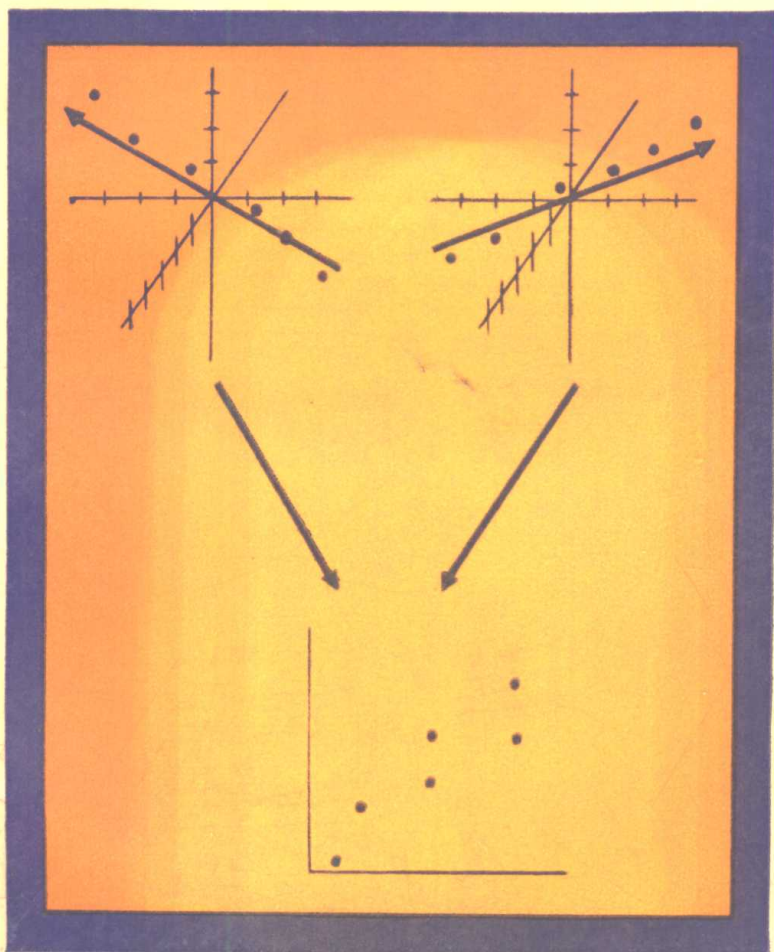


# 化学计量学方法

许禄著



科学出版社

(京)新登字 092 号

## 内 容 简 介

化学计量学是将数学和计算机科学应用于化学的一门新兴的交叉学科,是化学领域的一个重要分支。由于它是建立在多学科基础上的横向学科,所以在解决多种问题中显示了强大的生命力。

全书共分十一章:第一章对误差和数理统计作一扼要介绍;第二、三、五、六和七章为多元统计分析,内容包括回归分析、相关分析、主成分分析、偏最小二乘及模式识别;第四章介绍实验方案设计最优化方法;第八章为构效关系研究中的拓扑指数法;第九至第十一章侧重于非数值方法,如谱图检索、波谱模拟和结构解析中的人工智能等。

本书可供普通化学、生物化学、医学化学、环境化学及药物化学等专业的广大科技工作者和大专院校有关专业师生参考。

## 化 学 计 量 学 方 法

许 禄 著

责任编辑 操时杰

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

中 国 科 学 院 印 刷 厂 印 刷

新华书店北京发行所发行 各地新华书店经售

1995年2月第一版 开本:850×1168 1/32

1995年2月第一次印刷 印张:16 5/8

印数:平1—330 插页:精2

精1—530 总印数:438000

ISBN 7-03-004139-9/O·720(平)

ISBN 7-03-004140-2/O·721(精)

