

理化分析数据处理手册

漆德瑶 肖明耀 吴芯蕊 编著

中国计量出版社

内 容 提 要

本书介绍数理统计方法在理化分析测试中的应用。全书共分十二章,其中包括:测量数据的位数、可疑数据的取舍、平行多次测定分析结果的评定,两组及多组分析数据平均值和精度的评定,实验室之间分析数据的评定,分析测试中的标准曲线和质量控制图,各分析步骤的误差对分析结果的影响,几种简便的统计方法。为了使广大读者能正确掌握和应用数理统计方法解决具体问题,本书列举了129个例题,并附统计用表43个。在附录中还对数理统计中若干重要的基本概念作了简明叙述。

本书可供在分析、计量、测试、检验、质量管理等部门工作的科技人员参考应用,也可作理工科大学及有关培训班的学习教材。

理化分析数据处理手册

漆德瑞 肖明耀 吴芯芯 编著

责任编辑 栾桂芬

—*

中国计量出版社出版

北京和平里西街甲2号

中国计量出版社印刷厂印刷

新华书店北京发行所发行

—*

开本 850×1168/32 印张 11.5 字数 302 千字

1990年4月第1版 1990年4月第1次印刷

印数 1—5 000

ISBN 7-5026-0294-1/TB·244

定价 5.50 元

前 言

在分析测试中，由于分析方法和分析仪器不可能绝对准确，还有分析操作和环境条件等方面的原因，使得分析测试的结果总是带有误差。对于如何正确表达这种含有误差的分析结果，如何评定和比较分析结果以及现有的分析方法，如何从这有限的测量数据中获得更多的有用的信息等，就成为分析测试工作者非常关心的问题。特别是随着分析化学学科的迅速发展，分析数据的获得愈来愈快速，对分析测试准确度的要求也愈来愈高，因此，目前分析工作者，包括应用分析测试结果的科技人员，对了解以上所提的一些问题要求更为迫切。显然，要回答这些问题，除了需要本专业的知识以外，还需借助于数理统计的方法。

20 世纪初期，数理统计已应用于生物学和经济学中，至 50 年代，数理统计才开始应用于化学分析，在分析化学家和数理统计学家的共同配合和努力下，到 60 年代初已逐渐形成了一门系统的新的学科。目前化学计量学，图象认识，分析化学智能化等方面的新成就和这门学科的发展都有着密切关系。

至今为止，在国内已出版了许多种数理统计方法的参考书，有的还专门介绍了数理统计在分析化学中的应用。这些书籍在我国对推广在分析化学中应用数理统计方法都起到了很好的推广和促进作用。本人从 80 年代开始先后应上海市的测试学会、中华药学会、化学化工学会以及商品检验局、医药工业研究院，轻工业研究所等单位的邀请，办了 6 期讲习班，作了有关分析误差和分析数据处理的讲座，许多学员曾提出由于有些数理统计和应用方面

的参考书，都较系统地或较多地叙述和论证了数理统计的原理，推导了公式，章节也按数理统计学的要求来编排，这样对未学过高等数学和数理统计的工作人员，在短时间内希望较快地解决他遇到的有关数据处理问题感到困难和不便。因此他们希望能从解决实际问题出发，介绍数理统计的方法，使理化分析专业人员能较快地掌握并应用这一方法。根据这一指导思想，作者编写了以上讲习班用的讲义。并通过听取每期讲习班学员的意见，作了多次修改，不少学员希望能有机会将该讲义作适当修改补充后出版。现承中国计量出版社大力支持和热情鼓励，在原讲义的基础上，和多年从事数理统计工作的同志一起，共同编写并出版了此书。在此，本书作者向中国计量出版社深表谢意。

本书以实用为主，对数理统计学的原理不作系统介绍，也略去了数学公式的推导和论证，只介绍具体的分析测试数据的处理方法，各种方法后面还附有计算实例。为了查阅方便，全书章节均按分析测试工作者常遇到的一些有关数据处理的具体问题进行编写。由于仪器分析的使用日益增多，故在第九章中对标准曲线的绘制和评定叙述较多，为了使用方便，避免繁复的数学计算，本书最后一章还介绍了几种简便的统计检验方法。考虑到有的读者没有系统学习过数理统计学，故在本书附录中对数理统计中若干基本概念作了简要叙述。

我们希望这本书的出版能对我国广大从事分析测试的科技人员在实际工作中有所帮助，并欢迎专家和广大读者对本书不足之处提出宝贵意见。

漆德瑞

1988. 7. 28.

