工程数据处理

王广铨　罗传义　编著

吉林大学出版社

内 容 简 介

本书介绍了工程数据处理的常用方法。全书共分六章：误差理论、实验数据整理、统计检验、方差分析、回归分析、非线性模型参数估值。各章均附有习题。书末附有必要的统计用表。

本书可作为开设数据处理课的工科院校学生的教学用书或参考书，亦可供有关研究人员、工厂技术人员参考。

工 程 数 据 处 理

王广铨 罗传义 编著

吉林大学出版社出版	吉林大学出版社发行
（长春市东中华路29号）	东北师大校办印刷厂印刷
开本：787×1092毫米1/32	1990年8月第1版
印张：7.4375	1990年8月第1次印刷
字数：163千字	印数： 1—1080册

ISBN 7—5601—0567—X/O·65 定价：3.05元

前　　言

工程数据处理对广大科研工作者、工厂技术人员、企业管理人员是很有用的数学工具。在科学实验、工程设计、企业管理以及工科院校的大学生和研究生做毕业论文或毕业设计等实际工作中，经常遇到数据处理问题。因此，掌握数据处理的一些方法和技巧是十分必要的。随着科学技术的发展和电子计算器、电子计算机的普遍应用，人们对掌握数据处理的方法、技巧的要求愈来愈迫切。为了满足实际需要，我们编写了这本书。

本书较为详细地介绍了一些数据处理方法。在编写上，力求做到通俗易懂，在内容上尽量做到少而精，并注意从实际需要出发，侧重介绍结论和用法。这样做其目的是想使读者花费较少的时间和精力，掌握更多的方法。

本书第三章及全书的图稿由黄文焕同志完成。

作　者

1990年3月于吉林

目　　录

第一章 误差理论

1.1 误差研究的意义

对自然界所发生的量变现象的研究，常常需要借助于各式各样的实验和测量来完成。由于测量所用仪器或工具本身精度的限制，测试方法的不完善，测试环境的变化等客观因素的影响，也由于测试人员的技术水平、经验等主观因素的影响，实验中测得的值与真实值并不完全一致。这种差异在数值上的表现即为误差。

随着科学水平的提高和人们的经验、技巧、专门知识的丰富，实验中的误差可以逐渐减小，但实验工作始终不能做到没有误差。误差自始至终存在于一切科学实验的过程之中。我们研究误差的目的并不是使误差趋于零，或降低到不能再小的程度，因为这往往是办不到的。况且为了进一步减小误差，往往需要花费大量的人力和物力。我们研究误差的目的在于：

（1）正确处理实验数据。充分利用数据信息，在一定条件下，得到更接近真实值的最佳结果。

（2）合理选取所得结果的误差。既不能人为地将误差算得过份小，以免对生产造成危害，也不能算得过份大，以免造成人力物力的浪费。

（3）合理地选择实验仪器、条件和方法，以便以最经济的方式得到满足要求的结果。

对于从事实验科学的工作者，掌握实验数据的科学处理方法，熟悉有关误差的基本理论是十分必要的。

1.2 真值与平均值

1.2.1 真值

什么是真值？一般地说，由于自然界中的一切物体都处于永恒的运动中，客观存在的真值是不可能准确知道的。理论上说，真值是指测定次数无限多时求得的平均值。人们通常所说的真值有如下几种：

