

高等学校教材

# 实验设计 与数据处理

► 刘振学 黄仁和 田爱民 编著



化学工业出版社

教材出版中心

高等学校教材

# 实验设计与数据处理

刘振学 黄仁和 田爱民 编著



化学工业出版社  
教材出版中心

·北京·

(京)新登字 039 号

**图书在版编目 (CIP) 数据**

实验设计与数据处理/刘振学,黄仁和,田爱民编者. —北京:化学工业出版社,2005.3  
高等学校教材  
ISBN 7-5025-6547-7

I. 实… II. ①刘…②黄…③田… III. ①实验设计(数学)-高等学校-教材②实验数据-数据处理-高等学校-教材 IV. ①0212.6②N33

中国版本图书馆 CIP 数据核字 (2005) 第 012254 号

---

高等学校教材

**实验设计与数据处理**

刘振学 黄仁和 田爱民 编著

责任编辑:杨 菁

文字编辑:朱 磊 徐卿华

责任校对:陈 静 李 军

封面设计:潘 峰

\*

化学工业出版社 出版发行  
教材出版中心

(北京市朝阳区惠新里 3 号 邮政编码 100029)

发行电话:(010) 64982530

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

\*

新华书店北京发行所经销

化学工业出版社印刷厂印装

开本 787mm×1092mm 1/16 印张 13 字数 317 千字

2005 年 4 月第 1 版 2005 年 4 月北京第 1 次印刷

ISBN 7-5025-6547-7/G·1673

定 价:23.00 元

---

**版权所有 违者必究**

该书如有缺页、倒页、脱页者,本社发行部负责退换

# 前 言

数学是自然科学和社会科学最基础的学科。掌握了数学工具，也就拿到了开启成功之门的钥匙。

理工科大学生在大学阶段会学到不少数学原理和方法，但那是纯理论的。工科大学生也学习工程数学，但很少应用于实践；而理科大学生就学到很少或根本不学。由纯数学理论到生产实践，中间似乎还缺少一些过渡。因此，完成理论联系实际，由数学的纯理论转向实际应用，就需要开设另外的课程。

实验设计与数据处理课程是学习某些课程的必需知识，现将其独立了出来，成为化学工程和工艺、矿物加工工程等本科专业及硕士研究生教学独立的课程。又根据需要不断调整、完善。

本书共 10 章，由三部分组成。1~4 章为第一部分，主要介绍测量值与误差的分布、性质及检验；同时对方差分析这一纯数学的内容进行介绍，在简单介绍数学模型的基础上，通过简单推导出一系列方差分析公式，然后重点介绍方差分析的应用，并利用方差分析解决一些实际问题，这四章起着承前启后的作用。5~9 章为实验设计及统计方法应用内容，主要介绍正交实验设计、多因素序贯实验设计、随机化区组和拉丁方设计的方法和应用，线性回归方法及检验，主成分分析方法及应用，聚类分析方法及应用等内容；第 9 章为相对独立的内容，主要介绍利用误差理论进行分析化学质量管理。第 10 章单独成为第三部分，简单介绍目前最流行的大型统计软件之一 SPSS (Statistical Program for Social Sciences)，并结合前面各章内容对其使用进行介绍。

本书第 4 章为黄仁和编写，第 6 章为王力编写，第 7 章和 8.1、8.2 节为田爱民编写，其余各章节为刘振学编写，由刘振学对全书内容进行统稿。在编写过程中，还结合科研和教学经验，对部分内容进行了创作。马继红、汪兴隆和武艳菊等在读研究生参与了部分文字图表以及公式等的录入工作，在此向他们表示感谢。同时，在本书编著过程中，还得到了我院谭允楨院长、周仕学副院长等领导 and 同志们的关心和鼓励，在此一并表示深深的谢意。

