



普通高等教育“十三五”规划教材

PUTONG GAODENG JIAOYU “13·5” GUIHUA JIAOCAI

实验设计与数据处理 ——理论与实战

徐文峰 廖晓玲 主编



冶金工业出版社
www.cnmip.com.cn

普通高等教育“十三五”规划教材

实验设计与数据处理

——理论与实战

徐文峰 廖晓玲 主编
徐紫宸 杨梅 黄秋红 副主编

北京
冶金工业出版社
2019

内 容 提 要

本书共分5章, 主要内容包括: 实验分析中的误差及数据处理, 基本概念和名词术语, 并配有相应的实验训练项目; 实验设计基本概念、基本原则、方法和相应的实验项目; 正交实验理论和其科学性的原理特点, 配有简单的正交实验的验证性实验; 正交实验设计及具体影响因素, 正确设计正交实验的步骤, 同时, 配有正交实验的设计训练实验; Excel 在数据统计分析中的应用, 运用 Excel 功能的方法, 为实验设计与数据处理提供帮助。

本书可作为电子信息、生物医药、材料、机械等专业的教学用书, 也可供从事相关专业的工程技术人员和研究员参考。

图书在版编目(CIP)数据

实验设计与数据处理: 理论与实战 / 徐文峰, 廖晓玲
主编. —北京: 冶金工业出版社, 2019. 2
普通高等教育“十三五”规划教材
ISBN 978-7-5024-7981-7

I. ①实… II. ①徐… ②廖… III. ①试验设计—高等学校—教材 ②实验数据—数据处理—高等学校—教材
IV. ①O212. 6

中国版本图书馆 CIP 数据核字(2019)第 010150 号

出 版 人 谭学余
地 址 北京市东城区嵩祝院北巷 39 号 邮编 100009 电话 (010)64027926
网 址 www.cnmp. com. cn 电子信箱 yjchs@cnmp. com. cn
责任编辑 郭冬艳 美术编辑 吕欣童 版式设计 禹蕊
责任校对 郑娟 责任印制 牛晓波
ISBN 978-7-5024-7981-7

冶金工业出版社出版发行; 各地新华书店经销; 三河市双峰印刷装订有限公司印刷
2019年2月第1版, 2019年2月第1次印刷

787mm×1092mm 1/16; 11.75印张; 282千字; 177页

35.00元

冶金工业出版社 投稿电话 (010)64027932 投稿信箱 tougao@cnmp. com. cn
冶金工业出版社营销中心 电话 (010)64044283 传真 (010)64027893
冶金工业出版社天猫旗舰店 yjgycbs. tmall. com

(本书如有印装质量问题, 本社营销中心负责退换)

前 言

现今世界迎来了大数据时代，人们的日常工作、生活均与数据息息相关。理工科学生在走向社会后，不管是从事新产品开发、新工艺优化、生产过程质量管理，还是从事相关的技术服务、销售推广等工作，都会面临各种影响工作效果的因素以及各种大量相关的数据，怎样在众多纷繁复杂的因素或数据中，找到最主要的影响因素，为工作提供正确选择的参考依据，显得越来越重要。本书正是关注这个焦点，介绍实验设计与数据处理的理论与实践，不仅提供了需要学习解决问题的科学方法，同时介绍了培养解决问题和提高能力训练的实验。

实验设计方法是数理统计学的应用方法之一。大多数数理统计方法主要用于分析已经得到的数据，而实验设计方法却是决定数据收集方法科学的一个有效手段。在正确获得有效数据的同时，需要掌握数据的处理方法，两者不可分割。作者根据多年的教学实践经验和体会，在参考国内外相关资料及关注近年来教学动态的基础上，结合培养现代化创新应用型复合人才的需要，编写了本书。本书基于战略性新兴产业对人才的要求，开展“理论-项目”式实践教学实战训练，将现代科学数据收集方法的实验设计和数据处理理论与实验实践相结合，不仅使学生掌握相关专业理论知识，又培养其解决问题的实践及创新能力，为培养与时俱进的创新型、应用型复合人才，提供帮助。

