

“十二五”普通高等教育本科规划教材

实验设计与数据处理

SHIYAN SHEJI YU SHUJU CHULI 刘振学 王力 等编

第二版



化学工业出版社

“十二五”普通高等教育本科规划教材

实验设计与数据处理

SHIYAN SHEJI YU SHUJU CHULI 刘振学 王力 等编

第二版



化学工业出版社

本书主要包括三部分内容,即数据处理基础、实验设计与统计应用和计算机程序简介。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、主成分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、析因设计和响应面设计的方法与应用;最后,对现时最流行的大型统计软件之一SPSS (statistical product and service solutions) 进行了简要介绍,并用该软件对前述各章中的部分例题进行了分析处理。书末附有习题及常用的统计数表和常用正交表。

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。本书把误差与数据处理、质量控制和实验设计作为重点。

本书适合工艺、工程类大学生和理工类研究生教学使用,尤其适合化学化工、矿物加工、医学和环境学等学科的本科生和研究生使用,也可供广大分析化学工作者自学应用。

图书在版编目(CIP)数据

实验设计与数据处理/刘振学,王力等编. —2版. —北京:化学工业出版社,2015.2

“十二五”普通高等教育本科规划教材

ISBN 978-7-122-22593-1

I. ①实… II. ①刘…②王… III. ①试验设计-教材②实验数据-数据处理-教材 IV. ①O212.6

中国版本图书馆CIP数据核字(2014)第302911号

责任编辑:杨菁

文字编辑:徐雪华

责任校对:宋玮

装帧设计:韩飞

出版发行:化学工业出版社(北京市东城区青年湖南街13号 邮政编码100011)

印装:三河市延风印装厂

787mm×1092mm 1/16 印张15¼ 字数380千字 2015年3月北京第2版第1次印刷

购书咨询:010-64518888(传真:010-64519686) 售后服务:010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书,如有缺损质量问题,本社销售中心负责调换。

定 价:32.00元

版权所有 违者必究



前言

本教材是在 2005 年出版的《实验设计与数据处理》一书的基础上,加入了编者近几年的最新成果和经验,并对原有内容进行了扩充,进而修订而成。本教材第一版出版后,受到了社会的普遍关注,不少读者提出了中肯的意见和建议,为本教材的修订提供了大量思路,编者在此表示感谢;还有一些读者特意指出了教材中的个别文字、数字的错误,编者在此对他们的认真态度致以诚挚的敬意。

本教材第一版出版已有九年多的时间了,在此期间国内外形势都发生了巨大变化。这些变化是我们随时随地都能体会到的。由于我国目前所处的发展时期比较特殊,我国在管理政策上进行了很大的调整,因此本教材修订时充分考虑到这一特殊国情,对相关内容进行了修订,以做到与时俱进。如关于离群值的计算及判断,就采用了最新的国家标准。期间,本教材还被评为山东省优秀教材,这也督促我们在修订本教材时更加认真,对教材中的措辞、数据等进行了更加严谨的审查和更正。同时,这些年来,同仁们也把他们大量的成果予以分享,本教材修订过程也向他们进行了广泛的学习,以期能够在这一领域做出更多有益的事情,为祖国的建设和繁荣昌盛贡献更多的力量。

本教材修订时仍保留了原来的框架和定位。全书仍包含 3 篇,每一篇包含若干章节。其中第 1 篇主要对误差理论和有限数据统计处理进行了较大修订。第 2 篇除了在正交设计一章增加了大量内容外,根据读者建议结合我们自己的体会增加了析因设计和响应面设计两章内容;其中析因设计主要为响应面设计打基础,而响应面设计的宗旨是精细设计,是属于较高层次的设计,这一方法的掌握更加有助于研究人员科研能力的提高。第 3 篇仍只介绍 SPSS 软件。由于该软件更新很快,本次修订采用了现在比较流行的 19.0 版,而且本版软件比较容易获得。

本书原有的各章内容仍由原编写老师修订完成,新增加的两章中,第 8 章 析因设计初步和第 9 章 响应曲面设计分别由王力和刘振学编写,最后由刘振学对全书内容进行调整、完善。同时,在本书修订过程中,还得到了山东科技大学化学与环境工程学院周仕学院长、邵谦副院长和吕俊俊副院长等领导和同事们的关心和鼓励,在此一并表示深深的谢意。