由于作者水平所限，书中错误在所难免，望读者批评指正。

编者

2004 年 8 月

## 内 容 提 要

本书主要包括三部分内容,即数据处理基础、实验设计方法与应用和计算机数据处理软件简介。分别介绍测量值与误差、偶然误差的分布、误差传递等误差理论中的内容,介绍统计检验、方差分析、回归分析和聚类分析等数据处理方法与应用,介绍提高分析化学准确度的方法及质量控制方法,介绍正交实验设计、多因素序贯实验设计、随机化区组和拉丁方设计方法与应用,最后对现时最流行的大型统计软件 SPSS (Statistical Program for Social Sciences) 进行简要介绍,书末附有习题及常用的统计数表。

本书着重介绍基本概念和基本理论,并在此基础上结合专业特点,介绍了各种统计方法在化学化工、医药、环境检测、矿物加工等多方面的应用,本书把质量控制和实验设计作为重点。

本教材适于工艺、工程类本科生和理工类研究生教学使用,尤其适用于化学化工、矿物加工、医学和环境学等学科的本科生和研究生使用。本教材对于科研人员进行实验设计可提供很大的帮助,也可供广大分析化学工作者自学使用。

# 目 录

## 第 1 篇 数据处理基础

<b>1 绪论</b> .....	1
1.1 教学目的 .....	1
1.2 实验设计的性质和价值 .....	1
1.3 实验研究方法 .....	4
<b>2 误差和数据处理</b> .....	6
2.1 误差及其表示方法 .....	6
2.1.1 系统误差和偶然误差 .....	6
2.1.2 准确度和精密度 .....	7
2.1.3 误差和偏差 .....	7
2.1.4 标准偏差的计算 .....	11
2.2 偶然误差的正态分布 .....	12
2.2.1 频数分布 .....	12
2.2.2 分布函数 .....	13
2.3 误差传递 .....	15
2.3.1 系统误差的传递 .....	15
2.3.2 偶然误差的传递 .....	17
2.3.3 极值误差与误差分配 .....	20
2.4 有效数字及运算规则 .....	21
2.4.1 有效数字 .....	21
2.4.2 数字修约规则 .....	21
2.4.3 运算规则 .....	22
2.4.4 测量值的记录 .....	23
<b>3 有限数据统计处理</b> .....	24
3.1 总体的参数估计 .....	24
3.1.1 期望值和方差 .....	24
3.1.2 参数估计 .....	25
3.2 一般的统计检验 .....	29
3.2.1 离群值检验 .....	29
3.2.2 平均值检验 .....	33
3.2.3 $F$ -检验 .....	36
<b>4 方差分析</b> .....	39
4.1 概述 .....	39
4.1.1 基本概念 .....	39

4.1.2	方法	40
4.2	单因素方差分析	40
4.2.1	单因素方差分析基本公式	40
4.2.2	偏差平方和	41
4.2.3	方差分析统计量	42
4.3	无重复两因素方差分析	47
4.3.1	无重复两因素方差分析的数学模型	47
4.3.2	偏差分解	49
4.3.3	自由度	49
4.3.4	方差	50
4.3.5	$F$ -检验	50
4.3.6	方差分析表	51
4.4	无重复两因素方差分析	53
4.4.1	交互作用	53
4.4.2	偏差分解	54
4.4.3	自由度	55
4.4.4	方差	55
4.4.5	$F$ -检验	56
4.4.6	方差分析表	56

## 第 2 篇 实验设计与统计应用

<b>5</b>	<b>正交实验设计</b>	<b>62</b>
5.1	概述	62
5.1.1	正交表	62
5.1.2	正交表的特点	63
5.1.3	正交表的优点	64
5.1.4	正交表的分类	64
5.2	正交实验设计	65
5.2.1	正交实验设计步骤	65
5.2.2	正交实验的数据处理	65
5.3	多指标的实验	68
5.3.1	综合评分法	68
5.3.2	综合平衡法	69
5.4	有交互作用的设计	71
5.4.1	正交表的选择和表头设计	71
5.4.2	按方案进行实验	72
5.4.3	结果分析	72
5.5	正交实验的方差分析	73
5.5.1	总变差的分解	73
5.5.2	分析方法	74