本书可作为高等院校材料类、机械类、电子信息类、生物医药类等专业的大学教材，亦可供相关专业的工程技术人员参考。

本书由徐文峰教授、廖晓玲教授主编，重庆大学徐紫宸博士、重庆科技学院杨梅、黄秋红硕士担任副主编。

本书内容所涉及的研究，得到国家自然科学基金重点项目（11532004）、重庆市科委自然科学基金重点项目（CSTC2015JCYJBX0003）、重庆市科委自然

科学基金项目 (cstc2018jcyjAX0286)、重庆科技学院学科带头人人才项目、纳微智能材料重庆高校创新团队 (CXTDX201601032)、材料科学与工程重庆市高校重点学科建设项目, 以及纳微生物医学检测重庆市工程实验室的资助与大力支持, 在此表示衷心的感谢!

由于编者水平所限, 书中疏漏和不妥之处, 敬请广大读者批评指正。

编 者

2018年11月

目 录

1	实验分析中的误差及数据处理	1
1.1	平均数	1
1.1.1	算术平均数	1
1.1.2	几何平均数	2
1.1.3	调和平均数	2
1.2	分析中的误差	3
1.2.1	误差	3
1.2.2	误差的分类及减免误差的方法	6
1.3	标准差	7
1.3.1	标准差的意义	7
1.3.2	标准差的计算	8
1.3.3	标准差的运算性质	9
1.4	有效数字及运算规则	9
1.4.1	有效数字	9
1.4.2	运算规则	10
1.5	实验数据处理	11
1.5.1	随机误差的正态分布	11
1.5.2	定量分析数据的评价	15
1.5.3	数据相关性的判定	19
1.5.4	提高分析结果准确度方法	21
实验 1	A4 纸重量长度测量实验	22
实验 2	标准砝码测量实验	23
实验 3	酸碱滴定数据的处理应用实验	24
实验 4	氯化钠的提纯实验	28
实验 5	天然水总硬度的测定	29
实验 6	$\text{Fe}(\text{OH})_3$ 胶体电动势的测定	30
2	实验设计基本概念	33
2.1	实验设计方法的概念	33
2.2	实验设计三原则	34
2.2.1	重复	34
2.2.2	随机化	36

2.2.3 局部管理	37
2.3 全面实验法	38
2.4 简单比较法	39
2.5 单因素优化实验设计	40
2.5.1 单因素优化实验设计的适用场合	40
2.5.2 均分法	41
2.5.3 对分法	42
2.5.4 黄金分割法	43
2.5.5 分数法	45
2.5.6 分批实验法	48
实验7 荧光金纳米簇的合成	48
实验8 合成可降解生物膜的单因素优化工艺实验	49
实验9 银纳米簇水凝胶的制备	50
实验10 低碳钢强度及应变硬化指数测定	50
3 正交实验理论	52
3.1 基本概念	52
3.2 正交表的记号及含义	52
3.2.1 正交表的记号及含义	52
3.2.2 正交实验结果的分析	53
3.3 正交表的特点	53
3.3.1 正交表的特点	53
3.3.2 正交实验设计的基本步骤	53
3.4 多因素优化实验设计	54
3.4.1 多因素优化实验概述	54
3.4.2 因素轮换法	56
3.4.3 随机实验	58
3.4.4 2^2 设计	59
3.4.5 2^3 设计	61
实验11 辣椒红色素的提取	64
实验12 胞苷介导金-银复合纳米簇的合成	65
实验13 多因素实验法制备纳米级有序介孔氧化硅空心球	66
实验14 用正交法测定几种因素对蔗糖酶活力的影响	67
实验15 镁合金阳极氧化成膜实验	69
实验16 玻璃软化温度的测定	70
实验17 陶瓷坯料配方实验	71
实验18 铁电陶瓷材料制备	73
4 正交实验设计	75
4.1 正交实验设计概述	75