平 装 39.00 元  
定价: 布脊精装 42.00 元

## 前 言

化学计量学 (chemometrics) 是将数学和计算机科学应用于化学的一门新兴的交叉学科,是化学领域的一个重要分支。

数学是自然科学的语言,它在化学中的地位和作用日益突出和重要。特别是自 70 年代以来,随着计算机技术的迅速普及,数学和计算机科学在分析化学中的应用日益广泛,于是化学计量学的方法和内容得到了充实和扩展,使化学计量学成为化学、生物化学、医学化学、环境化学及药物化学中信息处理的强有力手段。1974 年,由美国的 B. R. Kowalski 和瑞典的 S. Wold 等发起,在美国华盛顿大学成立了国际化学计量学学会,开展了一系列的学术交流活 动,推进了化学计量学的迅速发展。从 1982 年起,在美国分析化学杂志 (Anal. Chem.) 两年一度的基础评论中开辟了化学计量学专题。同时,国际上,一些化学计量学专著及系列丛书相继问世。

化学计量学是建立在多学科基础上的横向学科,在解决多种问题中显示了它的强大生命力。目前,国内有众多的科技工作者在从事这一研究,不仅在应用上,而且在理论上也有新的发展。科学研究的实践使我们深深体会到,撰写一本关于化学计量学的专著,系统地介绍有关的理论和方法,以应客观发展之急需是很有意义的。于是,我们在自己实验室工作的基础上,参考新近出版的有关著作及文献写成此书。

全书共十一章。为便于阅读,第一章对误差和数理统计作了扼要的介绍。第二、三、五、六和第七章为多变元统计分析,内容包括回归分析、相关分析、主成分分析和因子分析、偏最小二乘及模式识别,其中偏最小二乘本应归入主成分分析和因子分析一章,但考虑到这种方法吸收了最小二乘法的优点,同时它是主成分回归的

进一步发展,因此我们将其单列一章,以利对该种方法的介绍。第四章介绍实验方案设计最优化方法。关于构效关系的研究有多种方法,其中由于拓扑指数法的简单性和有效性,所以多年来一直受到人们的关注。然而,在国内已出版的专著中介绍较少,故在第八章中侧重论述这一方法。第九至第十一章侧重于非数值方法,如谱图库检索、波谱模拟和结构解析中的人工智能等。

另外,考虑到本书的读者对象主要是化学工作者,故未作数学上的详细推导和证明,只侧重介绍方法的概念和结论,同时给出应用实例,以使方法的描述易读、易懂,在编写程序时亦可作为参考。

本书初稿承蒙中国科学院院士俞汝勤教授审阅,他提出了许多宝贵意见。中国科学院长春应用化学研究所的姚瑜元、胡昌玉、杨翌秋和姜旭宏等同志为本书的编写、定稿做了大量的工作。在此一并表示衷心的感谢。

在本书出版之时,还要特别感谢中国科学院院士汪尔康教授在本书编写过程中所给予的支持、帮助及与清华大学胡鑫尧教授所进行的有益的讨论。

由于水平有限,书中缺点和错误在所难免,敬请读者批评指正。

作者

1993年于中国科学院长春应用化学研究所

# 目 录

|                                   |     |
|-----------------------------------|-----|
| 前言                                | vii |
| 第一章 误差及数理统计基础                     | 1   |
| § 1.1 统计分析中的误差                    | 1   |
| § 1.1.1 均值及偏差                     | 1   |
| § 1.1.2 误差的分布                     | 1   |
| § 1.1.3 均值的置信限 (confidence limit) | 2   |
| § 1.1.4 结果的表示                     | 5   |
| § 1.1.5 置信限的其他应用                  | 6   |
| § 1.1.6 偶然误差的传递                   | 6   |
| § 1.1.7 系统误差的传递                   | 8   |
| § 1.2 显著性检验                       | 9   |
| § 1.2.1 $t$ 检验                    | 9   |
| § 1.2.2 $F$ 检验                    | 13  |
| § 1.2.3 $\chi^2$ 检验               | 16  |
| 参考文献                              | 17  |
| 第二章 回归分析                          | 18  |
| § 2.1 一元回归分析                      | 18  |
| § 2.1.1 一元回归方程的求法                 | 18  |
| § 2.1.2 相关系数和显著性检验                | 22  |
| § 2.1.3 一元线性回归的方差分析               | 24  |
| § 2.1.4 斜率 $b$ 和截距 $a$ 的区间估计      | 25  |
| § 2.1.5 $x$ 值之计算                  | 27  |
| § 2.1.6 标准加入法                     | 29  |
| § 2.1.7 借助回归线进行分析方法的比较            | 30  |
| § 2.1.8 权重回归分析                    | 33  |
| § 2.1.9 曲线回归                      | 36  |