5.5.3 适应范围	74
<b>6 多因素序贯实验设计</b>	76
6.1 最陡坡法	76
6.1.1 实例	76
6.1.2 数学原理	79
6.1.3 应用条件	80
6.2 调优运算和单纯形调优法	82
6.3 消去法	83
6.3.1 “0.618”法	84
6.3.2 分批实验法	84
6.4 小结	86
<b>7 随机化区组和拉丁方</b>	87
7.1 随机化区组	87
7.1.1 什么是随机化区组	87
7.1.2 随机化区组设计的数据处理	88
7.1.3 小结	91
7.2 拉丁方	91
7.2.1 概述	91
7.2.2 拉丁方设计的数据处理	92
7.2.3 拉丁方设计举例	93
7.3 正交方	94
<b>8 回归分析与聚类分析初步</b>	97
8.1 一元线性回归分析	97
8.1.1 概述	97
8.1.2 回归方法	97
8.1.3 一元线性回归方程的检验	100
8.2 二元线性回归	103
8.2.1 回归方法	103
8.2.2 二元线性回归方程的检验	104
8.3 主成分分析	106
8.3.1 概述	106
8.3.2 主成分的计算	107
8.3.3 主成分分析在企业效益中的应用	110
8.4 聚类分析	111
8.4.1 概述	111
8.4.2 样品间的距离	112
8.4.3 聚类方法	114
8.4.4 其他聚类方法	119
<b>9 质量控制</b>	121
9.1 分析化学质量控制方法	122



9.1.1	全程序空白实验值控制 .....	122
9.1.2	标准曲线 .....	122
9.1.3	平行双样 .....	123
9.1.4	加标回收 .....	124
9.1.5	标准参考物的应用 .....	125
9.1.6	方法对照 .....	125
9.2	质量管理图 .....	125
9.2.1	概念 .....	125
9.2.2	质量管理图的绘制与应用 .....	127
9.2.3	控制图的识别 .....	132

### 第3篇 计算机程序简介

<b>10</b>	<b>大型统计软件 SPSS 简介 .....</b>	<b>135</b>
10.1	SPSS 简介 .....	135
10.1.1	SPSS 概述 .....	135
10.1.2	SPSS 界面 .....	135
10.1.3	SPSS 的菜单 .....	136
10.1.4	SPSS 的其他窗口 .....	137
10.1.5	SPSS 的退出 .....	137
10.2	SPSS 基本操作 .....	137
10.2.1	数据的输入 .....	137
10.2.2	统计分析 .....	140
10.2.3	图形分析 .....	141
10.2.4	结果输出 .....	141
10.3	平均值检验 .....	142
10.3.1	Means 过程 .....	142
10.3.2	两组数据 $t$ -检验 .....	144
10.3.3	配对比较 .....	145
10.4	方差分析 .....	147
10.4.1	单因素方差分析 .....	147
10.4.2	两因素方差分析 .....	148
10.5	正交实验设计与方差分析 .....	150
10.5.1	正交实验设计 .....	150
10.5.2	正交实验的方差分析 .....	153
10.6	线性回归 .....	155
10.6.1	一元线性回归 .....	155
10.6.2	二元线性回归 .....	156
10.7	主成分分析 .....	158
10.8	聚类分析 .....	161
10.8.1	K-Means Cluster 过程 .....	161

10.8.2 Hierarchical Cluster 过程 .....	165
10.9 质量控制图 .....	170
<b>习题</b> .....	172
<b>附录</b> .....	179
附录 1 Dixon 检验统计量和临界值 .....	179
附录 2 Grubbs 检验 $T_{\alpha,n}$ 值表 .....	180
附录 3 Cochran 最大方差检验的临界值 .....	180
附录 4 标准正态分布表 .....	181
附录 5 $t$ -分布表 .....	182
附录 6 $F$ -分布表 .....	183
附录 7 相关系数临界值表 .....	186
附录 8 正交表 .....	187
<b>参考文献</b> .....	198