由于编者水平有限,加之时间紧迫,书中疏漏之处在所难免,望读者批评指正。

编者

2014 年 8 月



第一版 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。

对于理工科大学生来说，大学四五年中会学到不少数学原理和方法，但那是纯理论的，若要求学生们自己将之应用于实践，就需要他们花费大量时间去摸索；同时，一些相关概念和方法在大学数学（包括数理统计）中也不作介绍，因为那是专业教科书中的内容，比如一些误差、公差等。因此，虽然学生学了许多数学和统计方法，但会用的不多。而由于专业问题，让数学老师讲解一些专业性很强的东西也不现实。因此，完成理论联系实际，由纯理论的统计学理论过渡到实际应用，就需要开设另外的课程。

我院的《实验设计与数据处理》开设已经有了近 20 年的历史。最初是某些课程的需要，花费一定课时进行误差与数据处理方面的讲解和练习，尔后慢慢独立了出来，成为化学工程和工艺、矿物加工等本科专业及硕士研究生教学独立的课程。随着课程的进行，老师同学们又有了新的需要，于是慢慢形成了新的内容体系，尔后又逐步趋向完善，最终成为了现在的样子。

实验设计是一门实用性很强的课程。当我们进行科学研究、工艺改造和优化时，不先进行认真设计，不进行严格的数学运算，最后的结果仍然缺乏说服力。一个生产过程不可能停下来让你进行科学研究，你必须在不影响正常生产的前提下进行工艺过程改造，而且生产过程中的试验研究成本很高，没有认真严谨的设计和科学的数据处理是不行的。一个人一直埋头于实验室，他可能会干出一些成绩；但如果注意科研工作的方法论，他的工作效率会高得多，出成果也会快得多。当然，进行实验设计，对取得的实验数据进行科学处理以得出合理科学而又全面的解释，还要借助于数理统计方法。

本书第四章为黄仁和老师完成，第六章为王力老师完成，第七章和第八章的 1、2 节为田爱民老师完成，其余各章节为刘振学老师完成，最后由刘振学老师对全部内容进行调整、完善。除了参考其他教科书的内容外，我们还都结合自己的科研和教学经验，对各章节中的许多内容进行了创作。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，还得到了我院谭允祯院长、周仕学副院长等领导和同事们的关心和鼓励，在此一并表示深深的谢意。

由于编者水平所限，书中错误在所难免，望读者批评指正。

编者

2004 年 8 月



目录

第 1 篇 数据处理基础

第 1 章 概述	1
1.1 本教材的目的	1
1.2 实验设计的性质和价值	2
1.3 本课程将学到或用到的统计研究方法	4
第 2 章 误差和数据处理	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	8
2.1.4 标准偏差的计算	12
2.2 偶然误差的正态分布	13
2.2.1 频数分布	13
2.2.2 分布函数	14
2.3 误差传递	15
2.3.1 系统误差的传递	16
2.3.2 偶然误差的传递	18
2.3.3 极值误差与误差分配	21
2.4 有效数字及运算规则	22
2.4.1 有效数字	22
2.4.2 数字修约规则	22
2.4.3 运算规则	23
2.4.4 测量值的记录	23
第 3 章 有限数据统计处理	25
3.1 总体的参数估计	25
3.1.1 期望值和方差	25
3.1.2 参数估计	26
3.2 一般的统计检验	30
3.2.1 离群值检验	30

3.2.2	平均值检验	34
3.2.3	F -检验	38
第4章 方差分析		41
4.1	概述	41
4.1.1	基本概念	41
4.1.2	方法	42
4.2	单因素方差分析	42
4.2.1	单因素方差分析基本公式	42
4.2.2	偏差平方和	43
4.2.3	方差分析统计量	44
4.3	无重复两因素方差分析	49
4.3.1	无重复两因素方差分析的数学模型	49
4.3.2	偏差分解	51
4.3.3	自由度	51
4.3.4	方差	52
4.3.5	F -检验	52
4.3.6	方差分析表	53
4.4	有重复两因素方差分析	54
4.4.1	交互作用	55
4.4.2	偏差分解	56
4.4.3	自由度	57
4.4.4	方差	57
4.4.5	F -检验	58
4.4.6	方差分析表	58

第2篇 实验设计与统计应用

第5章 正交实验设计		66
5.1	概述	66
5.1.1	正交表	66
5.1.2	正交表的特点	67
5.1.3	正交表的优点	67
5.1.4	正交表的分类	68
5.2	正交实验设计	68
5.2.1	正交实验设计步骤	69
5.2.2	正交实验的数据处理	69
5.3	多指标的实验	72
5.3.1	综合评分法	72