4.2 正交表及其基本性质	75
4.2.1 正交表的构造	75
4.2.2 正交表的类型及特点	76
4.2.3 正交表的基本性质	77
4.3 正交实验设计的基本程序	78
4.3.1 正交实验方案设计	78
4.3.2 实验结果分析	82
4.4 实验方案设计	82
4.4.1 2^n 型正交表实验	82
4.4.2 3^n 型正交表实验	90
4.5 实验结果分析	93
4.5.1 实验设计结果的极差分析	93
4.5.2 正交实验设计结果的方差分析	105
实验 19 靶向血栓纳米球的制备实验	116
实验 20 高锰酸钾法测定蛋壳中 CaO 的含量	117
实验 21 烟碱的提取与定性分析	118
实验 22 从茶叶中提取咖啡因的工艺优化研究	119
实验 23 阿司匹林中乙酰水杨酸含量测定	120
实验 24 分光光度法测定铁强化酱油中铁含量的正交实验	121
实验 25 紫外可见分光光度法绘制 Fe 标准曲线	123
实验 26 正交实验最优参数的抗菌复合膜的抗菌效果验证实验	124
实验 27 锌酸盐镀锌实验	126
实验 28 固相反应法制备功能陶瓷	127
实验 29 功能陶瓷的介电性能测试	128
实验 30 玻璃的熔制实验	130
实验 31 溶胶-凝胶法制备铁电粉体	131
实验 32 铁电薄膜材料制备	133
5 Excel 在数据统计分析中的应用	135
5.1 样本统计量计算	135
5.1.1 常用统计量	135
5.1.2 统计量计算	135
5.2 统计假设检验	139
5.2.1 成对数据资料的假设检验	140
5.2.2 双样本假设检验	143
5.2.3 分析工具库的建立	147
5.3 方差分析	149
5.3.1 单因素方差分析	149
5.3.2 双因素方差分析	152

5.4 多元线性回归	160
5.4.1 相关性分析	160
5.4.2 建立回归模型	161
实验 33 用 Excel 对酸碱滴定数据处理 1	164
实验 34 用 Excel 对酸碱滴定数据处理 2	170
附表	174
附表 1	174
附表 2	174
附表 3	174
附表 4	174
附表 5	174
附表 6	174
附表 7	174
附表 8	174
附表 9	174
附表 10	174
附表 11	174
附表 12	174
附表 13	174
附表 14	174
附表 15	174
附表 16	174
附表 17	174
附表 18	174
附表 19	174
附表 20	174
附表 21	174
附表 22	174
附表 23	174
附表 24	174
附表 25	174
附表 26	174
附表 27	174
附表 28	174
附表 29	174
附表 30	174
附表 31	174
附表 32	174
附表 33	174
附表 34	174
附表 35	174
附表 36	174
附表 37	174
附表 38	174
附表 39	174
附表 40	174
附表 41	174
附表 42	174
附表 43	174
附表 44	174
附表 45	174
附表 46	174
附表 47	174
附表 48	174
附表 49	174
附表 50	174
附表 51	174
附表 52	174
附表 53	174
附表 54	174
附表 55	174
附表 56	174
附表 57	174
附表 58	174
附表 59	174
附表 60	174
附表 61	174
附表 62	174
附表 63	174
附表 64	174
附表 65	174
附表 66	174
附表 67	174
附表 68	174
附表 69	174
附表 70	174
附表 71	174
附表 72	174
附表 73	174
附表 74	174
附表 75	174
附表 76	174
附表 77	174
附表 78	174
附表 79	174
附表 80	174
附表 81	174
附表 82	174
附表 83	174
附表 84	174
附表 85	174
附表 86	174
附表 87	174
附表 88	174
附表 89	174
附表 90	174
附表 91	174
附表 92	174
附表 93	174
附表 94	174
附表 95	174
附表 96	174
附表 97	174
附表 98	174
附表 99	174
附表 100	174

