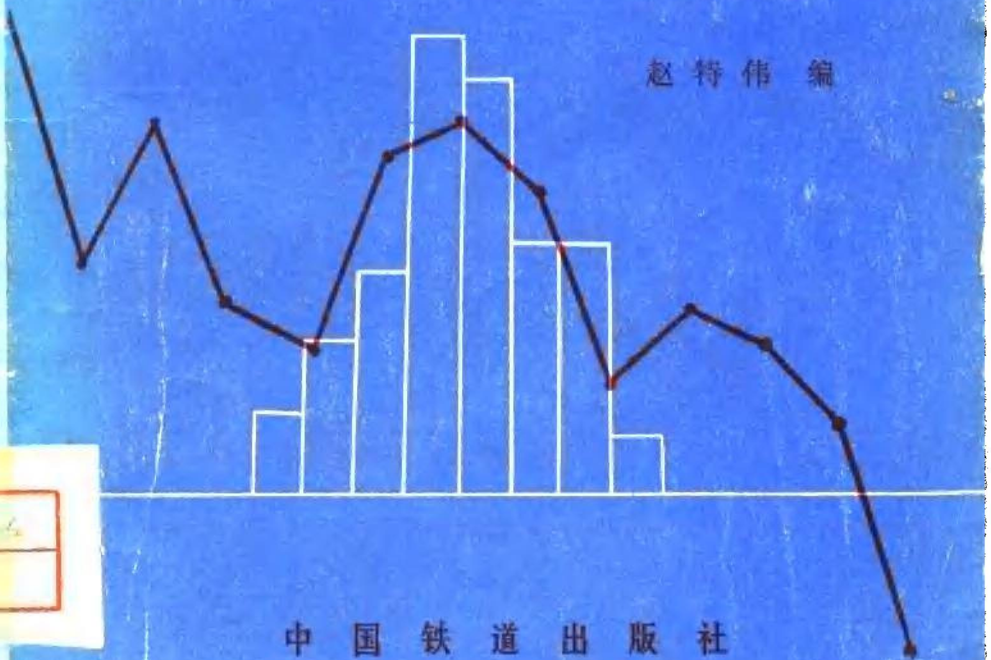
试验数据的整理与分析

赵特伟 编



中国铁道出版社

内 容 提 要

全书除四个附表外，共分十章。第一章讲测定值的数值计算。二至六章介绍概率论和数理统计内容，其中二至三章介绍预备知识，四、五、六三章叙述基本方法及其应用。七、八、九三章介绍误差理论与最小二乘法，其中第七章讲基本概念，八、九两章讲应用，为本书的重点。第十章讲根据数据进行质量控制的基本方法及应用。

本书可供科研工作者、工程技术人员及大专院校有关师生参考。

试验数据的整理与分析

赵特伟　编

中国铁道出版社出版

责任编辑　蒋传瀚

封面设计　翟　达

新华书店北京发行所发行

各 地 新 华 书 店 经 售

中国铁道出版社印刷厂印

开本：$787\times1092\frac{1}{32}$ 印张：12 字数：274 千

1981年4月 第1版　1981年4月 第1次印刷

印数：0001—12,000册　定价：1.60元

前　言

无论是在生产斗争还是在科学实验中，为了了解事物的特性及事物与事物之间的相互联系，常需要进行各种各样的试验。通过试验，就会得到一系列的试验数据。但是，在大多数情况下，一堆未经加工的试验数据（这样的数据有时被称之为“生的”），往往是什么问题也说明不了的。因此，如何科学地对试验数据进行整理与分析，以尽可能简明的方式把试验数据中所包含的有用信息提炼出来，最大限度地发挥它们的作用，是一个十分重要的问题。如果试验数据不能得到很好的利用，试验工作就会功亏于一篑，造成人力与物力的浪费。

自然，随着试验目的、测定方法和研究对象的性质的不同，整理分析试验数据的具体方法也不完全相同。本书的目的，就是要介绍这方面的一些基本原理与方法，而重点在于说明方法的应用。对于所介绍的每一种方法，都力求说清楚它们的基本思路、具体步骤及在什么情形下适用等问题。至于有关定理与结论在理论上的严格证明与推导，则不是本书叙述的重点。虽然文中有时也介绍一些理论证明与推导，但其主要目的不仅仅是在于要得出有关结论的本身，而是在于通过介绍证明与推导而使读者加深对概念与结论的理解。

