

试

验数据分析处理 与软件应用

王文健 许 荔 钱海挺 等编著
冼 进 审校

<http://www.phei.com.cn>



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

试验数据分析处理与软件应用

王文健 许 荔 钱海挺 王 卡 等编著

冼 进 审校

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书以试验数据为主题,全面地综合了试验数据分析处理中常用的理论方法和应用软件。

本书第一部分主要介绍数据分析处理的方法与理论,其特点是以理论介绍为基础,又强调分析处理方法的应用,内容主要涉及试验数据表示及误差分析、统计学分析与曲线拟合、方差与估计分析、回归与假设检验分析等基本理论及方法。第二部分主要介绍试验数据分析处理过程的软件实现,主要介绍了 Microsoft Office Excel 2003、OriginPro7.5、MATLAB7.1 和 SPSS 13.0 for Windows 四种常用数据处理软件在试验数据分析处理中的具体应用,最后对各种软件的应用与处理技巧进行了阐述。本书内容全面、叙述清楚,并且非常注重实用性。

本书体系完整、理论翔实、语言通俗易懂,既可作为高等院校专科生、本科生、研究生和实验员的学习教材,也可作为数据分析工作者和工程技术人员的参考用书。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

试验数据分析处理与软件应用 / 王文健等编著. —北京:电子工业出版社, 2008.2
ISBN 978-7-121-05843-1

I. 试… II. 王… III. 实验数据—数据处理—应用软件 IV. N33-39

中国版本图书馆 CIP 数据核字(2008)第 009665 号

策划编辑: 范子瑜

责任编辑: 宋兆武 侯丽平

印 刷: 北京市海淀区四季青印刷厂

装 订: 北京鼎盛东极装订有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1092 1/16 印张: 20 字数: 512 千字

印 次: 2008 年 2 月第 1 次印刷

印 数: 4000 册 定价: 38.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010) 88258888。

前 言

科学试验是科研工作者在研究过程中的重要方法和手段,通过试验可获得大量真实可靠的数据,只有对获得的试验数据进行一系列正确的分析与处理,才能得到可靠的试验分析结果,因此试验数据的分析处理是每位科研工作者必须掌握的一门技术手段。

本书力图全面、系统地介绍数据分析处理方法与软件实现的相关知识,让大家掌握多种有关试验数据分析处理的理论与计算机软件的应用。综观全书,其特点如下。

1) 完整性

本书对试验数据分析处理所用到的方法与理论进行了详细的介绍与说明,可根据需要选择使用,并且所有分析方法均能通过计算机软件进行实现。

2) 实用性

本书紧密结合数据分析处理的各个方面,分门别类地做了详细的介绍。本书语言通俗易懂,内容丰富翔实,突出了以实例为中心的特点。

3) 新颖性

本书先介绍数据分析处理的理论,然后介绍分析方法在计算机软件中的实现过程,这样可节省大量的查阅时间,轻松方便地完成数据的分析处理工作。

本书内容由试验数据分析处理理论和软件应用两部分组成。本书共有9章:第1~4章主要介绍试验数据表示及误差分析、试验数据统计学分析与曲线拟合、试验数据方差与估计分析、试验数据回归与假设检验分析等数据处理的基本理论和方法;第5章介绍了 Microsoft Office Excel 在试验数据分析处理中的具体应用;第6章介绍了 Origin 在试验数据分析处理中的具体应用;第7章介绍了 MATLAB 在试验数据分析处理中的具体应用;第8章介绍了 SPSS 在试验数据分析处理中的具体应用;最后,第9章阐述了各软件在数据处理中的应用特点和技巧。

本书体系完整、理论翔实、语言通俗易懂,既可作为高等院校专科生、本科生、研究生和实验员的学习教材,也可作为数据分析工作者和工程技术人员的参考用书。

本书由王文健、许荔、钱海挺、王卡和黄志平共同编写完成,同时参与本书编写的人员还有邹素琼、冼进、赵秋云、赵继军、彭艺、曲辉辉、周章、蒋波、徐留旺、曹振宇、张婷、温凌霜、鲁得翠、蒋泽平、魏乐、韩翔、程小英、谭小丽、卢丽娟、李小琼、周宏、罗吉、许翔燕、陈春、张忠、方小马、黄姹英、周明、宋晶、邓勇等,在此向他们表示衷心的感谢。在本书编写过程中,引用和参考了部分资料,在此对其作者一并表示感谢。

由于编者水平所限,书中不妥与疏漏之处在所难免,敬请广大读者批评指正。同时欢迎读者将图书使用过程中的问题与各种探讨、建议反馈给我们,本书作者会竭诚给您满意的答复。我们的联系方式 E-mail: china_54@tom.com。

编 者

2007年12月

目 录

第 1 章 试验数据表示及误差分析··· (1)

1.1 试验数据的测量与表示····· (1)

1.1.1 数据的测量····· (1)

1.1.2 数据的表示····· (3)

1.2 误差分析····· (4)

1.2.1 误差的含义····· (5)

1.2.2 误差的类型····· (6)

1.2.3 误差的表示方法····· (7)

1.2.4 精确度、正确度、精密度····· (11)

1.2.5 误差的来源····· (12)

1.2.6 误差的分配····· (13)

1.3 随机误差····· (16)