1

实验分析中的误差及数据处理

1.1 平均数

平均数 (Average) 是统计学中最常用的特征数, 用来描述资料的集中性, 即指资料中数据集中较多的中心位置。平均数可以分为算术平均数、几何平均数、调和平均数等。其中最常用的是算术平均数, 简称平均数。

1.1.1 算术平均数

将观测值 (Observed Value) 的总和除以观测值个数所得的商称为算术平均数 (Arithmetic Mean), 记为 \bar{x} 。算术平均数可根据样本大小及分组情况而采用直接法或加权法计算。

1.1.1.1 直接法

直接法主要用于样本含量 $n \leq 30$, 且未经分组资料平均数的计算。

设 \bar{x} 代表 x_1, x_2, \dots, x_n , 等 n 个变数的算术平均数, 则有如下关系式:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \sum_{i=1}^n \frac{x_i}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1-1)$$

式中, \sum 为加和符号 (summation); $\sum_{i=1}^n x_i$ 为 $x_1 + x_2 + \dots + x_n$ 的总和。在计算意义明确时,

$\sum_{i=1}^n x_i$ 可以简写成 $\sum x$ 。若欲求 2, 4, 6, 8, 15 的算术平均数, 其计算方法如下:

$$\bar{x} = \frac{2+4+6+8+15}{5} = 7$$

[例 1-1] 对某乳品厂生产的 10 袋小包装产品净重进行测定, 结果为 50.0g、52.0g、53.5g、56.0g、58.5g、60.0g、48.0g、51.0g、50.5g、49.0g, 求其平均数。

由于

$$\begin{aligned} \sum x &= 50.0 + 52.0 + 53.5 + 56.0 + 58.5 + 60.0 + 48.0 + 51.0 + 50.5 + 49.0 = 528.5 \\ n &= 10 \end{aligned}$$

所以, 算术平均数

$$\bar{x} = \frac{528.5}{10} = 52.85\text{g}$$

小包装产品的净重平均为 52.85g。

1.1.1.2 加权法

对于样本含量 $n \geq 30$ 且已分组的资料, 可在次数分布表的基础上采用加权法计算平均数, 计算公式为

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_k x_k}{f_1 + f_2 + \cdots + f_k} = \frac{\sum_{i=1}^k (f_i x_i)}{\sum_{i=1}^k f_i} = \frac{\sum (fx)}{\sum f} \quad (1-2)$$

式中, x_i 为第 i 组的组中值; f_i 为第 i 组的次数; k 为分组数。

由式 (1-2) 计算的平均数称为加权平均数 (weighted mean)。 f_i 是变量 x_i 所具有的“权”, 即变量 x_i 所占的比重。因为

$$\sum_{i=1}^k f_i = f_1 + f_2 + \cdots + f_k = n$$

故

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k (f_i x_i) = \frac{1}{n} \sum (fx) \quad (1-3)$$

1.1.2 几何平均数

当遇到计算平均增长率时, 常常以几何平均数表示其平均值。设 G 为 x_1, x_2, \dots, x_n , 这 n 个数据的几何平均数, 则有

$$G = (x_1 \cdot x_2 \cdot x_3 \cdots x_n)^{\frac{1}{n}} \quad (1-4)$$

对式取对数:

$$\lg G = \frac{\lg x_1 + \lg x_2 + \cdots + \lg x_n}{n} = \frac{\sum \lg x}{n} \quad (1-5)$$

可见, 几何平均数是变量对数的算术平均数的反对数值。

如果是一个分组数列时, 则几何平均数可用下列公式计算:

$$G = (x_1^{f_1} \cdot x_2^{f_2} \cdot x_3^{f_3} \cdots x_n^{f_n})^{\frac{1}{n}} \quad (1-6)$$