本书承许成业同志仔细地审阅了原稿，提出了宝贵的修改意见，在此，作者表示衷心的感谢。

作　者

1980年11月

目 录

绪　论

我们经常所进行的试验，按其目的的不同，大致可分为以下三类：

1）在固定条件下对事物的某些特性量所进行的测定或重复测定，其目的是要了解一定条件下事物的某些特性值。

例如，对物质密度的测定；分析化合物中的含量；对结构自由振动特性（自由振动的周期、阻尼等）的测定；等等。

2）将试验的条件有计划地进行改变，在不同的条件下对事物的某些特性量进行一次或多次重复测定，其目的是要了解条件变化对事物特性的影响。

例如，在进行结构试验时，往往就应考虑在不同的位置加载，而荷载也应考虑有不同的等级。又如，为了了解温度对某化学反应的影响，就要在不同的温度下进行实验。

在试验过程中能够变化的那些条件叫做因素。特别应该指出的是，我们经常把时间看作是一个特殊的因素。

若在整个试验过程中仅考虑一个因素的变化，则这样的试验称为单因素试验。若同时考虑几个因素的变化，则称为多因素试验。

3）在一定条件下，对两个或多个不同的事物特性量同时进行多次测定，目的是要了解这些事物特性彼此之间的相互关系。

例如，对河流的流量与水位同时进行测量；对机车车辆的振动与桥梁的振动同时进行测定；等等。

所谓测定值，指的是：在进行试验时，用测定器械对某

些事物的特性量进行测定，最后从记录装置上所读得的数值。测定值是试验数据的主要组成部分。这里所说的试验数据，是泛指与试验有关的数据，除测定值及通过直接对测定值进行数学运算而得到的数值而外，往往还包括诸如试验时间、测定的序号、试验荷载的位置与大小、配合比、所控制的温度等等一些从数量上反映试验条件变化情况的数值。在以后的讨论中，一般将仅认为测定值是含有误差的，而对于与测定值无直接有关的试验数据，则认为它们没有误差，或即使有误差，但相对于测定值的误差而言，它们的误差小到可以忽略不计的程度。

为了得到真正有用的试验数据，我们就必须在进行试验工作之前作好周密的计划，以做到有的放矢，避免徒劳无功。

在制订试验计划时，通常应考虑以下几点：

1）试验的目的与意义；

2）要达到试验的目的，需要哪些数据，应对哪些特性量进行测定；

3）用什么样的测定方法，需要什么样的仪器设备；

4）如何确定试验的条件，哪些因素需要变化，如何变化；

5）试验数据如何整理与分析，通过分析能够得出什么样的结果。

要搞好试验的计划，首先必须具有有关专业的业务知识与实践经验，同时，还应该掌握一定的数学方法。现在，在数理统计中已经发展形成了一个独立的重要分支——试验设计法。它就是专门讨论如何合理地安排试验以及由此而得到的试验数据应如何分析等问题的学科。关于这方面的内容，已有不少书籍作了专门介绍，本书中就不准备介绍了。

第一章　测定值的数值计算

由于在试验数据的整理分析中，不可避免地要对测定值进行数值计算，因此，在这里先介绍一下有关测定值数值计算的基本知识。

§1　舍入误差与有效数字

1.　有效数字

表示测定值的数值与通常数学上所说的数值，在概念上是不同的。例如，34.5，34.50与34.500等，在数学上都看作是同一个数值，但在表示测定值时就不一样了。为了说明这一点，下面首先从读取测定值时的舍入谈起。

我们知道，不管是用哪一种测定器械，最后从记录装置所得到的读数，其位数都是有限的，而不可能要求得到任意多位的读数。例如图1—1所示，指针所指处并无刻度（之

图1—1　读数的舍入

所以不再有进一步细分的刻度，其原因往往有两个：一是刻度已甚小，无法再细分；一是由于仪器本身的精度所限，再细分没有什么意义了），此时，很自然地，我们应将它读成离指针最近的那个数，即113。而实际上，准确的读数值 a 并不正好等于113，而只是满足下列不等式：