# 第 1 篇 数据处理基础

## 1 绪 论

### 1.1 教学目的

本教材主要讨论实验工作的设计方法、分析数据的统计处理、分析化学质量控制方法等，并对误差理论进行介绍。与分析化学不同，分析化学课本中包含少量数据处理方法，但非常简略。同样，与数理统计教材也具有很大区别。因为一般的数理统计只介绍理论而实用的例子很少，这对于专业课程来说，似乎存在一个不大不小的断层，学生们往往只知道一些理论却不知如何应用。在本书中，主要考虑与实验设计有关的分析并解释实验结果的统计方法的运用。

大部分化学和物理本科毕业生只学过非常简单的统计方法和知识，他们往往认识不到这些方法对其毕业后工作的重要性。他们没有掌握和运用，所以跟没有学过差不多，因此他们可能会认为，统计对于化学化工专业可有可无。因此，此处必须强调指出，凡是涉及到数据的问题，只要数据中包含有相当大的实验误差，则获得满意结果的惟一稳妥的处理办法便是统计方法，除此以外别无他择。同时可以知道，实验数据不存在误差是不可能的。若仅仅凭借自己的判断能力去评价一个人的工作结果，所采用或应当采用的判据可能与统计检验所依据的判据相类似，但由于没有利用数字进行定量描述与计算，结论会带有很大的主观随意性，用分析化学的语言描述，就是存在主观误差。若所研究的效应大于随机变差，这种做法或许是满意的。但是当与效应相比，随机变差显得相当大时，这种直观判断就可能导致错误，用它来替代严谨的统计检验，就不合适了。在大学实验室里，理工科大学生们用高纯度的试剂材料进行严格控制下的高精度实验，偏差可能很小。但工业生产中，这种理想条件是达不到的，即使能够达到，所取得的结果也往往不能直接应用于在现实中所遇到的更为复杂的情况。因此，统计方法应当作为从事工业生产的科技人员所必须掌握的一门技术，用来有效地处理工业生产中的各种问题。

本教材要求读者熟悉比较简单的统计技术。

### 1.2 实验设计的性质和价值

化学工业中的实验工作五花八门，但主要部分可归纳为下面几种类型的研究工作。

① 物理或物理化学研究 这类实验通常涉及物质的基本常数和特征的精确测定。这种工作与大学中进行的研究工作非常接近，实验的设计一般取决于它们的性质，除了有时用到误差评价外，一般不常用统计方法。

② 产品、原料等的常规分析 化学分析中遇到的系统误差往往大于随机误差。这些系统误差常常随不同的分析者而异，或随同一分析者的不同分析时间或采用不同方法、试剂而异。这种情况下，即使由同一分析者重复检验几次，也很难提高结果的精度，通常只能借以检出错误。除了实际的分析误差外，在抽取样品或把一个样品划分为适于分析的若干小的操作过程中，也会进一步引入相当的误差。所以，审查抽样或分析的方案了解所涉及的误差大小至关重要。对这类误差的研究需进行一定的设计，若想取得可靠的估计值，最好的方法是在实验模型的选择和结果的解释中都采用统计方法。

③ 材料实验 在对材料的机械和其他性质（如橡胶的耐磨强度）的实验中，或对那些在定义上还不明确的性质（如抗蚀性或耐久度）作评价时，每个观察值往往都带有可观的随机误差。因此，为了获得可靠的估计值，必须从相当数量的观测值中取均值。凡是涉及这类实验的研究工作，均需采用统计法的合理设计。

④ 化工过程的实验室研究 几乎所有化工新过程和大部分旧过程的改革，都是从小规模的实验室研究开始的。由于早期仅包括简单的和目的专一的实验，所以对设计要求不高。但随着研究的深入，将来就必须对各种材料或条件的变化对过程的效应进行系统性研究。这时，统计学是科研工作者选择最优设计的有力武器。

