面我们已指出，有效数字的相对误差仅与有效数字的位数有关，因此由上式可知对数的误差大小仅取决于真数 Y 的有效位数。例如数字 4967.23 有 6 位有效数字，求其对数时，因

$$\Delta \lg Y = \frac{1}{\ln 10} \cdot \frac{0.005}{4967.23} \approx 0.43 \times \frac{1}{1000000} = 0.00000043$$

6 位对数本身的最大误差为 0.0000005，可以看出它与真数误差引起的对数误差相仿，因而对数应取 6 位。可以得出结论：在对数计算中，真数与对数的有效位数应相同。

5. 三角函数计算法则

利用上面的分析方法可得出三角函数的有效位数应随角度误差减小而增加的结论。角度精度与三角函数位数的关系如表1-2所示。

表 1-2

角度精度	三角函数位数
10″	5 位
1″	6 位
0.1″	7 位
0.01″	8 位

由表 1-2 可以得出如下结论：在精度要求不高的一般性测量的坐标计算中，三角函数应取 5～6 位；在精度要求高的控制测量的坐标计算中，三角函数取 7～8 位即可满足精度要求。

第二章　测量误差理论与权

第一节　观测误差及其分类

一、测量及误差的概念

测量与观测有关，只有对某量进行了观测才会得到观测数据。因此，测量和观测这两个术语，具有相同的函意。测量实际上指的是操作或操作过程，其中包括这种操作过程的实际结果。

由于受仪器、观测者(即人)和外界条件的影响，使得一个被观测量的各个值之间出现差异。这种差异是一种必然现象，也就是说任何被观测量都不可能准确无误地予以测定。通常将某量的真值与其观测值之差称为观测值的误差。任何观测值都必然带有观测误差。

引起观测误差的原因很多，概括起来有以下三个方面原因产生的误差：仪器误差、人差及受外界条件影响的误差。仪器误差是由于仪器结构上的缺陷和精度的限制，使观测值含有误差。人差是指由于受观测者感官能力的限制及技术水平的影响等对观测带来的误差。外界条件影响的误差主要指不断变化着的空气温度、湿度、风力、大气折光等，导致观测结果产生误差。

二、误差种类

观测误差按其性质分为系统误差和偶然误差两种。

1. *系统误差*

在一定观测条件下，如果观测误差保持同一数值和符号，或在