1.3.1 随机误差的统计特征····· (16)

1.3.2 随机误差的正态分布曲线····· (17)

1.3.3 几种常用的非正态分布····· (17)

1.3.4 随机误差的合成····· (18)

1.4 系统误差····· (18)

1.4.1 基本概念····· (19)

1.4.2 系统误差的合成····· (20)

1.4.3 系统误差的减小与消除····· (21)

1.5 本章小结····· (23)

1.6 习题····· (23)

第 2 章 试验数据统计学分析与 曲线拟合····· (24)

2.1 统计学基础····· (24)

2.2 统计量····· (25)

2.2.1 样本平均值与样本偏差····· (25)

2.2.2 正态样本平均值和偏差 的分布····· (28)

2.3 多元变量统计学····· (29)

2.3.1 多元变量的描述统计量····· (29)

2.3.2 多元变量统计分析····· (31)

2.4 曲线拟合····· (33)

2.4.1 最小二乘原理····· (33)

2.4.2 线性情况的最小二乘拟合····· (33)

2.4.3 非线性情况的最小二乘拟合····· (35)

2.4.4 用最小二乘法做曲线拟合····· (35)

2.4.5 约束条件下的最小二乘拟合····· (38)

2.5 蒙特卡洛方法····· (38)

2.6 本章小结····· (40)

2.7 习题····· (40)

第 3 章 试验数据方差与估计

分析····· (42)

3.1 方差分析····· (42)

3.1.1 单因素方差分析····· (42)

3.1.2 两因素方差分析····· (45)

3.2 估计分析····· (50)

3.2.1 参数估计的基本概念····· (50)

3.2.2 贝叶斯方法····· (50)

3.3 点估计····· (55)

3.3.1 矩估计····· (55)

3.3.2 极大似然估计····· (57)

3.4 区间估计····· (58)

3.4.1 区间估计的意义····· (58)

3.4.2 分布参数的区间估计····· (59)

3.5 本章小结····· (62)

3.6 习题····· (62)

第 4 章 试验数据回归与假设

检验分析····· (64)

4.1 一元回归分析····· (64)

4.1.1 一元线性回归分析····· (64)

4.1.2 一元非线性回归分析····· (67)

4.2 多元回归分析····· (68)

4.2.1 多元线性回归分析····· (68)

4.2.2 曲线回归分析····· (70)

4.3 假设检验的基本概念····· (71)

4.3.1 假设检验问题	(72)
4.3.2 假设检验的原理	(72)
4.4 假设检验方法	(73)
4.4.1 正态总体均值的假设检验	(73)
4.4.2 正态总体方差的假设检验	(76)
4.4.3 非正态总体大样本均值的假设检验	(78)
4.4.4 分布 χ^2 拟合检验	(80)
4.5 本章小结	(80)
4.6 习题	(81)
第5章 Excel 在试验数据处理中的应用	(82)
5.1 Excel 的基础知识	(82)
5.1.1 数据输入	(82)
5.1.2 公式	(84)
5.1.3 工作表	(85)
5.2 数据的格式化	(86)
5.2.1 工作表的格式化	(86)
5.2.2 格式化数据	(89)
5.2.3 格式化日期	(91)
5.2.4 条件格式	(92)
5.3 公式及函数的应用	(94)
5.3.1 公式的引用	(94)
5.3.2 函数的调用	(94)
5.4 图表处理	(96)
5.4.1 图表类型	(96)
5.4.2 图表的建立	(98)
5.4.3 格式化图表	(100)
5.4.4 复杂图表的建立	(106)
5.4.5 自定义图表	(107)
5.5 数据分析处理	(109)
5.5.1 统计分析	(110)
5.5.2 方差分析	(112)
5.5.3 假设分析	(113)
5.5.4 相关分析	(114)
5.5.5 回归分析	(115)
5.6 综合实例	(117)
5.7 本章小结	(121)
5.8 习题	(121)

第6章 Origin 在试验数据处理中的应用 (123)

6.1 ORIGIN 的基础知识	(123)
6.1.1 Origin 工作环境	(124)
6.1.2 Origin 基本操作	(129)
6.2 ORIGIN 二维绘图	(132)
6.2.1 数据的导入	(132)
6.2.2 二维图形的绘制	(135)
6.2.3 二维图的定制	(150)
6.3 ORIGIN 三维绘图	(153)
6.3.1 矩阵窗口的转换	(153)
6.3.2 三维图的绘制	(154)
6.3.3 三维图的定制	(157)
6.4 ORIGIN 的数据分析	(158)
6.4.1 简单的数学运算	(158)
6.4.2 统计分析	(160)
6.4.3 快速傅里叶变换和滤波	(164)
6.4.4 峰值分析	(167)
6.5 ORIGIN 的函数拟合	(168)
6.5.1 函数拟合方法	(168)
6.5.2 非线性最小平方拟合	(170)
6.5.3 自定义拟合函数	(171)
6.6 在 ORIGIN 中使用 EXCEL	(172)
6.6.1 Origin 中打开和保存工作簿	(172)
6.6.2 使用 Excel 工作簿绘图	(173)
6.7 综合举例	(174)
6.8 本章小结	(175)
6.9 习题	(176)

第7章 MATLAB 在试验数据处理中的应用 (177)