⑤ 化工厂操作和生产过程的控制、研究 这类研究涉及过程效率的确定，或在正常工作条件下过程操作的研究，或由于各种操作条件、原材料的变化所起的效应等。这类实验可能很费钱，也可能需要多人协作，因而必须认真研究以确保采用最合适的设计。

研究任何指定的项目，尤其是④型和⑤型的项目时，研究人员可以从对所研究的特定过程有影响的各项基本关系入手，或者采用某种更为经验性的办法，直接在该过程中研究各种变化的效应。如果采用基本关系研究的方法，研究人员可能对各种物理参量加以研究。例如，与该过程有关的各种化学反应的速率和平衡常数，以及与之有关的各项传递系数、渗透度等，并以这一切为基础，推测出操作过程的最佳条件和其中任一条件发生变化的效应。但是，如果研究人员采用经验的方法，可以直接测定各种条件变化的效应，而不是深究那些导致他所观测到的效应的确切机制。在工业研究中，往往由于问题非常复杂，想要研究所观测的所有各种效应的潜在原因的工作量很难完成，因此就必须采用后一种经验方法。在这类经验法研究中，以统计原理为根据的实验设计为取得所需的实践资料提供了经济手段。即使采用的是基本关系的研究方法，往往也会发现由于它是以实际系统的简化为根据，要使实际过程最后调整到它的最佳条件，仍然需依靠经验法完成。

一项优良的实验设计，能以最少的实验工作量取得所需的资料。为此需做三件事。

- a. 必须正确地列出要通过实验解答的各项问题，即明确实验的目的；
- b. 必须在兼顾所要求的精度和可能碰到的实验难点的条件下正确地选用实验方法；
- c. 必须正确选定实验的一般数学模型，即历次观测的数目、周期和相互关系。

实验设计的统计理论所论述的就是这种包括在一组观测值中各项数据的数目及其相互关系的一般模型。利用数学理论，可获得一个实验方案所提供资料的定量度量，然后将它与不同的方案对比，评价它们对任一给定项目的适应性。通过这类研究，加上运用它们所取得的实践经验，就可以发展成为一门涉及到各种类型实验设计的学科，用来指导科技人员从中选择出适应于特定项目的设计。

任何一组观测数据，都是在所研究的各种条件下的一种模型或排列。如果在研究过程中，以杂乱无章的方式取得结果，那么最终的排列也是不平衡和不规则的。但若予以审查，

往往会发现当初要是设计得当，用较少的工作量也能取得等量的资料。在这个领域里，统计学的作用是提供各项判据，是研究者可据以判断所建议的某项设计的效率，并为他们提供一些经过理论和实践证明了的、对处理某些类型的问题特别有效的某种标准设计模型。用统计方法设计实验的一个附带的好处是，迫使运用统计方法的研究人员必须预先考虑他们在寻求哪些目标，必须采用哪些步骤去探索这些目标。此外，还迫使他们考虑所有可能来源误差的大小。这种先期的考虑本身就具有重大价值，因为由此可以引导研究人员认识到并从中回避可能的难点和失误。如果没有先期的考虑，这些难点或失误可能在研究工作的后期才被发现，那就会带来很多不便。

统计学的重要作用之一，是为确定观测数据的数目提供合理的依据。由于任何度量都带有一定程度的随机误差，因此必须综合好多观测数据的结果，才能使所得的资料具有足够的精度。在工业工作中，这种随机误差相当可观，因此在为一项指定目标而设计的研究工作中，要求以某些最小数目的观测值就可获得必需的精度。如果观测值超过了这个最小数目，那么原可更好地用与其他工作的实验时间和精力、物力等于是浪费了。这相当于工程师们所谓的超需设计，在工业研究中，其虚耗的程度往往抵得上一件工业设备的设计代价。反过来，如果实验属欠需设计的话，即观测数目太少，也可能得出虚假的结论。除非研究人员掌握了若干可供判断实验正确数目的明确判据，这类超需设计或欠需设计就很难免。经验证明，正确的设计类型和恰到好处的观测数目两者的综合效果，可对一个科研项目的实验工作量带来经济上很大的节省。这种经济上节省的价值，可大大补偿为设计实验所花费的额外时间和脑力劳动。由于研究工作的主要开支，几乎都花在实验设计上，所以实验开支的减少会降低整个研究工作的成本。