两边取对数, 则

$$\lg G = \frac{1}{n} \sum_{i=1}^k (f_i \lg x_i) \quad (1-7)$$

式中, $n = \sum_{i=1}^k f_i$ 。

由式 (1-6) 计算的几何平均数称为加权几何平均数。

1.1.3 调和平均数

计算平均速率时需用调和平均数, 用 H 表示。

通常, 先计算倒数平均数 $\frac{1}{H}$, 即

$$\frac{1}{H} = \frac{1}{n} \sum \frac{1}{x}$$

那么

$$H = \frac{n}{\sum \frac{1}{x}}$$

可见, 调和平均数是变量倒数的算术平均数的倒数。

1.2 分析中的误差

1.2.1 误差

1.2.1.1 误差 (Error) 与准确度 (Accuracy)

A 误差

测定值 x_i 与真实值之差。真实值 (True Value): 在一定的时间和空间条件下, 被测量的物质的客观存在值, 它是可趋进而不可达到的哲学概念。真值是客观存在的, 它分为科学规定真值、标准真值、理论真值。统计学中常用总体均值 μ (Mean) 替代真实值。

误差的大小可用绝对误差 E (Absolute Error) 和相对误差 RE (Relative Error) 表示。

$$E = x_i - \mu \quad (1-8)$$

$$RE = \frac{x_i - \mu}{\mu} \quad (1-9)$$

相对误差表示误差占真值的百分率或千分率。

B 准确度

- (1) 测定平均值与真值接近的程度;
- (2) 准确度高低常用误差大小表示, 误差小, 准确度高。

[例 1-2] 分析天平称量两物体的质量各为 1.6380g 和 0.1637g, 假定两者的真实质量分别为 1.6381g 和 0.1638g, 则两者称量的绝对误差分别为:

$$(1.6380 - 1.6381)\text{g} = -0.0001\text{g}$$

$$(0.1637 - 0.1638)\text{g} = -0.0001\text{g}$$

C 结论

- (1) 绝对误差相等, 相对误差并不一定相同;
- (2) 同样的绝对误差, 被测定的量较大时, 相对误差就比较小, 测定的准确度也就比较高 (选分子量大的基准物质);
- (3) 用相对误差来表示各种情况下测定结果的准确度更为确切;
- (4) 绝对误差和相对误差都有正值和负值。正值表示分析结果偏高, 负值表示分析结果偏低;
- (5) 实际工作中, 真值实际上是无法获得的。

常用纯物质的理论值、国家标准局提供的标准参考物质证书上给出的数值或多次测定结果的平均值当作真值。

1.2.1.2 偏差 (Deviation) 与精密度 (Precision)

A 偏差

- (1) 偏差: 个别测定结果 x_i 与几次测定结果的平均值的差。

- (2) 绝对偏差 d_i : 测定结果与平均值之差;

$$d_i = x_i - \bar{x} \quad (1-10)$$

- (3) 相对偏差 d_r : 绝对偏差在平均值中所占的百分率或千分率

$$d_r = \frac{|x_i - \bar{x}|}{\bar{x}} \times 100\% \quad (1-11)$$

各偏差值的绝对值的平均值,称为单次测定的平均偏差,又称算术平均偏差 (Average Deviation):

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n |d_i| = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (1-12)$$

单次测定的相对平均偏差表示为:

$$\bar{d}_r = \frac{\bar{d}}{\bar{x}} \times 100\% \quad (1-13)$$

B 标准偏差 (Standard Deviation)

标准偏差又称均方根偏差、标准差,是测定次数趋于无限多时的样本总体标准偏差,用 σ 表示。即公式如下:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x - \mu)^2}{n}} \quad (1-14)$$