（1）理论真值。例如，平面三角形的三内角之和恒为180°。又如，同一量值自身之差为零，而自身之比为1。此外还有理论设计和理论公式表达值等等。

（2）计量学约定真值。它指的是国际会议和标准化组织或国际上公认的量值。

（3）相对真值。如国家标准样品的标称值或用标准仪器所测得的值等。

此外，不可能准确知道一个量的真值。因为我们平时测定的次数总是有限的，所以其平均值只能是真值的估计量或称近似值。

1.2.2 平均值

在处理实验结果时，常用的平均值有以下几种：

1. 算术平均值

算术平均值是常用的一种平均值。设x_1，x_2，…，x_n代表各次测定值，n代表测定次数，则算术平均值

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{1.1}$$

算术平均值的一个重要性质，就是若测定值的分布服从正态分布，则算术平均值即为一组等精度测量中的最佳值，或称为最可信赖值。

2．加权算术平均值

设x_1，x_2，…，x_n代表各次测定 值，ω_1，ω_2，…，ω_n代表各次测定值对应的权，则加权算术平均值（或简称加权平均值）

$$\bar{x}_\omega=\frac{\sum_{i=1}^{n}\omega_i x_i}{\sum_{i=1}^{n}\omega_i} \tag{1.2}$$

加权平均值中的权应该酌情而定，如在不等精度的测量中，认为某一精度的测定值较好，即可给以较大的权值，加重它在平均值中的份量。

我们也可以将权理解为测定值x_i在很大的测量总数N中出现的频率n_i/N，如代之以概率p_i来表示，则式(1.2)可改写为

$$\bar{x}_\omega=\sum_{i=1}^{n}p_i x_i \tag{1.3}$$

3．对数平均值

在化学反应、热量传递及质量传递中，其分布曲线多具有对数的特性，在这种情况下表征平均值的量就应该用对数平均值。

设两个量x_1、x_2，其对数平均值是这两个量的差与它们的自然对数的差之比，即

$$x_l=\frac{x_1-x_2}{\ln x_1-\ln x_2}=\frac{x_1-x_2}{\ln\frac{x_1}{x_2}} \tag{1.4}$$

可以指出，变量的对数平均值总小于算术平均值。如果变量相差甚小时，可以用算术平均值代替对数平均值，而并无多大误差。x_1与x_2间之差愈小，其误差亦愈小。对于这一点，可作如下证明：设

$$\frac{x_1}{x_2}=\frac{1+a}{1-a}$$

则

$$a=\frac{x_1-x_2}{x_1+x_2}$$

当x_1与x_2相差很小时，$a\ll 1$，此时可采用近似公式

$$\ln(1+a)=a,\ln(1-a)=-a$$

由对数平均值的定义式及上面几个公式

$$\begin{aligned}x_l&=\frac{x_1-x_2}{\ln\frac{x_1}{x_2}}\\&=\frac{x_1-x_2}{\ln\left(\frac{1+a}{1-a}\right)}=\frac{x_1-x_2}{\ln(1+a)-\ln(1-a)}\\&\approx\frac{x_1-x_2}{2a}=\frac{x_1-x_2}{2\left(\frac{x_1-x_2}{x_1+x_2}\right)}\\&=\frac{x_1+x_2}{2}\end{aligned}$$

即在所讨论的条件下，对数平均值近似地等于算术平均值。

通常，若比值（x_1/x_2）<2，则可以用算术平均值来代替对数平均值，引起的误差不超过4%。

4．几何平均值

几何平均值是将n个测定值相乘后再开n次方所得的值。即

$$x_g=(x_1\cdot x_2\cdot\cdots\cdot x_n)^{1/n} \tag{1.5}$$

或以对数形式表示

$$\ln x_g=\frac{1}{n}\sum_{i=1}^{n}\ln x_i \tag{1.5}$$

5．均方根平均值

简称均方根值，在统计学学中被广泛应用。其定义式为

$$u=\left(\frac{\sum_{i=1}^{n}x_i^2}{n}\right)^{1/2} \tag{1.7}$$

以上介绍的几种平均值都是想从一组测定数据中找出最接近真值的那个值。平均值的选择主要决定于实验数据的分布类型。如当实验数据服从正态分布（参见1.6）或近似服从正态分布时，应采用算术平均值；当实验数据服从对数正态分布时，则应采用几何平均值。本书所讨论的实验数据均假定服从正态分布或近似服从正态分布，故后面所讨论的平均值均指算术平均值。