为一项实验做出最佳最省的设计，需要对有关的误差和需观测的效应大小有所认识。这似乎很困难，但实践中却很少要求研究人员去完成那种有关误差的先期资料一无所有的实验。这类资料有时是粗略和不确定的，即使如此，利用这些资料进行设计，虽然比资料充分时的设计差一些，但无疑比不加设计的实验好得多。如果实验工作的合理设计是有规律的定期进行的，应尽快建档收集这类资料，以供以后设计类似实验时参考。如果运用得当，合理设计的主要特点之一是鼓励研究人员在日常工作中充分运用全部知识和经验，帮助他们做好本职工作。

除了早期投资利用的资料外，伴随着工作的进行可以获得新的资料。例如，在化工研究中，各项实验或试制通常是一次一个结果或一次几个结果，因此在一项实验的过程中，先前实验的结果在其后的实验未开始前即可利用。这时，可以运用已有的资料，确定实验设计和进一步实验的次数。这种利用一系列观测中的先前部分获得的资料，涉及其后作业的概念，称作对一个问题的序贯分析。其最简单的形式是，先只做一次小规模中间实验，取得有关误差或效应的一般性质的概况资料，然后利用这些资料去设计整个实验。如果系统地运用序贯分析，其优越性更为明显。

在解释实验结果时，常常用到显著性统计检验。在这种检验中，通常假设某种想要探明的效应并不存在，然后去考察偏离此假设的差异是否有理由归因于偶然性的原因。如果发现出现这种差异的概率小于某一水准，例如小于  $1/20$ ，即可认为该假设不成立，并称该效应是显著的。但若概率大于这一规定的显著性水准，则称没有足够的证据去说明该效应的存在。这里当然必须应用各种显著性检验并进行审慎解释。虽然本书中大多数设计都是为了使显著性检验套用得上而设计的，但是它们的价值却与这些检验无关。但若多次实验的变异性

与所考查的效应相比显得过大时,就需取得大量的观测数据,才能得出肯定的结论。这类情况下,在对任何效应确认其真实性质之前,最好先做一下显著性的检验。此外,由于检验结果在很大程度上取决于随机误差的大小,因此最好采用这样的设计,即可从实际的实验估计出这些误差。

### 1.3 实验研究方法

① 误差理论 由于《分析化学》中已学习过关于误差的基本知识,本书中的这部分内容就是在以前所学知识的基础上,深入了解误差及其相关概念和计算等。主要掌握误差的概念、分类和表示、误差的计算、误差传递以及有效数字运算等。其中,误差理论是这门课的基础。此外,一种完美的分析方法及结果表述应注意有效数字。

② 有限数据统计处理 这是概率论知识的基本应用之一,在《概率论》学习中,由于有许多较抽象的数学理论,往往是学生对所学知识印象不深,而且有些时候往往不知道哪些是重点。化学化工中只用到不多的概率统计知识,通过本部分内容的学习,可更好地掌握这些知识,同时也能指导同学们将其他知识应用到实践中去。

③ 方差分析 在统计的实际应用中许多地方用到方差分析,方差分析也是进行实验设计所需掌握的基本内容之一。本章中论述了因素、水平、交互作用、指标等许多具体概念,并且对误差的概念进行了重新定义。可以说这一章内容是承上启下的,它用了前几章学到的知识,又为下一步实验设计的学习打下了基础,这是本课程的难点,也是重点之一。

④ 正交实验设计 正交实验设计是科研和生产中应用最多的实验研究方法之一,尤其用于生产改造、最优配方及最优工艺过程的研究。由于它方便、简洁而得到研究人员的认可。

⑤ 多因素序贯实验设计 在多因素多水平的情况下,为了减少实验工作量,除了可采用正交实验设计以外,更多地应用的是序贯实验设计法。多因素组合实验时,全面实验的工作量很大,采用序贯实验法就可明显地减少试点总数,缩短实验周期。