|                           |     |
|---------------------------|-----|
| § 2.2 多元回归分析              | 40  |
| § 2.2.1 多元回归分析方程的求法       | 40  |
| § 2.2.2 多元线性回归的方差分析及显著性检验 | 42  |
| § 2.2.3 多元回归的计算步骤         | 48  |
| § 2.3 逐步回归方法              | 53  |
| § 2.3.1 逐步回归的基本思想         | 53  |
| § 2.3.2 逐步回归的计算步骤         | 57  |
| § 2.3.3 逐步回归的计算例子         | 59  |
| § 2.4 回归分析中几个问题的讨论        | 67  |
| § 2.4.1 变量的评估             | 67  |
| § 2.4.2 回归分析方法            | 73  |
| § 2.4.3 回归模型的评估           | 78  |
| 参考文献                      | 83  |
| <b>第三章 相关分析</b>           | 84  |
| § 3.1 导言                  | 84  |
| § 3.2 协方差和相关系数            | 86  |
| § 3.3 相关和回归               | 88  |
| § 3.4 方差-协方差矩阵            | 89  |
| § 3.5 随机变量的时间序列           | 92  |
| § 3.5.1 自协方差和自相关          | 92  |
| § 3.5.2 具有高相关的时间序列        | 93  |
| § 3.5.3 一些特征过程的自相关谱       | 95  |
| § 3.5.4 实际例子              | 98  |
| 参考文献                      | 100 |
| <b>第四章 最优化方法</b>          | 101 |
| § 4.1 改变单因子法              | 101 |
| § 4.2 单纯形法                | 102 |
| § 4.2.1 基本单纯形法            | 102 |
| § 4.2.2 改良单纯形(变步长)法       | 105 |
| § 4.2.3 改良单纯形法的例子         | 110 |
| § 4.2.4 超改良单纯形法           | 116 |

|                            |            |
|----------------------------|------------|
| § 4.2.5 超改良单纯形计算举例         | 120        |
| § 4.3 响应曲面法                | 125        |
| § 4.3.1 引言                 | 125        |
| § 4.3.2 两因子响应曲面            | 127        |
| § 4.3.3 响应曲面的解释            | 130        |
| § 4.3.4 应用举例               | 135        |
| 参考文献                       | 141        |
| <b>第五章 主成分分析和因子分析</b>      | <b>143</b> |
| § 5.1 主成分分析                | 143        |
| § 5.1.1 两维空间中的主成分分析        | 143        |
| § 5.1.2 $m$ 维空间中的主成分分析     | 146        |
| § 5.1.3 主成分分析的操作步骤、应用和数字例子 | 149        |
| § 5.2 因子分析                 | 154        |
| § 5.2.1 因子分析算法简介           | 154        |
| § 5.2.2 渐进因子分析             | 159        |
| 参考文献                       | 173        |
| <b>第六章 偏最小二乘方法</b>         | <b>174</b> |
| § 6.1 多元线性回归 (MLR)         | 174        |
| § 6.2 主成分回归                | 177        |
| § 6.3 偏最小二乘 (PLS)          | 179        |
| § 6.3.1 基本原理               | 179        |
| § 6.3.2 偏最小二乘算法            | 181        |
| § 6.4 非线性偏最小二乘             | 188        |
| 参考文献                       | 191        |
| <b>第七章 模式识别方法</b>          | <b>192</b> |
| § 7.1 导言                   | 192        |
| § 7.2 数据的表示及预处理            | 192        |
| § 7.3 特征的提取和压缩             | 194        |
| § 7.4 相似系数和距离              | 197        |
| § 7.4.1 相似系数               | 197        |
| § 7.4.2 距离                 | 198        |