式中, μ 为样本总体的均值 (Mean),在校正了系统误差的情况下, μ 即代表真值; n 为测定次数。

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (1-15)$$

样本标准差 (Sample Standard Deviation): 在有限次数的测定时,标准偏差称为样本标准差,以 S 或 s 表示。

式中, $(n-1)$ 表示 n 个测定值中具有独立偏差的数目,又称为自由度 (Degrees of freedom) 用 ν 表示。即,当 $n \rightarrow \infty$ 时, $\bar{x} \rightarrow \mu$, $S \rightarrow \sigma$ 。

样本标准差 S 用式 (1-16) 计算更为方便。

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \Rightarrow S = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1}} \quad (1-16)$$

变异系数 CV (Coefficient of Variation): S 与平均值之比称为相对标准偏差,以 S_r (或 RSD) 表示:

$$S_r = \frac{S}{\bar{x}} \times 100\% \quad (1-17)$$

式 (1-17) 也可用千分率表示 (即式中乘以 1000%)。如以百分率表示又称为变异系数 CV (Coefficient of Variation)。

C 精密度

(1) 精密度: 在确定条件下,将测试方法实施多次,求出所得结果之间的一致程度。精密度的大小常用偏差表示。

(2) 精密度的高低还常用重复性 (Repeatability) 和再现性 (Reproducibility) 表示。

1) 重复性 (r): 同一操作者,在相同条件下,获得一系列结果之间的一致程度。

2) 再现性 (R): 不同的操作者, 在不同条件下, 用相同方法获得的单个结果之间的一致程度。

(3) 用标准偏差比用算术平均偏差更合理。

对比: 有两组测定值, 判断精密度的差异, 见表 1-1。

甲组 2.9 2.9 3.0 3.1 3.1

乙组 2.8 3.0 3.0 3.0 3.2

表 1-1 甲乙计算结果表

组别	平均值 \bar{x}	平均偏差 d	标准偏差 s
甲组	3.0	0.08	0.08
乙组	3.0	0.08	0.14

从表 1-1 中可以看出平均偏差相同、标准偏差不同、两组数据的离散程度不同; 在一般情况下, 对测定数据应表示出标准偏差或变异系数。

1.2.1.3 准确度与精密度的关系

准确度是指测量值与真实值之间的差异大小, 准确度越高, 则测量值和真实值之间的差异就越小。

精密度是指多次平行测量的测量值之间的接近程度, 精密度越高, 则多次平行测量的测量值之间就越接近。

准确度与精密度的关系如图 1-1 所示。

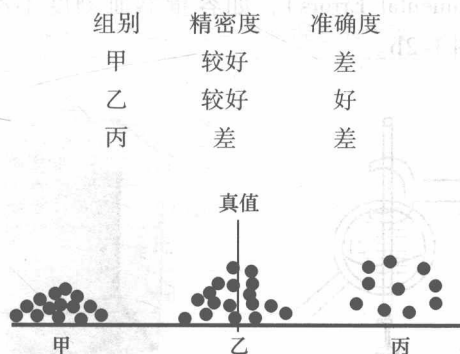


图 1-1 三组测定结果的数值分布图

由此, 我们能够得出结论:

(1) 精密度是保证准确度的先决条件;

(2) 精密度高不一定准确度高;

两者的差别主要是由于系统误差的存在而造成, 并不是离散程度大的偶然误差造成的。

[例 1-3] 分析铁矿中铁含量, 得如下数据: 37.45%, 37.20%, 37.50%, 37.30%, 37.25% 计算此结果的平均值、平均偏差、标准偏差、变异系数。

$$\bar{x} = \frac{37.45\% + 37.20\% + 37.50\% + 37.30\% + 37.25\%}{5} = 37.34\%$$

$$\bar{d} = \frac{\sum_{i=1}^n |d_i|}{n} = \frac{0.11\% + 0.14\% + 0.16\% + 0.04\% + 0.09\%}{5} = 0.11\%$$

$$S = \sqrt{\frac{\sum_{i=1}^n d_i^2}{n-1}} = \sqrt{\frac{0.11^2 + 0.14^2 + 0.16^2 + 0.04^2 + 0.09^2}{5-1}} \% = 0.13\%$$