⑥ 相关和回归 一般认为,线性回归并不重要也不难掌握。其实,说它不难掌握是对的,但它在实际工作中却有重要用途。除了用它进行预测和控制外,对于一个化学化工工作者,产品质量控制及实验室管理都是不可缺少的工作。

⑦ 多元统计和聚类分析 一般的数学教科书(当然也包括数理统计教材)中很少介绍多元统计的内容,而实际上多元统计对于化学化工、矿物加工、环境科学、医学与生命科学以及许多社会科学等的科学研究与结果分析都有着重要用途。因此,本教材将对这种统计方法进行简单介绍。

⑧ 质量管理与控制 有些数学教科书中包含质量管理内容,但很粗糙,而且只包含基本的统计内容。本课程是从化学化工角度,利用分析化学中的理论,对这一问题进行了较详细的论述。对于科研或化工过程的控制、产品质量把关等均具有较大的实用价值。

⑨ 随机化区组和拉丁方 对于少数几个因素的实验或结果受区域、温度影响及样品数量限制的实验研究,可以应用随机化区组或拉丁方设计。这是方差分析基本应用之一,对于一些专门的检测或设计,可以得到满意的结果,它具有正交设计的优点,又可方便地进行少因素多水平实验。

⑩ 大型统计软件——SPSS应用简介 对于统计分析工作者而言,SPSS——Statistical

Program for Social Sciences 无疑是最优秀的统计软件之一，而且目前已有多个版本面世。但对于一个科技工作者，要掌握它的使用绝非易事。本教材对它的使用进行简要介绍，掌握了这些方法后，再在实践中进行一些较为深入的摸索，就可很快掌握它。那时就会发现，它的好处实在是无与伦比。

## 2 误差和数据处理

### 2.1 误差及其表示方法

在定量分析中，分析结果应具有一定的准确度，因为不准确的分析结果会导致产品报废、资料浪费，甚至在科学上得出错误的结论。但是，在分析过程中，即使是技术很熟练的人，用同一方法对同一试样仔细地进行多次分析，也不能得出完全相同的分析结果，而是分析结果在一定的范围内波动。这就是说，分析过程中的误差是客观存在的。

分析工作应该做到既快速又准确。但是，两者同时达到是不太现实的，这就要求分析者对两者进行判断。快速与准确两者之间谁是主要方面，则需视实际需要才能确定，有时快速是主要的，如钢铁厂的炉前快速分析等；有时准确度是主要的，如原子量和各种常数的测定等。由于各种分析方法所能达到的准确度不同，因此，应根据具体情况，设法使分析结果达到一定的准确度，以适应各种工作的需要。

#### 2.1.1 系统误差和偶然误差

在定量分析中，对于各种原因导致的误差，根据其性质的不同，可以区分为系统误差和偶然误差两大类。

##### (1) 系统误差

系统误差是由某种确定的因素造成的，使测定结果系统偏高或偏低；当造成误差的因素不存在时，系统误差自然会消失。当进行重复测量时，它会重复出现。系统误差的大小，正负是可以测定的，至少在理论上说是可以测定的，所以是可测误差。系统误差的最重要特性是它具有“单向性”。

根据系统误差的性质和产生的原因，可将其分为如下几种。

① 方法误差 这种误差是由分析方法本身造成的。例如，在重量分析中，沉淀的溶解、共沉淀现象、灼烧时沉淀的分解或挥发等；在滴定分析中，反应进行不完全、干扰离子的影响、化学计量点和滴定终点不一致及副反应的发生等，都会系统地导致测定结果偏高或偏低。

② 仪器和实际误差 仪器误差来源于仪器本身不够精确，如砝码质量、容量器皿刻度和仪表刻度不准确等。实际误差来源于试剂不纯。例如，试剂和蒸馏水中含有被测物质或干扰物质，使分析结果系统偏高或偏低。