|                                    |            |
|------------------------------------|------------|
| § 7.5 模式识别方法                       | 200        |
| § 7.5.1 有管理的方法                     | 200        |
| § 7.5.2 无管理方法                      | 236        |
| § 7.5.3 显示方法                       | 244        |
| § 7.5.4 人工神经网络                     | 251        |
| § 7.5.5 综合性数据例子                    | 262        |
| 参考文献                               | 275        |
| <b>第八章 构效关系研究和分子拓扑指数</b>           | <b>277</b> |
| § 8.1 图论的基本概念                      | 277        |
| § 8.1.1 引言                         | 277        |
| § 8.1.2 图论基础                       | 278        |
| § 8.2 化学结构和图论                      | 280        |
| § 8.2.1 引言                         | 280        |
| § 8.2.2 化学图                        | 281        |
| § 8.2.3 分子的矩阵表示                    | 282        |
| § 8.3 化学结构的二维联接表                   | 285        |
| § 8.4 主要的拓扑指数及其应用                  | 287        |
| § 8.4.1 引言                         | 287        |
| § 8.4.2 距离矩阵指数                     | 288        |
| § 8.4.3 邻接矩阵指数                     | 304        |
| § 8.4.4 中心指数 (center index)        | 323        |
| § 8.4.5 信息论指数                      | 330        |
| § 8.4.6 量子论拓扑指数 $a_N$ 及广义 $a_N$ 指数 | 339        |
| § 8.4.7 几个问题的讨论                    | 349        |
| § 8.4.8 拓扑指数的应用                    | 355        |
| 参考文献                               | 369        |
| <b>第九章 谱图库检索</b>                   | <b>373</b> |
| § 9.1 导言                           | 373        |
| § 9.2 质谱谱图库                        | 373        |
| § 9.2.1 质谱数据源及检索系统                 | 373        |
| § 9.2.2 谱图质量控制                     | 376        |
| § 9.2.3 谱图简化                       | 377        |



|  |            |
|--|------------|
| § 9.2.4 检索算法                               | 379        |
| § 9.3 碳-13 谱图库                             | 387        |
| § 9.3.1 主要的碳-13 检索系统                       | 387        |
| § 9.3.2 检索算法                               | 388        |
| § 9.4 红外光谱谱图库                              | 398        |
| § 9.4.1 红外光谱数据源及一些检索系统                     | 399        |
| § 9.4.2 IR 谱图的简化                           | 401        |
| § 9.4.3 检索方法                               | 405        |
| § 9.5 综合谱图信息                               | 412        |
| § 9.5.1 概况                                 | 413        |
| § 9.5.2 IR 和 MS 的联合运用                      | 413        |
| § 9.5.3 <sup>13</sup> C-NMR 谱与 GC/MS 的联合检索 | 419        |
| § 9.6 与谱图库相关的检索                            | 421        |
| § 9.6.1 结构检索                               | 422        |
| § 9.6.2 化学名称检索                             | 425        |
| § 9.6.3 杂原子检索                              | 426        |
| § 9.7 文件组织                                 | 427        |
| 参考文献                                       | 431        |
| <b>第十章 碳-13NMR 波谱模拟</b>                    | <b>433</b> |
| § 10.1 导言                                  | 433        |
| § 10.2 分子力学模型的建立                           | 435        |
| § 10.2.1 势能函数                              | 435        |
| § 10.2.2 MM2, MMP2 程序简介                    | 439        |
| § 10.2.3 实例                                | 441        |
| § 10.3 分子结构特征的提取                           | 442        |
| § 10.4 化学环境相似度的计算                          | 444        |
| § 10.5 数学模型的选取                             | 452        |
| § 10.6 波谱模拟                                | 458        |
| 参考文献                                       | 464        |
| <b>第十一章 谱图解析中的专家系统</b>                     | <b>465</b> |
| § 11.1 专家系统                                | 465        |