7.1 MATLAB 的基础知识	(177)
7.1.1 命令窗口	(177)
7.1.2 命令历史	(177)
7.1.3 帮助浏览器	(179)
7.1.4 当前目录浏览器	(179)
7.1.5 工作空间浏览器	(179)
7.1.6 数组编辑器	(180)
7.1.7 编程小知识	(180)

7.2	MATLAB 的数值计算	(181)	8.1.1	SPSS 界面	(232)
7.2.1	基本运算符	(181)	8.1.2	变量定义	(233)
7.2.2	特殊变量、标准初等数学 函数和数据分析基本函数	(181)	8.1.3	数据输入	(236)
7.2.3	数理统计分析与方差分析 及计算机求解	(182)	8.1.4	数据编辑	(237)
7.2.4	关系运算与逻辑运算	(188)	8.1.5	数据保存	(238)
7.3	矩阵运算	(188)	8.1.6	数据读取	(239)
7.3.1	矩阵的生成和一般基本 运算	(188)	8.1.7	SPSS 参数设置	(239)
7.3.2	矩阵的一般操作	(191)	8.1.8	输出窗口	(239)
7.3.3	特殊矩阵函数	(193)	8.2	SPSS 数据预处理	(240)
7.4	MATLAB 基本绘图	(194)	8.2.1	数据排序	(240)
7.4.1	基本绘图操作和二维绘图	(194)	8.2.2	行列转置	(241)
7.4.2	三维绘图	(205)	8.2.3	数据选取	(241)
7.4.3	图形窗口的编辑	(210)	8.2.4	数据分类汇总	(243)
7.5	回归分析和数据插值	(212)	8.2.5	变量加权	(245)
7.5.1	多项式拟合	(212)	8.2.6	数据文件分组	(245)
7.5.2	最小二乘曲线拟合	(213)	8.2.7	数据文件合并	(245)
7.5.3	数据插值	(214)	8.2.8	数据次序的确定	(247)
7.6	快速傅里叶变换的 MATLAB 实现	(217)	8.2.9	变量的变换与计算	(248)
7.6.1	数字信号的 FFT 变换和 逆变换	(217)	8.3	SPSS 的统计分析	(249)
7.6.2	带噪声信号的频率识别	(219)	8.3.1	Frequencies 分析	(249)
7.6.3	卷积运算中的 FFT 的应用	(220)	8.3.2	Descriptives 分析	(254)
7.6.4	功率谱估计	(221)	8.3.3	Explore 分析	(255)
7.7	其他常用信号处理、分析的 MATLAB 实现	(222)	8.3.4	Crosstabs 分析	(259)
7.7.1	滤波	(222)	8.3.5	Ratio 分析	(262)
7.7.2	相关函数	(224)	8.4	SPSS 的统计图	(264)
7.7.3	相干函数	(225)	8.4.1	条形图	(264)
7.7.4	时谱分析	(227)	8.4.2	线图	(265)
7.8	综合练习	(228)	8.4.3	面积图	(266)
7.9	本章小结	(230)	8.4.4	圆图	(267)
7.10	习题	(230)	8.4.5	高低图	(267)
8	SPSS 在试验数据处理 中的应用	(232)	8.4.6	箱图	(269)
8.1	SPSS 13.0 FOR WINDOWS 软件介绍	(232)	8.4.7	散点图	(270)
8.1.1	SPSS 界面	(232)	8.4.8	质量控制图	(271)
8.1.2	变量定义	(233)	8.4.9	直方图	(272)
8.1.3	数据输入	(236)	8.4.10	正态概率图	(272)
8.1.4	数据编辑	(237)	8.4.11	序列图	(273)
8.1.5	数据保存	(238)	8.5	SPSS 的参数检验	(273)
8.1.6	数据读取	(239)	8.5.1	均值	(273)
8.1.7	SPSS 参数设置	(239)	8.5.2	单样本 t 检验	(273)
8.1.8	输出窗口	(239)	8.5.3	两独立样本 t 检验	(275)

8.5.4 两配对样本 t 检验	(276)
8.6 SPSS 的方差分析	(277)
8.6.1 单因素方差分析	(277)
8.6.2 多因素方差分析	(280)
8.6.3 协方差分析	(283)
8.7 SPSS 的回归分析	(283)
8.7.1 一元线性回归分析	(283)
8.7.2 多元线性回归分析	(286)
8.7.3 非线性回归分析	(286)
8.7.4 曲线估计	(287)
8.7.5 加权回归分析	(289)
8.8 SPSS 的非参数检验	(290)
8.8.1 单样本非参数检验	(290)
8.8.2 两独立样本非参数检验	(293)
8.8.3 两相关样本非参数检验	(294)
8.8.4 多独立样本非参数检验	(294)
8.8.5 多相关样本非参数检验	(295)
8.9 SPSS 的相关分析	(296)
8.9.1 二元变量相关分析	(296)

8.9.2 偏相关分析	(297)
8.9.3 距离相关分析	(298)
8.10 综合实例	(300)
8.11 本章小结	(304)
8.12 习题	(304)

第9章 软件应用与技巧 (306)