$$112.5 \leqslant a \leqslant 113.5$$

例如，上述的 a 也可能是 112.8……，也可能是 113.37……，

等等,此时我们说，测定值113的数值直至最后一位都是有意义的，其最后一位数 3 是经过舍入得来的，至于其后面位上的数字就不清楚了。在表示测定值的数值中，有意义的数字就叫做有效数字。例如上面的113,它有三位有效数字。同样，我们若说某量的测定值为34.50，则意味着其准确的读数值 a 满足

$$34.495 \leqslant a \leqslant 34.505$$

也就是说，测定值 34.50 的有效数字有四位。而若说测定值为34.5，则意味着

$$34.45 \leqslant a \leqslant 34.55$$

测定值的有效数字为三位。

在表示测定值时，有效数字最好能明确地表示出来。以光速为例，若光速表示为

$$C = 29979600000\text{厘米/秒}$$

则看不出到底有几位有效数字，若记为

$$C = 2.99796 \times 10^{10}\text{厘米/秒}$$

则就可知道该测定值的有效数字为六位。若再把它近似为

$$C = 3.00 \times 10^{10}\text{厘米/秒}$$

则有效数字只有三位了。

2. 舍入误差与舍入规则

由于测定器械的分辨能力所限，只能通过舍入而读取一定位数的测定值，由此而造成的误差即称之为舍入误差或凑整误差。设舍入误差为ε_r，则

$$\varepsilon_r = x - a \qquad (1-1)$$

式中 x ——测定值（经过舍入后的读数值）；

a ——准确的读数值。

实际上，由于准确的读数值一般是不知道的，因此，舍入误差也是无法知道的。在处理具体问题时，我们往往只是设

法去确定它们的一个范围。例如，对于上面所谈到的测定值113，由

$$x=113$$

及
$$112.5\leqslant a\leqslant 113.5$$

即推得：
$$|\varepsilon_r|\leqslant 0.5$$

而对于上述的 $x=34.50$，则有：

$$|\varepsilon_r|\leqslant 0.005$$

除了上述的那种在测定过程中因测定器械的分辨能力所限而造成的舍入误差以外，我们还常会遇到另一种舍入误差——在计算过程中因计算工具所能表示的数值的位数有限而引起的舍入误差。例如，我们用计算尺进行运算，其结果一般只能得到三位数，这里实际上就可能产生舍入误差。即使是位数很多的大型电子计算机，其所能表示的数的位数也是有限的，因此，照样也有舍入的问题。

例如，在进行运算中要用到圆周率π，而π的准确值应为

$$\pi=3.14159265358979323846264 33\cdots$$

而实际进行计算时，必须对它进行舍入。若取三位，即 $\pi\approx 3.14$，则舍入误差为

$$\varepsilon_r=-0.001592653589793\cdots$$

若取五位，$\pi\approx 3.1416$，则舍入误差为

$$\varepsilon_r=0.000007346\cdots$$

下面，接着来谈一谈如何对数值进行舍入的问题。

过去习惯的做法是“四舍五入”，凡小于或等于 4 的数就“舍”（即弃去），凡大于或等于 5 的数则“入”（即在前一位上进 1）。其实，这样做有一个缺点，就是在要舍入的那位数为 5 时，总是造成正的舍入误差。我们知道，要舍入的那位数可能取的数值是

1，2，3，4，5，6，7，8，9

若用四舍五入，则该舍的是四个数——1，2，3，4，而该入的是五个数——5，6，7，8，9。在处理大量数值时，这些数字中每一个出现的机会都是彼此相同的。再从舍入误差的大小来看，当需要舍入的末位分别为1，2，3，4，5，6，7，8，9时，进行四舍五入后产生的舍入误差在这一位上的数值相应地为

－1，－2，－3，－4，5，4，3，2，1

由此可见，1与9，2与8，3与7，4与6，这每两个数之间，彼此的舍入误差大小相等，而符号相反，而5在四舍五入时则只产生正的舍入误差，且舍入误差的绝对值比其他数的大。因此，若遇5就入，则将会造成正的舍入误差出现的机会多。而这一点往往对于数值计算是不利的。例如，在对大量的测定值（或近似值）求和时，由于正的舍入误差占优势，就可能会使计算结果偏大。