|          |                                 |     |
|----------|---------------------------------|-----|
| § 11.1.1 | 专家系统的基本结构 .....                 | 465 |
| § 11.1.2 | 硬件和专家系统 .....                   | 466 |
| § 11.1.3 | 软件和专家系统 .....                   | 467 |
| § 11.1.4 | 专家系统的研制 .....                   | 469 |
| § 11.1.5 | 专家系统在化学中的应用 .....               | 470 |
| § 11.1.6 | 专家系统的发展趋势 .....                 | 472 |
| § 11.2   | 谱图解析专家系统 .....                  | 474 |
| § 11.2.1 | 谱图解析专家系统的工作过程 .....             | 474 |
| § 11.2.2 | 知识库 .....                       | 475 |
| § 11.2.3 | 子结构-子光谱相关性统计 .....              | 478 |
| § 11.2.4 | 整体结构对接 .....                    | 486 |
| § 11.2.5 | 候选化合物的确证 .....                  | 492 |
| § 11.3   | 一个分布式专家系统的介绍 .....              | 494 |
| § 11.3.1 | 系统的设计和框图 .....                  | 494 |
| § 11.3.2 | 光谱数据专家 .....                    | 495 |
| § 11.3.3 | 控制器 (controller reasoner) ..... | 499 |
| § 11.3.4 | 知识库 (library expert) .....      | 499 |
|          | 参考文献 .....                      | 500 |
|          | 附录 .....                        | 502 |
|          | 主题索引 .....                      | 520 |

# 第一章 误差及数理统计基础

因本书中有不少章节要用到数理统计方面的知识，所以在这一章首先概略地介绍数理统计的基本概念，以利于其他章节的讨论。

## § 1.1 统计分析中的误差<sup>[1-3]</sup>

### § 1.1.1 均值及偏差

用于测定值的比较，其判据主要有两个：均值和数据的离散度。均值的表达式为

$$\bar{x} = \left( \sum_{i=1}^n x_i \right) / n \quad (1.1)$$

用于离散度的量度主要是标准偏差  $s$ ：

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (1.2)$$

$s$  的另一表达式为

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n - 1} - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n(n - 1)}} \quad (1.3)$$

在统计分析中，常常用到的是  $s^2$ ，即标准方差。

### § 1.1.2 误差的分布

标准偏差可以表征测定结果相对于均值的离散程度，但却不能指示这些数据的分布情况。而表征数据的分布情况要用直方图（或频谱图）。如对某一溶液作 50 次测定，其均值为  $0.50\mu\text{g}/\text{ml}$ 。其

中,  $0.46\mu\text{g/ml}$  出现 1 次,  $0.47\mu\text{g/ml}$  出现 3 次,  $0.48\mu\text{g/ml}$  出现 5 次,  $0.49\mu\text{g/ml}$  出现 10 次, 等等. 将每一测定值出现的频率对测定值作图即为直方图(或频谱图).

若对某样本作无限多次测定, 所得数据称为该样本的总体(population), 总体的均值常用  $\mu$  表示. 若无系统误差, 称为该样本的真值. 事实上不可能作无限次测定. 若作  $n$  次测定, 其均值为  $\bar{x}$ ,  $\bar{x}$  是  $\mu$  的估计. 同样, 若总体的标准偏差为  $\sigma$ , 有限次(如  $n$  次)测定的标准偏差为  $s$ , 则  $s$  为  $\sigma$  的估计. 当  $N$  趋于无穷大时,  $s$  将趋近于  $\sigma$ .

在数学上, 常用正态分布或高斯分布来描述某样本的总体:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1.4)$$

其中,  $x$  为样本测量值,  $y$  为测量值的概率密度.

正态(或高斯)分布具有如下重要性质:

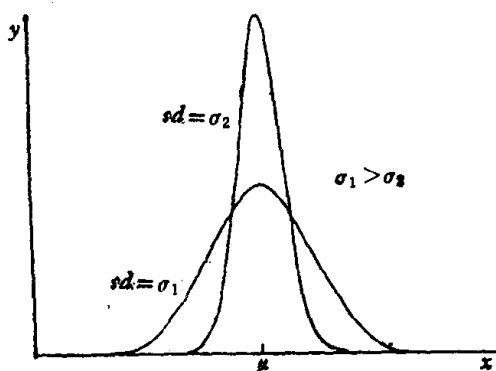


图 1.1 均值相同, 标准偏差不同的正态分布  
 $sd$  为标准偏差.