$$CV = \frac{S}{\bar{x}} = \frac{0.13}{37.34} \times 100\% = 0.35\%$$

中位数：数据从小到大排列，测量值个数 n 为奇数时，正中间的那个数为中位数，当 n 为偶数时，中间相邻两个测量值的平均值为中位数。

极差：一组测量数据中最大值与最小值之差

$$R = X_{\max} - X_{\min}$$

1.2.2 误差的分类及减免误差的方法

1.2.2.1 系统误差产生的原因、性质及减免

系统误差或称可测误差 (Determinate Error)。

A 产生的原因

- (1) 方法误差 (Method Errors)：如反应不完全、干扰成分的影响、指示剂选择不当；
- (2) 试剂或蒸馏水纯度不够；
- (3) 人为误差 (Personal Errors)，如观察颜色偏深或偏浅，第二次读数总是想与第一次重复等造成，见图 1-2a；
- (4) 仪器误差 (Instrumental Errors)，如容量器皿刻度不准又未经校正，电子仪器“噪声”过大等造成，见图 1-2b。

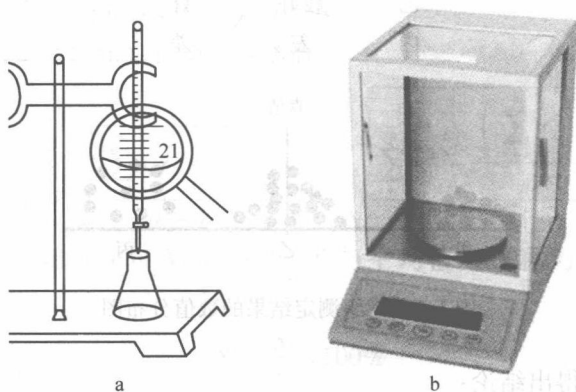


图 1-2 读数 and 仪器产生误差

B 系统误差的性质

- (1) 重复性：同一条件下，重复测定，重复地出现。
- (2) 单向性：测定结果系统偏高或偏低。
- (3) 恒定性：大小基本不变，对测定结果的影响固定。
- (4) 可校正性：其大小可以测定，可对结果进行校正。

C 系统误差的校正方法

选择标准方法、提纯试剂和使用校正值等办法加以消除。常采用对照实验和空白实验的方法。

1.2.2.2 偶然误差产生的原因、性质及减免

偶然误差或称未定误差、随机误差 (Indeterminate Errors)。

A 产生的原因

由一些无法控制的不确定因素引起的。

(1) 如环境温度、湿度、电压、污染情况等的变化引起样品质量、组成、仪器性能等的微小变化；

(2) 操作人员实验过程中操作上的微小差别；

(3) 其他不确定因素等所造成。

B 性质

时大时小，可正可负。

C 减免方法

无法消除。通过增加平行测定次数，降低。

1.2.2.3 过失误差 (粗差): 认真操作, 可以完全避免。

例如: 沉淀的溅失或玷污; 试样溶解或转移不完全或损失; 称样时试样撒落在容器外; 读错刻度; 记录和计算错误; 加错试剂等。

系统误差与随机误差的区别如表 1-2 所示。

表 1-2 系统误差与随机误差的比较

项 目	系 统 误 差	随 机 误 差
产生原因	固定的因素	不定的因素
分类	方法误差、仪器与试剂误差、主观误差	
性质	重现性、单向性 (或周期性)、可测性	服从概率统计规律、不可测性
影响	准确度	精密度
消除或减小的方法	校正	增加平行测定的次数

1.3 标准差

1.3.1 标准差的意义

用平均数作为样本的代表, 其代表性强弱受样本资料中各观测值变异程度的影响。如果各观测值变异小, 则平均数对样本的代表性强; 如果各观测值变异大, 则平均数对样本的代表性弱。因而仅用平均数对一个资料的特征作统计描述是不全面的, 还需引入一个表示资料中观测值变异程度大小的统计量。