为了使正负舍入误差出现的机会大致均等，现已普遍采用如下的舍入规则：

1）四舍六入；

2）若要进行舍入的那位数为5时：

① 若其后还有数字，则入；

② 若其前一位为奇数，则入；

③ 若其后不再有数字，且前一位为偶数，则舍。

例如，要求将下列数值均舍入成三位数字，则

345.6⟶346　　　0.5234⟶0.523

1.2054⟶1 21

9.375⟶9.38

12.45⟶12.4

§2 计算结果的误差分析

举一个例子。设我们要计算矩形截面的惯性矩 I，计算公式为

$$I=\frac{1}{12}bh^3$$

式中 b ——矩形截面的宽；

h ——矩形截面的高。

若已知 $b\approx 43$ 厘米， $h\approx 109$ 厘米，前者有两位有效数字，后者有三位有效数字。将这些数值代入计算公式，得

$$I=\frac{1}{12}bh^3\approx 4640521\text{厘米}^4$$

这里一共给出了七位数字，这些数字是否都有意义呢？为了回答这个问题，我们根据 b 与 h 的误差来估计 I 可能取值的范围。由已知条件有

$$42.5\leqslant b\leqslant 43.5$$

$$108.5\leqslant h\leqslant 109.5$$

从而可推出

$$4.5237\cdots\times 10^6\leqslant\frac{1}{12}bh^3\leqslant 4.759\cdots\times 10^6$$

可见， I 的真正数值最小可能为 $4.5\cdots\times 10^6$，最大可能为 $4.7\cdots\times 10^6$，因此，根本没有必要写成 $I=4640521$ 厘米^4 那样的答案，而将结果表示成 4.6×10^6 就可以了，或至多是表示成 4.64×10^6。

由此可见，测定值的误差会给计算结果带来一定的影响。下面将会看到：有时候，这种影响还是相当大的。又因为测定值本身就是一种近似值，故我们干脆直接对一般的近似值进行讨论。

为了叙述方便，我们把准确值记为$a_1, a_2, \cdots$，而相应的近似值记为$x_1, x_2, \cdots$，误差记为$\Delta x_1, \Delta x_2, \cdots$；把用准确值$a_1, a_2, \cdots$进行计算所得到的准确结果记为$b$，而把用近似值$x_1, x_2, \cdots$进行计算所得到的近似结果记为$y$，其相应的误差记为$\Delta y$。亦即

$$x_i = a_i + \Delta x_i \qquad i = 1, 2, \cdots \qquad (1-2)$$

$$y = b + \Delta y \qquad (1-3)$$

此外，我们把$\frac{\Delta y}{y}$，$\frac{\Delta x_1}{x_1}$，$\frac{\Delta x_2}{x_2}$，…等分别称为y，x_1，x_2，…等的相对误差，常以百分率来表示。而为强调计，有时也把$\Delta y, \Delta x_1, \Delta x_2$，…等称为绝对误差。

下面，我们将分别对各种运算进行讨论。

A、加法与减法

先考虑最简单的情形：$b = a_1 \pm a_2$。此时，用近似值计算所得的结果为

$$\begin{aligned} y &= x_1 \pm x_2 = (a_1 + \Delta x_1) \pm (a_2 + \Delta x_2) \\ &= (a_1 \pm a_2) + (\Delta x_1 \pm \Delta x_2) = b + (\Delta x_1 \pm \Delta x_2) \end{aligned}$$

从而得到

$$\Delta y = \Delta x_1 \pm \Delta x_2 \qquad (1-4)$$

类似地，若考虑

$$b = a_1 \pm a_2 \pm \cdots \pm a_n \qquad (1-5)$$

则有

$$\Delta y = \Delta x_1 \pm \Delta x_2 \pm \cdots \pm \Delta x_n \qquad (1-6)$$

若要估计误差Δy的限界，则可得

$$|\Delta y| \leqslant |\Delta x_1| + |\Delta x_2| + \cdots + |\Delta x_n| \qquad (1-7)$$

B、乘法与除法

先考虑乘法：$b = a_1 \cdot a_2$。此时，由近似值所得的计算结果为$y = x_1 \cdot x_2$，从而y的误差为

$$\begin{aligned}\Delta y&=y-b=x_1\cdot x_2-a_1\cdot a_2\\&=x_1\cdot x_2-(x_1-\Delta x_1)(x_2-\Delta x_2)\\&=x_2\cdot\Delta x_1+x_1\cdot\Delta x_2-\Delta x_1\cdot\Delta x_2\end{aligned}$$