5.3.2	综合平衡法	72
5.4	有交互作用的设计	74
5.4.1	正交表的选择和表头设计	75
5.4.2	按方案进行实验	76
5.4.3	结果分析	76
5.5	正交实验的方差分析	77
5.5.1	总变差的分解	77
5.5.2	分析方法	77
5.5.3	适应范围	78
5.6	正交表的改造	79
第 6 章	多因素序贯实验设计	82
6.1	最陡坡法	82
6.1.1	实例	82
6.1.2	数学原理	85
6.1.3	应用条件	86
6.2	调优运算和单纯形调优法	88
6.3	消去法	90
6.3.1	“0.618” 法	90
6.3.2	分批试验法	91
6.4	小结	92
第 7 章	随机化区组和拉丁方	93
7.1	随机化区组	93
7.1.1	什么是随机化区组	93
7.1.2	随机化区组设计的数据处理	94
7.1.3	小结	97
7.2	拉丁方	98
7.2.1	概述	98
7.2.2	拉丁方设计的数据处理	99
7.2.3	拉丁方设计举例	99
7.3	正交方	100
第 8 章	析因设计初步	103
8.1	基本原理	103
8.2	析因设计的优点	105
8.3	二因子析因设计	105
8.3.1	简单析因设计	105
8.3.2	统计分析	106
8.3.3	一般的析因设计	107
8.3.4	响应曲线和曲面的拟合	107

8.4	2^k 析因设计	108
8.4.1	简介	108
8.4.2	2^2 设计	109
第9章	响应曲面设计	111
9.1	响应面概述	111
9.2	最速上升法	112
9.3	二阶响应曲面的分析	115
9.3.1	稳定点的位置	115
9.3.2	响应曲面的特征	116
9.3.3	岭系统	119
9.4	响应曲面法实验设计	119
9.4.1	拟合一阶模型的设计	120
9.4.2	拟合二阶模型的设计——中心复合设计	121
9.4.3	二阶响应曲面的 Box-Behnken 设计	124
第10章	回归分析与聚类分析初步	126
10.1	一元线性回归分析	126
10.1.1	概述	126
10.1.2	回归方法	127
10.1.3	一元线性回归方程的检验	129
10.2	二元线性回归	132
10.2.1	回归方法	132
10.2.2	二元线性回归方程的检验	133
10.3	主成分分析	135
10.3.1	概述	135
10.3.2	主成分的计算	136
10.3.3	主成分分析在企业效益中的应用	139
10.4	聚类分析	140
10.4.1	概述	140
10.4.2	样品间的距离	141
10.4.3	聚类方法	143
10.4.4	其他聚类方法	148
第11章	质量控制	150
11.1	分析化学质量控制方法	151
11.1.1	全程序空白值实验值控制	151
11.1.2	标准曲线	151
11.1.3	平行双样	153
11.1.4	加标回收	153
11.1.5	标准参考物的应用	154

11.1.6 方法对照	154
11.2 质量管理图	154
11.2.1 概念	154
11.2.2 质量管理图的绘制与应用	157
11.2.3 控制图的识别	161

第3篇 计算机程序简介

第12章 大型统计软件 SPSS 简介	163
12.1 SPSS 简介	163
12.1.1 SPSS 概述	163
12.1.2 SPSS 界面	164
12.1.3 SPSS 的菜单	165
12.1.4 SPSS 的其他窗口	165
12.1.5 SPSS 的退出	165
12.2 SPSS 基本操作	165
12.2.1 数据的输入	166
12.2.2 统计分析	168
12.2.3 图形分析	169
12.2.4 结果输出	170
12.3 平均值检验	170
12.3.1 均值过程 (分组求均值)	170
12.3.2 两组数据 t -检验	172
12.3.3 配对比较	174
12.4 方差分析	176
12.4.1 单因素方差分析	176
12.4.2 两因素方差分析	176
12.5 正交实验设计设计与方差分析	179
12.5.1 正交实验设计	179
12.5.2 正交实验的方差分析	181
12.6 线性回归	185
12.6.1 一元线性回归	185
12.6.2 二元线性回归	187
12.7 主成分分析	188
12.8 聚类分析	192
12.8.1 K-均值聚类过程	192
12.8.2 系统聚类	195
12.9 质量控制图	201