极差是表示资料中各观测值变异程度大小最简便的统计量。极差大, 则资料中各观测值变异程度大; 极差小, 则资料中各观测值变异程度小。但是极差只利用了资料中的最大

值和最小值,并不能准确表达资料中各观测值的变异程度,比较粗略。当资料很多而又要迅速对资料的变异程度作出判断时,可以利用极差这个统计量。标准差广泛用来度量样本各观测值间的变异程度和平均数的代表情况。标准差充分考虑了各个变数与平均数的离差。每个变数与平均数相差越小,则样本变异程度越小;反之越大。如果每个变数与平均数之差为零,这表示每个变数与平均数没有差异。标准差是用各变数与平均数差的大小来度量变异程度的一个统计量。

1.3.2 标准差的计算

为了准确表示样本内各个观测值的变异程度,人们首先会考虑以平均数为标准,求出各个观测值与平均数的离差,即 $(x - \bar{x})$,称为离均差。虽然离均差能表达一个观测值偏离平均数的性质和程度,但因为离均差有正、有负,离均差之和为零,即 $\sum(x - \bar{x}) = 0$,因而不能用离均差之和,即 $\sum(x - \bar{x})$ 来表示资料中所有观测值的总偏离程度。为了合理地计算出平均差异,我们先将各个离均差平方,即 $(x - \bar{x})^2$,再求离均差平方和,即 $\sum(x - \bar{x})^2$,简称平方和,记为 SS 。

由于离差平方和常随样本大小而改变,为消除样本大小的影响,用平方和除以样本大小,即 $\sum(x - \bar{x})^2/n$ 。为了使所得的统计量是相应总体参数的无偏估计量,统计学证明,在求离均差平方和的平均数时,分母为自由度 $n - 1$ 。于是,我们采用统计量 $\sum(x - \bar{x})^2/(n - 1)$ 表示资料的变异程度。统计量 $\sum(x - \bar{x})^2/(n - 1)$ 称为均方(Mean Square, MS),又称样本方差(Sample Variance),记为 S^2 ,即

$$S^2 = \frac{\sum(x - \bar{x})^2}{n - 1} \quad (1-18)$$

相应的总体参数称为总体方差,记为 σ^2 。对于有限总体而言, σ^2 的计算公式为

$$\sigma^2 = \frac{\sum(x - \mu)^2}{n} \quad (1-19)$$

统计学上把 S^2 的平方根称为样本标准差(Sample Standard Deviation),其单位与观测值度量单位相同。由样本资料计算标准差的定义公式为

$$S = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \quad (1-20)$$

这里 $n - 1$ 为自由度(Degrees of freedom),应用自由度的目的是为了减少抽样误差的影响。例如,一个样本含有 n 个变数,从理论上说, n 个变数与 \bar{x} 之差得到 n 个离均差,但是,其中 $n - 1$ 个是可以自由变动的,最后一个离均差受 $\sum(x - \bar{x}) = 0$ 这一条件的限制而不得自由变动。例如,有5个变数,其中有4个离均差为 -2 、 -1 、 1 、 2 ,则第5个离均差值必须等于0。如果其中前4个的离均差为 -1 、 0 、 1 、 2 ,则第5个离均差只能等于 2 ,这样才能使得离均差的总和等于0。所以,这5个离均差中,只有4个能自由变动,自由度 ν 就是 $n - 1 = 4$ 。通常,自由度 ν 等于样本变数的总个数减去计算过程中使用的条件数。在计算标准差时,条件只有一个,即 $\sum(x - \bar{x}) = 0$,故自由度为 $n - 1$ 。如果计算其他统计量时,要是应用2个条件,自由度 $\nu = n - 2$,应用 k 个条件,则自由度 $\nu = n - k$ 。若样本很大时,可不用自由度,直接用 n 亦可。依照国家标准,统计学中自由度用符号 ν 表示。