9.1 数据处理软件的优缺点	(306)
9.1.1 Microsoft Office Excel	(306)
9.1.2 Origin	(307)
9.1.3 MATLAB	(308)
9.1.4 SPSS for Windows	(309)
9.2 数据处理软件的选择与技巧	(309)
9.2.1 数据分析与处理的基本步骤	(309)
9.2.2 软件的选择与技巧	(310)
9.3 本章小结	(311)
9.4 习题	(311)

参考文献 (312)

第1章 试验数据表示及误差分析

知识点

- 试验数据的测量与表示
- 误差的来源、含义及表示方法
- 误差的传播定律与误差的分配
- 随机误差
- 系统误差

本章导读

测量是人们认识和改造客观世界的一种不可或缺的主要手段。在科学研究、试验分析及工程实践中经常需要对客观事物的某种特性给出一定量的数字表征,并用科学的表述尽量给出其“真值”。然而,测量出的数值不可能完全等于其真值,它只能无限地接近于这个真值。因此,测量值与真值之间的误差是必然存在的,也是不可避免的。为了减小这样的误差,降低误差的消极影响,提高科学试验的可信度及工程实践的准确度和安全性,就必须了解误差,认识误差的来源,从而找到减小误差的方法来提高测量值的精确度、正确度和精密密度。本章将主要介绍试验数据的测量、表示及误差分析的一些基本概念,为后续各章对试验数据处理和分析的深入研究奠定基础。

1.1 试验数据的测量与表示

在研究自然科学领域时,经常需要进行一些有关客观世界假设和证明的试验。在这些试验中,将不可避免地要对其中一些表征量进行测量。与此同时,还必须准确、有效、规范地记录下所测量到的数据结果,提高它们经过公式计算和变换后的精确度,以便更好地证明试验所提出的假设。

1.1.1 数据的测量

测量就是对某种量的值的定量确定,比如物理对象(或研究对象)的特性,或者描述可能再现物理状态所必需的参数的定量确定等。这也就是说,测量活动是可以重复进行的,并且每一次测量都是相互独立地来确定被测量的值。

根据测量精度、测量方法、测量条件及被测对象在测量过程中所处的状态,可以把测量的方法分为四种类型。当然,这只是对测量方法进行粗略分类,根据研究的需要还会有其他不同的分类方法。

1. 按照对测量结果精确度的要求分类

1) 工程测量

工程测量是一般工作中所进行的测量,测量结果只要求取得测量值就能够满足测量的要求,并不需要考虑测量结果误差的大小或测量值的可信程度。因此,对于进行这种测量的设

备或仪器,其敏感度和精度都会比较低,并且对进行测量的环境和条件没有什么特殊的要求和限制,给出的测量值通常也会比较稳定。所以,单次测量或多次测量所给出的测量结果将会是完全一样的,故这种测量不需要考虑测量误差问题。

在一般生产现场或一般工程试验中所进行的测量,多为工程测量。

2) 精密测量

凡是经过测量取得测量结果后,要求估计出测量结果的误差确切值的测量,即为精密测量。这种测量在误差理论的指导下,需要经过反复多次的重复测量,所使用的测量仪器和设备应该具有较高的精度和敏感度,并且在每次测量中都能够反映出测量误差的变化和存在。在测量完成后把所获取到的数据根据误差理论进行处理,计算出最佳的测量结果,并估计出表示测量误差的确切值。进行精密测量的条件要求比工程测量要严格得多,大多数精密测量是根据测量仪器的使用条件在实验室中进行的,所以它也被称为实验室测量。

在测量之前,首先应该明确对测量结果的精度要求,确定是属于哪一种测量。这对于提出测量方案、选用测量仪器和设备及对测量值进行处理等都是很重要的。采用精密测量得到的测量结果精度较高,因而所付出的代价也相对较大。而工程测量所得到的测量结果精度较低,所使用的测量设备通常都很简单,价格便宜,操作也比较简便,故付出的代价也比较小。所以,要根据实际需要,合理地确定采用哪种测量方法。

2. 按照取得测量结果的方法分类

1) 直接测量

把被测量和作为测量标准的量直接进行比较,或用预先按标准校对好的测量仪器对被测量进行测量,通过测量能直接得到被测量的数量大小,此种测量称为直接测量。直接测量可以用下面的一般公式来表示:

$$y = x$$

式中, y 为被测量, x 为测得值。

在工程测量中,对时间、长度、质量进行的测量和使用专用仪表对压力、温度、湿度进行的测量都是直接测量。

2) 间接测量

被测量不能用直接测量方法得到,而必须通过一个或多个直接测量值,利用一定的函数关系进行运算才能最终得出,这种测量称为间接测量。间接测量可以用下面一般公式来表示:

$$y = f(x_1, x_2, \dots, x_m)$$

式中, y 为被测量, $x_1, x_2, \dots, x_m$ 为各直接测量值。

间接测量在科学研究中用得最多,因为在大多数情况下,仅仅使用直接测量方法不能得到被测量值;或者能够测量得到被测量值,但测量过程非常复杂,不如采用间接测量方法更方便、更精确。例如,天文学方面各种参数的测量,原子内部结构参数的测量等。