(1) 数据关于  $\mu$  为对称分布;

(2)  $\sigma$  值越大, 数据的离散程度越大(图 1.1);

(3) 分布在  $\mu \pm \sigma$  的范围内, 约为总体的 68%; 在  $\mu \pm 2\sigma$  的范围内约为总体的 95%; 在  $\mu \pm 3\sigma$  的范围内约为总体的 99.7% (图 1.2).

在分析化学中, 绝大部分情况下其测定符合正态分布.

### § 1.1.3 均值的置信限 (confidence limit)

根据正态分布的性质我们可以对真值定义一范围, 该范围称为置信限 (confidence limit). “置信限”的意思是, 当一置信度 (confidence degree) 即一确定的概率被指定之后, 则真值将落

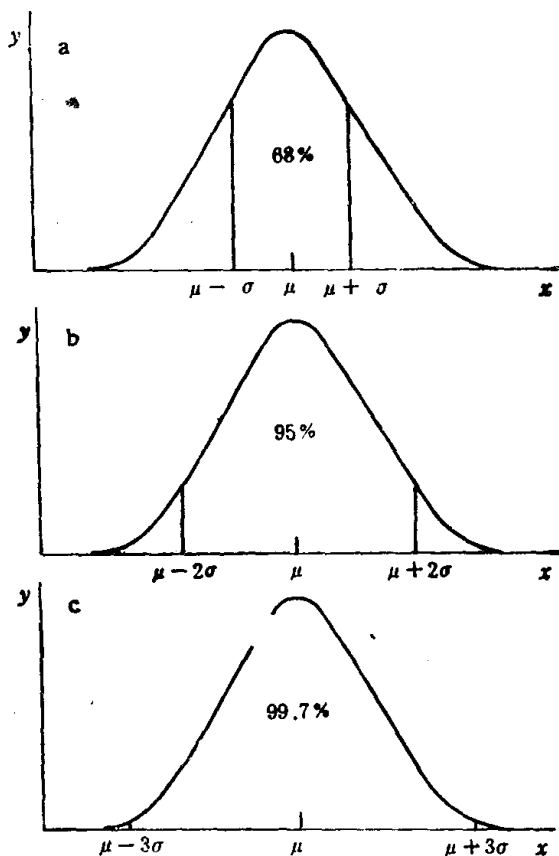


图 1.2 正态分布的性质

在置信区间之内。置信区间的大小依赖于所指定的确定性 (certainty), 也就是说, 确定性越大, 所需要的置信区间也越大。

对于一样本, 可分  $M$  次采样。相应于第  $m_i (i = 1, 2, \dots, M)$  次采样, 有  $n_i$  个测定数据, 这  $n_i$  个测定数据的均值为  $\bar{x}_i$ 。  $\bar{x}_i$  的分布称为采样均值分布 (sampling distribution of the mean)。若  $\sum_{i=1}^M n_i = n$ , 则  $\bar{x}_i$  的均值和总体的单次测定值相同, 但其标准偏差 (此时称为均值的标准误差——standard error of the mean, 简称为 s. e. m.) 为

$$\text{s.e.m.} = \frac{\sigma}{\sqrt{n}} \quad (1.5)$$

那么, 采样均值(以  $\bar{x}$  表示) 95% 的范围为

$$\mu - 1.96(\sigma/\sqrt{n}) < \bar{x} < \mu + 1.96(\sigma/\sqrt{n}) \quad (1.6)$$

请注意, 式中用的为准确值 1.96, 而不是近似值 2. 式(1.6)可改写为:

$$\bar{x} - 1.96(\sigma/\sqrt{n}) < \mu < \bar{x} + 1.96(\sigma/\sqrt{n}) \quad (1.7)$$