习题	204
附表	212
附表 1 Nair 检验临界值表	212
附表 2 单侧 Dixon 检验统计量和临界值	212
附表 3 双侧 Dixon 检验统计量和临界值	213
附表 4 Grubbs 检验临界值表	214
附表 5 Cochran 最大方差检验的临界值	214
附表 6 标准正态分布表	215
附表 7 t -分布表 (双侧)	216
附表 8 F -分布表 (单侧)	217
附表 9 相关系数临界值表	220
附表 10 正交表	221
参考文献	232

第 1 篇

数据处理基础

第 1 章 概 述

1.1 本教材的目的

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，学生们往往知道一些理论但却不知如何应用。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、没有运用，所以跟没有学过也差不多，因此他们可能会认为，统计对于化学化工专业纯属可有可无。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，则获得满意结果的唯一稳妥的处理办法便是统计方法，除此以外别无他择。同时我们知道，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。

本教材要求读者已熟悉了一般的统计技术,并将不时引用相应的内容。

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,不常用统计方法。但这种工作需要最现代化的仪器设备,以及对这些设备的熟练操作和细致的工作态度。

(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几次该检验,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以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太明确的性质(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或原材料的变化所起的效应,等等。这类实验可能很费钱,也可能需要许多人的协作,因而必须认真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,他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之潜在原因的工作量大得不可能完成,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。即使所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,所以要使实际过程最后调整到它的最佳条件,仍然须依靠经验法去完成。



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:第一,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;第二,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;第三,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,可从而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,已经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。如果我们在研究工作过程中,以杂乱无章的方式取得结果,则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,和必须采用哪些步骤去探索这些目标。此外,还迫使他们考虑对于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,因为往往能够由此而引导研究人员认识到并从中回避可能的难点和失误。而且要不是这样做,这些难点或失误也可能要到研究工作的后期才被发现,那就晚了。

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,因此往往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,这种随机误差可能相当可观,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,那么原可更好地用之于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的正确数目用的明确判据,这类超需设计或欠需设计就属难免。经验证明,正确的设计类型和采取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而有余。由于任何研究工作的主要开支,几乎总是花在实验设计上,所以实验开支的减少会立即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,需要具备有关的误差和需观测的效应大小的认识。乍看起来这是很困难的,但实践中却极少要求一位研究人员去完成那种带有神秘性质的实验——有关误差的先期资料一无所有。这类资料有时是粗略的和不确定的,即使如此,利用这些资料进行设计,虽然可比资料充分时的设计差一些,但无疑会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,以供今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,帮助他规划好本职工作。

除了早期可资利用的资料外,伴随着工作的进行可以获得新的资料。例如,在化工研究中,各项试验或试制通常是一次一个或一次几个结果,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之。这时,我们可以运用已有的资料,确定实验设计和必要的进一步实验的次数。这种利用一系列观测中的先前部分获得的资料,涉及其后的作

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试验，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：可从实际的实验估计出这些误差。

1.3 本课程将学到或用到的统计研究方法

(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算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，往往使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它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。本章中讨论了因素、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研究中。由于它方便、简洁而得到研究人员的普遍喜爱。

(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，除了可采用上章介绍的正交实验设计以外，更多地应用的是正交设计基础上的序贯实验设计法。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) 析因设计 是一种多因素、多水平、单效应的交叉分组实验设计,又称完全交叉分组实验设计。通过量化各因子及其交互作用对指标的效应,在筛选大量因子研究的初期阶段,析因设计具有明显效果。

(11) 响应曲面法设计 是统计、数学和软件紧密结合和发展的结果,它将响应受多个变量影响的问题进行建模和分析并以此来优化响应。

(12) 大型统计软件——SPSS 应用简介 对于统计分析工作者而言,SPSS——Statistical Product and Service Solutions 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,要掌握它的使用决非易事。本教材中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,就可很快将它掌握。那时可能就会发现,它的好处实在是无与伦比。

第2章 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,而是在一定范围内波动。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,两者同时达到却是不太现实的,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。

2.1.1.1 系统误差

系统误差是由某种确定的因素造成的,是测定结果系统偏高或偏低;当造成误差的因素不存在时,系统误差会自然消失。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。系统误差的最重要的特性,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:

(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,由于沉淀的溶解、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,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,如砝码重量、容量器皿刻度和仪表刻度不准确等。试剂误差来源于试剂不纯。例如,试剂或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