3) 组合测量

被测量不能通过直接测量或间接测量得到,而必须通过直接测量的测得值或间接测量的测得值建立起联立方程组,通过求解联立方程组,才能得到最后的测量结果。这样的测量方法称为组合测量,它可以用下面的一般公式来表示:

$$F_1(y_1, y_2, y_3, \dots, y_m, x_{11}, x_{21}, x_{31}, \dots, x_{i1}, \dots, x_{n1}) = 0$$

$$F_2(y_1, y_2, y_3, \dots, y_m, x_{12}, x_{22}, x_{32}, \dots, x_{i2}, \dots, x_{n2}) = 0$$

.....

$$F_j(y_1, y_2, y_3, \dots, y_m, x_{1j}, x_{2j}, x_{3j}, \dots, x_{ij}, \dots, x_{nj}) = 0$$

.....

$$F_m(y_1, y_2, y_3, \dots, y_m, x_{1m}, x_{2m}, x_{3m}, \dots, x_{im}, \dots, x_{nm}) = 0$$

式中: $F_1, F_2, \dots, F_m$ 为组合测量中 y 与 x 构成的已知函数关系; $y_1, y_2, y_3, \dots, y_m$ 为组合测量中的 m 个被测量值; x_{ij} 为组合测量中第 j 个直接(或间接)被测量的第 i 次测得值, 其中 $i=1, 2, 3, \dots, n, j=1, 2, 3, \dots, m$ 。

上述联立方程可通过改变测量条件的方法列出, 也可以通过被测量 y 以不同组合参加的测量过程列出, 使方程的数目 m 与被测量 y 的个数相等, 这样就可以解此联立方程, 从而求得各被测量的值。

3. 按照测量条件分类

1) 等精度测量

对某一固定被测量进行重复测量, 所取得的测量数据, 可以认为是在相同的测量精度条件下得到的, 这种测量称为等精度测量。对一固定被测量做等精度测量, 所得测量数据允许有一定范围误差, 但对偏大或偏小的数值, 不能判定哪种数值更加接近被测量的真实值, 只能同等对待, 即对取得数据的信赖程度是相同的, 这是判定是否为等精度测量的重要依据。

2) 不等精度测量

对被测量进行测量所得到的数据, 其精确度可判定是不等的, 这种测量称为不等精度测量。不等精度测量造成精度不等的原因, 可能是由于条件的改变, 所用测量设备的不同, 也可能是由于数据的来源不同等。对不等精度测量的数据应当采取特殊的处理方法。

4. 按照被测对象在测量过程中所处的状态分类

1) 静态测量

被测量在测量过程中可以认为是固定不变的, 对这种被测量进行的测量称为静态测量。实际上, 静态测量不需要考虑时间因素对测量的影响, 可以把被测量或测量误差作为随机变量来进行研究。

2) 动态测量

被测量处在随时间不断变化的状态中, 对这种被测量进行的测量称为动态测量。进行这种测量和处理这种测量得到的数据, 就需要考虑时间因素对测量的影响, 即把测得值或测量误差作为随机过程来研究。

上面介绍的四种测量的分类方法, 只涉及在以后研究测量数据处理和误差问题首先要弄清楚的基本概念, 以及各种专业进行测量时所具有的共性问题。

1.1.2 数据的表示

1. 近似值和有效数

含有误差的数值均为近似值, 它包括:

- 测量值;
- 数学常量的有限位小数表示, 如 $\pi=3.1416$, $e=2.718$;

- 物理量值, 如传热系数, 汽化潜热, 空气密度等;
- 运算结果, 数字运算过程中舍入误差的积累。

一般情况下, 近似值与真值之差的绝对值总是小于近似值最末一位数的一半, 这就是近似值的误差限。由此, 可以定义若误差限 Δ_n 不超过数值 L 中某一位的半个单位时, 则从该位起算, 到数值 L 的第一位非零数字为止, 共有 n 位, 则称 L 为具有 n 位有效数字的有效数, n 称为有效位数。例如:

3.1416	五位有效数字	2.718	四位有效数字
0.09160	四位有效数字	6.150×10^{10}	四位有效数字

2. 数字的舍入规则

若近似数的位数较多, 则应该在适当确定有效位数后, 舍去多余的尾数, 其规则是: 以保留的最末位为单位, 后面的数大于 0.5 则进 1, 小于 0.5 则舍去; 等于 0.5 时, 则以使末位凑成偶数为准, 即当末位为奇数时进 1, 为偶数时舍去, 这也称为“五留双”。下面的各数为取三位有效数时的处理示例:

3.1416	3.14 (舍)	2.718	2.72 (入)
4.4350	4.44 (入)	6.1250	6.12 (舍)

这种以规定位数的奇偶性作为舍入判据的方法, 在大量的数字运算中, 可使舍入的概率相等, 舍入误差的影响达到最小。

3. 试验结果的数字表达

测量结果有效位数的确定, 应依据相应的误差值。若已知测量误差为 0.008mL, 而测量结果为 0.796mL, 这时取位就是适宜的; 若给出 0.79633mL, 则毫无意义; 若只给出 0.80mL, 则损失了精度。因为, 通常将测量结果最末位的数值认为是含有误差的欠准数值。

为了避免在不恰当的数字运算时损失试验精度, 一般的做法如下:

- 为减小舍入误差的影响, 在运算过程中可多保留一位有效数字;
- 在进行加减运算时, 和或差在小数点后的取位, 应以运算值当中小数点后位数最少的为准。例如, $1.367+3.79531+6.8527$, 其中 1.367 小数点后的位数最少, 故计算结果只宜写成 12.015。
- 积或商的有效位数, 应当与运算值中有效位数最少的相同。例如, $1.457 \times 0.0974 \times 85.319$, 其中 0.0974 为三位有效数, 故计算结果只宜写成 12.1。

一个试验结果, 最后的表达应该包括结果的最佳值、误差区间, 以及该误差区间的置信概率, 如 $Y=7.42 \pm 0.02$ ($P=98\%$)。

1.2 误差分析

在科学研究的工作中, 需要经常探讨和改进试验的物理模型及测量方法, 只有这样才能得到更为准确的试验结果; 一个没有考虑误差大小的试验结果, 一个不知准确度如何的测量值, 是不能够说明问题的, 有时候甚至是毫无意义的。对测量误差进行研究, 其目的就是希望确切地了解测量误差的大小范围, 再把测量误差控制在能够满足需要的程度, 并以误差理论为依据对测量的结果做出科学的、合理的评定。

1.2.1 误差的含义

根据通常习惯的概念和理解, 误差应该存在于各种测量的测得值之中, 即:

$$\text{误差} + \text{真实值} = \text{测得值}$$

作为表征测量误差大小的指标, 通常采用三种形式来表示: 绝对误差、相对误差和百分误差。下面将分别介绍这三种误差。

1. 绝对误差

利用测量器具或仪器对被测对象进行测量, 所得到的测量数据称为测量值或示值。测量值与被测量客观真实值之间的差值称为绝对误差 (简称误差), 即:

$$\text{绝对误差} = \text{测量值} - \text{真值}$$

误差具有与被测量相同的量纲, 且可以为正值或负值, 误差可以反映测量值对真实值偏离的大小。

在实际工作中, 这种理论上的公式是很难实现的, 因为被测量的真实值是无法得到的, 所以也就无法从测得值中把测量误差分离出来。然而, 正是因为不知道被测量的真实值, 才需要试验进行测量, 并希望通过误差值来评价测得值的可靠性, 即测得值与被测量真值的近似程度。为了解决在实际工作中绝对误差的计算问题, 可根据理论和实际需要, 用测量应得到的数值, 即应得值 (目的值) 来代替被测量的真实值进行绝对误差的计算。

2. 相对误差

由于绝对误差只能表示测量误差的绝对大小, 当测量不同数量级的被测量时, 利用绝对误差就不能确切地表示出所有测量的精确程度。例如, 对两个数量级不等的量分别进行测量, 虽然得到的绝对误差相同, 但测量精度是不一样的。其中数量级大的, 测量精度就较高; 数量级小的, 测量精度则较低。因此, 如果对测量的精度或对测量仪器所具有的精度进行比较时, 必须要用相对误差作为相互比较的技术指标。

误差与真实值之比称为相对误差, 即:

$$\text{相对误差} = \text{误差} / \text{真值}$$

相对误差是一个无量纲数, 并且有正负之分, 相对误差可以用来反映误差对真实值的影响程度。但由于真实值难以获得, 而且测量值与真实值又总是接近的, 所以在工程实际中, 相对误差可用误差与测量值之比来近似表达, 即:

$$\text{相对误差} \approx \text{误差} / \text{测量值}$$

例 1-1 常量滴定管 (25mL) 读数时可估读到 $\pm 0.01\text{mL}$, 若要求滴定的相对误差小于 0.1%, 在滴定时, 耗用体积应控制为多少?

解: $\because 2 \times 0.01 / V < 0.1\%$

$$\therefore V > 20\text{mL}$$

答: 耗用体积应控制在大于 20mL, 小于等于 25mL 范围内。

例 1-2 分析天平称准确至 $\pm 0.0001\text{g}$, 要使称量误差不大于 0.1%, 至少应称取多少试样?

解: $\because 2 \times 0.0001 / m_s \leq 0.1\%$

$$\therefore m_s \geq 0.2\text{g}$$

答：至少应称取 0.2g 试样。

3. 百分误差

百分误差就是相对误差乘以 100。

如果用 μ 表示某个量的真值，而用 $\Delta\mu$ 表示测量或计算得到的近似值与真值 μ 的绝对误差，那么：

$$\Delta\mu/\mu = \text{近似值的相对误差}$$

$$100 \times \Delta\mu/\mu = \text{近似值的百分误差}$$

如果一个数准确到 n 位有效数字，则它的绝对误差不会大于第 n 位数字的一个单位的一半。例如，9.15616 这个数准确到了第六位数字，那么它的绝对误差将不大于 $0.00001 \times 0.5 = 0.000005$ 。

1.2.2 误差的类型

为了对误差能有一个系统和概括的了解，现在介绍一种被普遍采用的误差分类方法，即根据误差所具有的性质和特点来进行分类。这样就可以把误差分为随机误差、系统误差和粗大误差（即错误）三大类。一般来说，试验误差都是这三类误差的某种相加函数，而粗大误差则可以立即消除。