式(1.7)给出 95% 的均值置信区间. 相类似, 99.7% 的置信区间由下式给出:

$$\bar{x} - 2.97(\sigma/\sqrt{n}) < \mu < \bar{x} + 2.97(\sigma/\sqrt{n}) \quad (1.8)$$

通常, 我们习惯于采用 99% 的置信区间:

$$\bar{x} - 2.58(\sigma/\sqrt{n}) < \mu < \bar{x} + 2.58(\sigma/\sqrt{n}) \quad (1.9)$$

事实上, 我们并不知道  $\sigma$ . 但当测定次数  $n$  足够大时, 在计算中常用标准偏差  $s$  代替  $\sigma$ . 如某溶液中硝酸根离子 50 次测量的均值为  $0.500\mu\text{g/ml}$ ,  $s = 0.00165\mu\text{g/ml}$ , 则其 95% 的置信区间为:

$$0.500 - 1.96 \times 0.00165/\sqrt{50} < \mu < 0.500 \\ + 1.96 \times 0.00165/\sqrt{50}$$

其置信限为:

$$\mu = 0.500 \pm 0.00046\mu\text{g/ml}$$

当  $n$  不够大时, 则由  $s$  代替  $\sigma$  所引进的误差将较大. 此时计算置信限可用下式表示:

$$\mu = \bar{x} \pm t(s/\sqrt{n}) \quad (1.10)$$

式中  $t$  (即  $t$  分布) 依赖于自由度  $(n-1)$ . 自由度为计算  $s$  中独立偏差  $(x_i - \bar{x})$  的个数, 在此种情况下为  $(n-1)$ . 因为

$$\sum_{i=1}^n (x_i - \bar{x}) = 0,$$

所以只要  $(n-1)$  个  $(x_i - \bar{x})$  为已知, 则第  $n$  个  $(x_i - \bar{x})$  可由上式计算出. 某些  $t$  值列于表 1.1.

由表 1.1 易于看出, 当  $n$  大于 50 时,  $t$  将非常接近于式(1.7)中的 1.96(对应于 95% 的置信区间)和式(1.9)中的 2.58(对于 99%

表 1.1 相应于某些置信区间的  $t$  值

| 自由度 | $t$ 值     |           |
|-----|-----------|-----------|
|     | 置信区间, 95% | 置信区间, 99% |
| 1   | 12.71     | 63.66     |
| 2   | 4.30      | 9.92      |
| 3   | 3.18      | 5.84      |
| 4   | 2.78      | 4.60      |
| 5   | 2.57      | 4.03      |
| 10  | 2.23      | 3.17      |
| 20  | 2.09      | 2.85      |
| 30  | 2.04      | 2.75      |
| 50  | 2.01      | 2.68      |
| 100 | 1.98      | 2.63      |

的置信区间)。这就证明了在上述硝酸根离子浓度计算中所作假设(即用  $s$  替代  $\sigma$ ) 的正确性。

如用离子选择性电极进行了尿中钠离子的测量, 结果为: 102, 97, 98, 99, 101, 106mmol. 试分别计算 95% 和 99% 的置信限。

自由度 = 6 - 1 = 5, 由表 1.1 可知对应于 95% 和 99% 置信限的  $t$  值分别为 2.57 和 4.03。6 次测量均值为 100.5mmol, 标准偏差为 3.27, 故得:

$$\begin{aligned} \mu &= 100.5 \pm 2.57 \times 3.27 / \sqrt{6} \\ &= 100.5 \pm 3.4 \text{mmol (95% 置信限)} \end{aligned}$$

和

$$\begin{aligned} \mu &= 100.5 \pm 4.03 \times 3.27 / \sqrt{6} \\ &= 100.5 \pm 5.4 \text{mmol (99% 置信限)} \end{aligned}$$