1.3 误差的表示方法

常见的误差表示方法有以下几种：

1．绝对误差

某量值的绝对误差定义为该量的给出值x（包括实验值、计算近似值等要研究和给出的非真值）与真值A（包括理论真值、约定真值和相对真值等）之差，即

$$\Delta_x = x - A \tag{1.8}$$

如果误差的绝对值愈小，表示结果与真值愈接近，给出值的准确度愈高；反之，误差的绝对值愈大，给出值的准确度愈低。若给出值大于真值，误差为正值；反之，误差为负值。绝对误差是反映给出值偏离真值大小的。

请读者注意，严格地说，绝对误差并非误差的绝对值。例如，用标准仪器测得某物理量的值为1.728（可看作是真值A），而用另一台普通仪器测得该物理量的值为1.730，则测量值的绝对误差为

$$\Delta_x = 1.730 - 1.728 = 0.002$$

若另一次测量值为1.725，其绝对误差为

$$\Delta_x = 1.725 - 1.728 = -0.003$$

绝对误差是有单位的，其单位与给出值的单位相同。

在绝大多数情况下，由于真值无法知道，常常需要借助于误差范围来表示误差。任何测量都有一定的误差范围，因此可以用测量的误差范围来确定一个量的真值范围。例如，用分析天平测得一块金属重3.2803(g)，已知分析天平称量的误差范围是±0.0001(g)，则该金属块的真值范围是3.2803±0.0001(g)。这里所说的误差范围又称为最大绝对误差。习惯上，人们又把最大绝对误差也称为绝对误差。最大绝对误差的量值前面一般都加“±”号，这是同式(1.8)所定义的绝对误差不同的。

2. 相对误差

相对误差是指绝对误差在真值中所占的百分率，即

$$E_x = \frac{\Delta_x}{A} \times 100\% \tag{1.9}$$

在误差较小时，测定值x与真值A接近，故人们常常将绝对误差与测定值之比作为相对误差，即

$$E_x = \frac{\Delta_x}{x} \times 100\% \tag{1.10}$$

例1-1 用分析天平测得某一金属块的重量为4.1854(g)，而用半微量天平测得该金属块的重量为4.18544(g)（可看作真值），试求相对误差。

解 $\Delta_x = x - A$

$= 4.1854 - 4.18544$

$= -0.00004(\text{g})$

相对误差

$$E_x = \frac{\Delta_x}{A} \times 100\%$$

$$= \frac{-0.00004}{4.18544} \times 100\% = -0.001\%$$

或

$$E_x = \frac{\Delta_x}{x} \times 100\%$$

$$= \frac{-0.00004}{4.1854} \times 100\% = -0.001\%$$

用最大绝对误差计算出的相对误差称为最大相对误差。如例1-1中测量值4.1854(g)的最大绝对误差为±0.0001(g)，则最大相对误差为

$$\frac{\pm 0.0001}{4.1854} \times 100\% = \pm 0.0025\%$$

在用同一仪器进行测量时，不论被测量的量值多大，它们的最大绝对误差都是相同的。但相对误差却有不同。例如用台称称量两个物体的重量分别为102(g)和5(g)，称量的最大绝对误差为±1(g)，而相对误差分别为

$$\frac{\pm 1}{102}\times 100\% = \pm 1\%$$

$$\frac{\pm 1}{5}\times 100\% = \pm 20\%$$

可见用相对误差来表示实验结果的误差，才能准确地表明实验结果是否有意义。

在实际工作中，往往遇到这种情况，为了获得更准确的结果，在相同条件下需进行多次重复测定。这种情况下的重复测定又称为平行测定或等精度测定。这时若测量仪器已经进行过校准，则测定值的误差往往用下面的几种方法表示。