1. 随机误差

在测量的过程中，必然会受到一些随机因素的影响，从而造成具有随机性质的测量误差，这种误差被称为随机误差。随机误差的大小和方向（误差的正负）是无法预测、无确定规律的。即使在尽可能相同的条件下，对同一固定量进行重复的测量，在极力排除和改正一切明显的、有规律的偏差之后，每次测量所得到的测得值，总是会在一定的范围内随机性地波动变化，这些就是由随机误差造成的。

在测量的统计分析中，随机误差（也称统计误差）是经常会遇到的。因为随机误差产生于众多因素的微小波动，这些影响既难发现又难排除，是伴随整个测量过程而不能消除的误差。随机误差具有随机变量的一切特征，所以，需要采用数理统计的方法，来研究随机误差的统计特征，以判断它对测量结果的影响。

随机误差对个体来说，也就是重复测量中的任一次测量所产生的误差，它是没有规律的、不能控制的，同时也是无法消除的。但对于总体，即经过多次测量所得到的测得值而言，随机误差是服从一定统计规律的，所以对随机误差可以采用概率统计的方法进行处理。

2. 系统误差

当测量条件一定时，误差的大小和方向是恒定的；而当测量条件变化时，误差则按某一确定的规律随之发生变化，这类误差称为系统误差，也称确定误差，简称系差。所谓确定的规律，指的是误差的变化可以用函数式或曲线图形来进行描述。系统误差的产生，一般是由一个或几个因素而引起的，因此也是有规律可循的。根据已知的变化规律，又可以将系统误差进一步划分为：对于固定不变的系统误差，称为恒定系统误差，简称恒差；对于按照一定的已知规律变化的系统误差，称为变值系统误差，简称变差。因此，对系统误差的处理，就是发现它，并且想方设法地消除它。

上述的确定误差，从其可知性方面来说是比较典型的系统误差，统称为已定系统误差；

和它相对应的,如果系统误差带有一定的随机性质,则称此系统误差为未定系统误差。

由于系统误差一般来说都并不遵守统计处理规则,因此这种误差降低了平均值的可靠性,而降低的程度只能被估计。

下面就一些引起系统误差的问题进行说明:

- 在仪器仪表刻度中的判断误差。例如,观测者在进行多次观测时,估计最小刻度的分数偏高,那么得到结果的系统误差将趋于偏高的值。仪器仪表不正确的校正(或不合理的假定)或仪器仪表运动部分的磨损也可能引起系统误差。
- 试验条件的改变也会引起系统误差。例如,仪器是在某一大气压下进行校正的,而在另一大气压下使用,这时也将导致系统误差。

3. 粗大误差

明显地歪曲测得值的误差称为粗大误差,亦即错误。这类误差通常是由于人员疏忽而造成的操作错误、读数错误、运算错误及记录错误等,既然是由于疏忽或失误所造成的,因此也可以叫做疏失误差。比如,某数据要求精确度达到六位有效数字,而使用的仪表仪器仅可以保证五位有效数字的准确性;抄记读数的时候,将0.916记成0.619等。当然,有时尽管在测量过程中很细心,操作也很认真,但同样会因为随机性的缘故而出现较大的误差。这种情况从概率论的观点来分析也是完全可能的,即相对数值较大的误差出现的概率尽管微乎其微,但不等于绝对不会出现。

确切地说,所谓“粗大误差”其实已不再属于误差的范畴,所以对于含有粗大误差的测量值一般又称为坏值,在对试验结果进行数据处理之前,必须先行剔除这些坏值。但对原因不明的可疑值,在处理时应采取慎重、小心的态度,尽管它对测量的影响较大,但在不能判定为不可信的情况下,绝不能主观臆断轻易地把它剔除,而是应当根据一定的准则来加以判断,最终决定是否将该数据剔除。

在测量时应该对必须修正和舍弃的一个或多个错误做出一些估计。错误包括下列几种情况:

- 仪器的读数错误;
- 数据的抄写错误;
- 单位的混乱;
- 运算错误;
- 小数点的位置错误。

通常而言,对于所取得的一组测量结果,首先需要判断其中是否有坏值并进行剔除,之后再根据系统误差和随机误差的不同特点分别进行处理,最终才能获得较好的测量结果。

1.2.3 误差的表示方法

作为表征测量误差大小的指标,通常会采用两种形式来表示。用无量纲的相对值来表示的误差称为相对误差;与前者相对应的,用有量纲的数值表示的误差称为绝对误差。下面将分别介绍这两种误差的几种表示方法。

1. 绝对误差的表示方法

由于在误差计算中所用应得值的不同,得到的绝对误差也就有不同的名称,常见的绝对误差有以下几种表示方法。

1) 真误差

把被测量的真实值作为应得值, 所获得的测得值与真实值之间的差值即为真误差, 简称真差。

$$\delta = x - \mu$$

因为通过测量是得不到被测量的真实值的, 所以真差 δ 也是无法得到的, 因此真差经常只能在理论推导中应用。但在一些特殊情况下, 如在理论上可以求到真值 A_0 或把指定值 A_s 、传递值 A 看成是被测量的真实值, 这样做的话则可以认为真差是能够求得的。