将上式两边除以y，

$$\frac{\Delta y}{y}=\frac{\Delta x_1}{x_1}+\frac{\Delta x_2}{x_2}-\frac{\Delta x_1}{x_1}\cdot\frac{\Delta x_2}{x_2}$$

由于$\frac{\Delta x_1}{x_1}$与$\frac{\Delta x_2}{x_2}$一般均很小，故上式中最后一项与前两项相比可忽略不计，从而

$$\frac{\Delta y}{y}\approx\frac{\Delta x_1}{x_1}+\frac{\Delta x_2}{x_2} \tag{1—8}$$

或

$$\left|\frac{\Delta y}{y}\right|\leqslant\left|\frac{\Delta x_1}{x_1}\right|+\left|\frac{\Delta x_2}{x_2}\right| \tag{1—9}$$

现考虑除法：$b=a_1/a_2$。此时，

$$\Delta y=y-b=y-\frac{a_1}{a_2}=y-\frac{x_1-\Delta x_1}{x_2-\Delta x_2}$$

$$=y-\frac{x_1}{x_2}\cdot\left(\frac{1-\frac{\Delta x_1}{x_1}}{1-\frac{\Delta x_2}{x_2}}\right)=y-y\left(1-\frac{\Delta x_1}{x_1}\right)\Big/\left(1-\frac{\Delta x_2}{x_2}\right)$$

$$\approx y-y\left(1-\frac{\Delta x_1}{x_1}\right)\left(1+\frac{\Delta x_2}{x_2}\right)$$

$$=y\cdot\frac{\Delta x_1}{x_1}-y\cdot\frac{\Delta x_2}{x_2}+y\cdot\frac{\Delta x_1}{x_1}\cdot\frac{\Delta x_2}{x_2} \tag{1—10}$$

注意，这里利用了近似公式（见§4表1—3）

$$\frac{1}{1-\delta}\approx1+\delta\quad（当\delta很小时）$$

将（1—10）式两边除以y，并略去最后一项，得

$$\frac{\Delta y}{y}\approx\frac{\Delta x_1}{x_1}-\frac{\Delta x_2}{x_2} \tag{1—11}$$

或
$$\left|\frac{\Delta y}{y}\right| \leqslant \left|\frac{\Delta x_1}{x_1}\right| + \left|\frac{\Delta x_2}{x_2}\right| \qquad (1-12)$$

更一般地，若

$$b = a_1 \cdot a_2 \cdots a_m / (a_{m+1} \cdot a_{m+2} \cdots a_n) \qquad (1-13)$$
$$(1 \leqslant m \leqslant n)$$

则可推得

$$\frac{\Delta y}{y} \approx \frac{\Delta x_1}{x_1} + \frac{\Delta x_2}{x_2} + \cdots + \frac{\Delta x_m}{x_m} - \frac{\Delta x_{m+1}}{x_{m+1}} - \frac{\Delta x_{m+2}}{x_{m+2}} - \cdots - \frac{\Delta x_n}{x_n} \qquad (1-14)$$

或
$$\left|\frac{\Delta y}{y}\right| \leqslant \left|\frac{\Delta x_1}{x_1}\right| + \left|\frac{\Delta x_2}{x_2}\right| + \cdots + \left|\frac{\Delta x_n}{x_n}\right| \qquad (1-15)$$

C、乘幂与方根

这里我们讨论：

$$b = a_1^m \qquad (m > 0)$$

利用当 $x \approx a$ 时成立的近似公式

$$f(x) - f(a) \approx f'(x)(x-a) \qquad (1-16)$$

$y = x_1^m$的误差为

$$\Delta y = y - b = x_1^m - a_1^m \approx (x^m)'_{x=x_1}(x_1 - a_1)$$
$$= m x_1^{m-1} \Delta x_1$$

上式两边除以 y，

$$\frac{\Delta y}{y} \approx m \frac{\Delta x_1}{x_1} \qquad (1-17)$$

由此可见，x_1的 m 次乘方的相对误差等于x_1本身的相对误差的m倍。而x_1的 n 次方根的相对误差则等于x_1本身相对误差的$1/n$，这是因为：

$$b = \sqrt[n]{a_1} = a_1^{\frac{1}{n}}$$

D、对数