#### § 1.1.4 结果的表示

测量结果最常用的表示方式是均值和标准偏差。前者表征测试量的大小, 后者表征测试的精密度。

与之有关的是有效位的取舍。所谓有效位是指某种测量所达到的精度。如下列测试值: 10.09, 10.11, 10.09, 10.10 和 10.12, 其

均值为 10.102, 标准偏差为 0.0130。但测试值仅准确到小数点后面第一位, 而第二位为可疑位, 故结果的表示为:

$$\bar{x} \pm s = 10.10 \pm 0.01 \quad (N = 5)$$

但也有人建议表示为:

$$\bar{x} \pm s = 10.10_2 \pm 0.01, \quad (N = 5)$$

其下脚为避免信息的丢失而加。

另外, 对于小数点后面数字的取舍一般遵循“四舍五入”的规则。但有人建议, “5”的入或舍, 应使它前一位数成为与“5”最近的偶数。如 9.65 应为 9.6, 而 9.75 应为 9.8, 余类推。

### § 1.1.5 置信限的其他应用

置信限可以用于系统误差的测试。如一分光光度计对某标准溶液在某一波长处测试, 其吸收值应为 0.470。现在进行 9 次测定, 均值  $\bar{x} = 0.461$ , 标准偏差  $s = 0.003$ 。置信区间若为 95%, 则

$$\begin{aligned} \mu &= \bar{x} \pm t(s/\sqrt{n}) \\ &= 0.461 \pm 2.26 \times 0.003/\sqrt{9} \\ &= 0.461 \pm 0.002 \end{aligned}$$

由于 0.470 并不落在所得置信区间范围内, 所以仪器有系统误差。

置信限还可用于试样的测定。如有一大批药物的片剂, 欲知片剂的重量, 则不可能对每片一一称重。另外, 若想知道片剂的组分和含量, 特别是采用破坏性分析方法, 如原子吸收, 则亦不可能对每片药物进行分析。在这种情况下, 可从中取出试样, 测得均值和标准偏差, 继而得到测定量的置信区间。

### § 1.1.6 偶然误差的传递

#### 1. 线性组合

如  $y$  为测量量  $a, b$  和  $c$  等的线性组合:

$$y = K + K_a a + K_b b + K_c c + \dots$$

式中  $K_a, K_b$  和  $K_c$  等为常数, 则加和或差值的方差 (即标准偏差



的平方)是各量方差的加和:

$$\sigma_y = \sqrt{(K_a\sigma_a)^2 + (K_b\sigma_b)^2 + (K_c\sigma_c)^2 + \dots}$$

如滴定中,移液管的初值和终值分别为 3.51 ml 和 15.67 ml, 其标准偏差均为 0.02ml, 则用去滴定液的体积及标准偏差分别为:

$$\text{消耗的滴定液体积} = 15.67 - 3.51 = 12.16(\text{ml})$$

$$\text{标准偏差} = \sqrt{(0.02)^2 + (0.02)^2} = 0.028(\text{ml})$$

此例说明,组合的标准偏差大于单个读数的标准偏差,但小于各量标准偏差之和。

## 2. 乘除表达式

若计算  $y$  的表达式为:

$$y = kab/cd$$

式中  $a, b, c$  和  $d$  分别为测量量,  $k$  为常数, 则相对标准偏差有如下关系:

$$\frac{\sigma_y}{y} = \sqrt{\left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2 + \left(\frac{\sigma_c}{c}\right)^2 + \left(\frac{\sigma_d}{d}\right)^2}$$

如荧光的量子产率可用下式计算:

$$\Phi = I_f/kcLI_0e$$

式中各量的相对标准偏差是:

$I_0$  为入射光强度, 0.5%;

$I_f$  为荧光强度, 2%;

$e$  为摩尔吸收, 1%;

$c$  为浓度, 0.2%;

$L$  为液槽长度, 0.2%;

$k$  为仪器常数。

则  $\Phi$  的相对标准偏差为:

$$rs_d = \sqrt{2^2 + 0.2^2 + 0.2^2 + 0.5^2 + 1^2} = 2.3(\%)$$

由此可见, 最终结果的相对标准偏差略大于上述分量中具有最大相对标准偏差的那个分量 ( $I_f$ )。这一结果给我们的暗示是, 若拟