③ 操作误差 操作误差是由分析人员所掌握的分析操作与正确的分析操作有差别引起的。例如，分析人员在称取试样时未注意防止试样吸湿，滴定分析中移液管或滴定管不洁净，洗涤沉淀时洗涤过分或不充分，灼烧沉淀时温度过高或过低，称量沉淀时坩埚及沉淀未完全冷却等。

④ 主观误差 主观误差又称个人误差。这种误差是由分析人员本身的一些主观因素造成的。例如，分析人员在辨别滴定终点的颜色时，有的人偏深，有的人偏浅；在读取刻度值时，有的人偏高，有的人偏低等。在实际工作中，有的人还有一种先入为主的习惯，即在得到第一个测量之后，在读取第二个测量值时，主观上尽量使其与第一个测量值相符合，这样也容易引起主观误差。主观误差有时列入操作误差中。



## (2) 偶然误差

偶然误差又称随机误差，它是由一些随机的、偶然的原因造成的。例如测量时环境温度、湿度和气压的微小波动，仪器的微小变化，分析人员对各份试样处理时的微小差别等，这些不可避免的偶然原因，都将使分析结果在一定范围内波动，引起偶然误差。由于偶然误差是由一些不确定的偶然原因造成的，因而是可变的，有时大，有时小，有时正，有时负，所以偶然误差又称不定误差。偶然误差在分析操作中是无法避免的。例如一个很有经验的人，进行很仔细的操作，对同一试样进行多次分析，得到的分析结果却不能完全一致，而是有高有低。偶然误差的产生难以找出确定的原因，似乎没有规律性，但如果进行很多次测定，便会发现数据的分布符合一般的统计规律。这种规律是“概率统计学”研究的重要内容。由于它们涉及到较多的数学处理，而这些内容不属于本课程的内容，故在下面将只引用结论而不讨论其数学处理。如有必要可参考有关教材或专著。

## (3) 过失误差

在分析化学中，除系统误差和偶然误差外，还有一类“过失误差”。过失误差是工作中的差错，是由于工作粗心马虎，不按操作规程办事等原因造成的。例如读错刻度、记录和计算错误及加错试剂等。在分析工作中，当出现很大误差时，应分析其原因，如是过失所引起，则在计算平均值时舍去。通常，只要加强责任感，对工作认真细致，过失是完全可以避免的。

过失误差一般归于系统误差。

### 2.1.2 准确度和精密度

分析结果和真实值之间的差值叫误差。误差越小，分析结果的准确度越高，就是说，准确度表示分析结果与真实值接近的程度。

在实际工作中，分析人员在同一条件下平行测定几次，如果几次分析结果的数值比较接近，表示分析结果的精密度高。也就是说，精密度表示各次分析结果相互接近的程度。在分析化学中，有时用重复性 (repeatability) 和再现性 (reproducibility) 表示不同情况下分析结果的精密度。前者表示同一分析人员在同一条件下所得分析结果的精密度，后者表示不同分析人员或不同实验室之间在各自的条件下所得分析结果的精密度。

精密度高不一定准确度高，因为这时可能有较大的系统误差。例如甲、乙、丙三人同时测定

一铁矿石中的含量 (真实含量 54.36%)，各分析四次，测定结果 (%) 如表 2-1 所示。

所得分析结果绘于图 2-1 中。

由图 2-1 可见，甲的分析结果的精密度很高，但平均值与真实值相差颇大，说明准确度低，即分析结果存在很大的系统误差；乙的分析结果精密度不高，准确度也不高；只有丙的分析结果的精密度和准确度都比较高。精密度是保证准确度的先决条件，精密度低说明所测结果本身就不可靠，当然其准确度也就不高。因此，如果一组测量数据的精密度很差，自然失去了衡量准确度的前提。

### 2.1.3 误差和偏差

前面已经提到，测定结果 ( $X$ ) 与真实值 ( $\mu_0$ ) 之间的差值称为误差 ( $E$ )，即

$$E = X - \mu_0 \quad (2-1)$$