2) 剩余误差

把有限 n 次测量后所得到测得值的算术平均值作为应得值而求出的绝对误差, 称为剩余误差, 简称残差。

$$v_i = x_i - x_0$$

式中, v_i 为第 i 次测得值残差; x_i 为第 i 次测量的测得值; x_0 为 n 次测得值的算术平均值。

因为剩余误差 v_i 是可以由测得值算出的, 所以它在误差计算中经常使用到。

3) 最大绝对误差

因为通过测量并不能得到真实值 μ , 从而也就无法求到真差。如果能想办法找出一个界限值 U , 且能做出判断:

$$U \geq |x - \mu|$$

亦即

$$U = \sup |\delta|$$

式中, $\sup$ 表示测得值 x 的真差 δ 的绝对值绝不超过 U , 则称 U 为最大绝对误差。因为在实际中很少会用到绝对误差 δ , 所以人们习惯上也把最大绝对误差 U 简称为最大误差。这样即使省略了“绝对”二字, 也不会造成理解上的混乱。

根据实际条件规定的最大绝对误差 (即误差的界限值), 绝不是凭空想出或任意决定出来的, 而是要凭一定的依据。例如, 在用数字常项 π 进行计算的时候, 如果取 3.14 进行计算, 则由 π 值引起的绝对误差为:

$$|\delta| = |\pi - 3.14| = 0.00159 \cdots < 0.0016 = U$$

即最大绝对误差为:

$$U = 0.0016$$

4) 标准误差 (标准偏差)

标准误差的定义是: 对任一固定量进行无穷次的测量, 各次测量真差平方的算术平均值再开方所得到的数值。根据其数学运算关系, 又可称为方均根误差。

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \Big|_{n \rightarrow \infty} = \sqrt{\frac{1}{n} \sum_{i=1}^n \delta_i^2} \Big|_{n \rightarrow \infty}$$

式中, μ 为真值, x_i 为第 i 次测量得到的测得值, δ_i 为第 i 次测得值与真实值之间的真差。

但在测量实践中观测次数不可能无限多, 因此实际应用中, 以有限次观测次数 n 计算出标准差的估值定义为中误差 m , 作为衡量精度的一种标准, 计算公式为:

$$m = \pm \hat{\sigma} = \pm \sqrt{\frac{1}{n} \sum_{i=1}^n \delta_i^2}$$

例 1-3 有甲、乙两组，各自用相同的条件观测了 6 个三角形的内角，得三角形的闭合差（即三角形内角和的真误差）分别如下。

甲：+3″ +1″ -2″ -1″ 0″ -3″

乙：+6″ -5″ +1″ -4″ -3″ +5″

试分析两组的观测精度。

解：用中误差公式计算得

$$m_{\text{甲}} = \pm \hat{\sigma} = \pm \sqrt{\frac{1}{6} \sum_{i=1}^6 \delta_{\text{甲}i}^2} = \pm 2.0''$$

$$m_{\text{乙}} = \pm \hat{\sigma} = \pm \sqrt{\frac{1}{6} \sum_{i=1}^6 \delta_{\text{乙}i}^2} = \pm 4.3''$$

从上述两组结果中可以看出，甲组的中误差较小，所以观测精度高于乙组。而直接从观测误差的分布来看，也可看出甲组观测的误差比较集中，离散度较小，因而观测精度高于乙组。所以在测量工作中，普遍采用中误差来评定测量结果的精度。

5) 或然误差

或然误差又称概差。它是根据误差出现的概率来进行定义的。在一组测得值中如果不计较误差的正负号，则误差大于 ρ 的测得值与误差小于 ρ 的测得值将各占一半。假如考虑测量误差的正负号，或然误差 ρ 同样可以把带有正误差的测得值及带有负误差的测得值按测量误差的大小被 $+\rho$ 或 $-\rho$ 等分。

6) 极限误差

一般地，精密计量中对于服从正态分布的随机误差常用三倍标准误差来作为极限误差：

$$\delta_{\text{lim}} = 3\sigma$$

从理论上已证明得到，若测得值服从正态分布，则测得值的误差小于极限误差 δ_{lim} 的概率为 99.73%，即测量误差只有 3/1000 能超过极限误差 δ_{lim} ，在实际中就相当于不会出现超过极限误差 δ_{lim} 的误差。

最大误差 U 应当与极限误差 δ_{lim} 有所区别，因为求得 U 的方法并没有受到限制，而极限误差 δ_{lim} 是从概率论中引入的概念，它与标准误差具有某种固定的关系。另外，最大误差的定义符号 $\sup$ 是绝对不会超过的意思，而极限误差 δ_{lim} 的 3σ 定义是说测量误差还有可能超过极限误差 δ_{lim} ，只是这样的概率非常小而已。

目前，也有取两倍标准误差 (2σ) 来作为极限误差的，只是它的可靠性比 3σ 稍微差一些。然而究竟是取 3σ 还是 2σ 作为极限误差，这要根据实际的需要来最终确定，现在还没有被一致公认地明确规定。不过一般多采用 3σ 来作为极限误差。

上面一共介绍了六种表示绝对误差的方法。根据它们在理论上的定义和计算公式可以看出，最后得到的表示绝对误差的数值，有的可以得到明确的正、负值（如 δ 和 v_i ），而有的却总是正值（如 U 、 σ 、 ρ 和 δ_{lim} ）。后面的四种表示方法，尽管计数值永远都为正值，但实际使用这些表示方向不定的随机误差时，前面应该冠以 ($\pm$) 号，否则就会被误认为误差的方向是