3．极差

是一组测定值中最高值与最低值之差，用以说明误差变化的范围。极差的定义式为

$$R = x_{max} - x_{min} \tag{1.11}$$

式中x_{max}，x_{min}分别为测定值中的最高值和最低值。

极差的最大缺点是显示不出与测量次数的关系，而测量的结果与真值的近似程度是同测量次数密切相关的。

4．算术平均误差

算术平均误差（或称为平均偏差）简称为平均误差。其定义为

$$\delta = \frac{\sum_{i=1}^{n} | x_i - \bar{x} |}{n} \tag{1.12}$$

式中n为测定次数，$\bar{x}$为测定值的算术平均值，x_i为第i次测定值。

一直到现在，人们常用算术平均误差来表示测定值的误差，其主要原因是手算简单。事实上，在袖珍电子计算器已经普及的今天，算术平均误差的计算要比接下来要讲的标准差计算麻烦得多，并且表示测定值的误差用算术平均误差不如用标准差更好些。

5．标准差

标准差是标准误差的简称，又称为标准偏差。当测定次数为无穷时，其定义为

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1.13}$$

在有限次测定中，标准差可用下式表示，即

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1.14}$$

标准差是现在文献中常用的一种表示误差的较好的方法。它不但与一组测定值中每个数据有关，而且对其中较大误差或较小误差敏感性很强，能明显反映出较大的个别误差，这是算术平均误差所不及的。实验愈精确，标准误差愈小。用袖珍电子计算器计算标准差是相当容易的。

1.4 误差的来源及分类

在本章开始时我们说过，实验工作始终不能做到没有误差，所以真值是无法准确知道的，测定值永远是真值的近似

值。这是什么原因呢？分析一下实验数据的误差来源和性质便可明了。

误差根据其性质可分为系统误差、随机误差和过失误差。

1.4.1 系统误差

系统误差也叫可定误差，它是由于实验过程中某些经常发生的原因所造成的，它对测定结果的影响比较固定，在同一条件下重复测定时，它会重复出现。因此，该误差的大小往往可以估计，并可设法减小或加以校正。系统误差产生的主要原因有：

1．方法误差

方法误差亦称为理论误差。这是由于测量方法本身形成的误差，或者由于测量所依据的理论本身不完善等原因而导致的误差。例如用伏安法测电阻时，电表内阻的影响；在化学分析中的重量分析中，由于沉淀物质的溶解度较大，造成沉淀溶解损失等。

2．仪器误差

主要是仪器本身不够准确或未经校准所引起的。如仪器的零点不准，仪器的制造工艺所导致的仪器精密度不高，长期使用引起的磨损等原因所产生的误差等。

3．操作误差

由于操作人员的主观原因所造成的。如用停秒表计时，有的人习惯超前或滞后，都会带来计时的误差。

1.4.2 随机误差

随机误差也称为偶然误差。当在测量中，已经消除了引

起系统误差的一切因素，并且观测者又正确细心地进行测量，但重复测定时，所得的结果并不完全一致。这就是由于随机误差所造成的。随机误差是由于很多无法估计的、各种各样的随机原因所引入的误差。在单次测定中，随机误差的大小及其符号是无法预言的，没有任何规律性。但在多次等精度测定中，随机误差的出现还是有规律的。它具有统计规律性（参见1.6）。

随机误差量值的大小，往往用标准误差来表示。随机误差是不可避免的，实验工作者可设法将它大大减小，但不可能完全消除它。

1.4.3 过失误差

过失误差是由于实验工作中粗枝大叶，操作不正确而引起的误差。其鉴别依据是观测结果与事实不符。只要实验工作人员加强责任心，严格遵守操作规程，一丝不苟，认真细致地进行实验，过失误差是可以避免的。

应该指出，在进行误差分析时要估计的误差只有系统误差和随机误差这两类。

1.5 准确度与精密度的概念

在前面的讨论中，曾用到准确度与精密度的概念，以后的叙述中还将用到。所以，有必要介绍一下准确度与精密度的概念。

1.5.1 准确度

准确度是表示所得测定结果与真实值接近的程度。所得