于上海工业大学

目 录

第一章 测量数据的位数	(1)
§ 1 有效数字	(1)
§ 2 有效数字计算	(3)
§ 3 分析结果数据的位数	(5)
第二章 可疑数据的取舍	(7)
§ 1 可疑数据及处理原则	(7)
§ 2 偏大或偏小的可疑数据的处理方法	(8)
§ 3 偏大和偏小的可疑数据同时出现的处理方法	(21)
第三章 平行多次测定分析结果的评定	(32)
§ 1 分析结果的可信范围	(32)
§ 2 分析结果的偏离及其评定	(44)
§ 3 标准偏差的评定	(48)
§ 4 测定次数的确定	(54)
第四章 两组分析数据平均值的评定	(56)
§ 1 用选定的标准来评定一组分析数据平均值的准确度	(57)
§ 2 两组分析结果的评定	(65)
§ 3 两组成对数据所得平均值 $\bar{x}_1$ 及 $\bar{x}_2$ 的评定	(82)
§ 4 秩和检验法	(84)
§ 5 分析次数的选择	(92)
第五章 多组分析数据平均值的评定	(106)
§ 1 在多组分析数据中无对比组时的检验	(106)
§ 2 在多组分析数据中有对比组时的检验	(114)
§ 3 秩和检验法	(117)
§ 4 多组分析数据平均值中可疑值的检验	(120)

第六章	两组及多组分析数据精度的评定	(124)
§ 1	分析数据的精度表示法	(124)
§ 2	一组分析数据精度的评定	(127)
§ 3	两组分析数据精度的评定	(134)
§ 4	评定两组分析数据精度时测定次数的确定	(135)
§ 5	多组分析数据精度的评定	(139)
第七章	实验室之间分析数据的评定 (一)	(148)
§ 1	实验室之间用同一分析方法得到的分析结果的评定	(149)
§ 2	重复性及重现性的表示法	(164)
§ 3	个别分析数据缺落时的补救方法	(166)
§ 4	用成对分析数据评定实验室之间的分析结果	(169)
§ 5	分析数据的单因素方差分析步骤	(177)
第八章	实验室之间分析数据的评定 (二)	(182)
§ 1	实验室之间分析含量不等的试样时分析数据的评定	(182)
§ 2	实验室和试样含量的交互影响的评定	(189)
§ 3	用两因素方差分析评定分析数据时的几点假设及验证	(194)
§ 4	重复性和重现性的评定	(198)
§ 5	两因素方差分析法的几个应用实例	(201)
第九章	分析测试中的标准曲线	(211)
§ 1	标准曲线呈直线的评定	(211)
§ 2	标准曲线的评定	(220)
§ 3	标准曲线的斜率与截距的可信范围	(225)
§ 4	根据标准曲线所求得的分析结果的可信范围	(230)
§ 5	标准曲线的比较	(236)
§ 6	标准曲线平移的检验	(245)
§ 7	标准曲线的几种应用	(248)
§ 8	分析结果的精度和欲测物含量有关时的标准曲线	(258)
§ 9	不呈直线的标准曲线	(267)
第十章	分析测试中质量控制图的绘制	(276)
§ 1	分析结果控制图	(276)
§ 2	根据分析数据的最大差值作分析结果控制图	(279)
§ 3	分析精度控制图	(283)
§ 4	分析结果范围控制图	(285)

第十一章 各分析步骤的误差对分析结果的影响	(288)
§ 1 各操作步骤的系统误差对分析结果的影响	(288)
§ 2 各分析步骤的随机误差对分析结果的影响	(290)
§ 3 误差传播定律的应用实例	(292)
第十二章 几种简便的统计检验方法	(306)
§ 1 平均值的简便估计	(306)
§ 2 成对数据的符号检验法	(307)
§ 3 瓦尔特-沃福威兹 (Wald-Wolfowitz) 检验法	(309)
§ 4 极差检验法	(311)
§ 5 威尔科克松 (Wilcoxon) 符号秩和检验法	(316)
§ 6 威尔科克松 (Wilcoxon) 秩和检验和曼-惠特尼 (Mann-Whitney) U -检验	(319)
§ 7 秩号相关检验法	(322)
§ 8 简便线性回归法 (塞尔法-Theil's Method)	(324)
§ 9 拟合优度检验 (科尔莫戈罗夫法-Kolmogorov Test)	(327)
附录	(333)
A. 统计学基本概念	(333)
B. 自由度	(354)
主要参考文献	(358)

第一章 测量数据的位数

在分析测试中，直接测量需要记录测量值；间接测定需要对有关测量值进行计算，最后得到分析测试的结果。这里都遇到用几位数字来表示这些测量值或分析结果的问题。应该注意，多写一位或少写一位数字都是不妥的。

§1 有效数字

能够代表一定的物理量的数字，称为有效数字。在分析测试中所记录的数字都应该是有效数字。但由于测量仪表刻度的限制，最后一位数字往往需要估计才能读出，因此测量数据的最后一位数字是可疑数字，例如：用万分之一的天平所称得的某试样重 1.256 3 g，其中 1.256 g 都是从所加的砝码标值直接读得的，它们都代表了一定的物理量（重量），而最后一位数字“3”则是从天平刻度上估计得到的，不同的观察者或同一观察者在不同时间，它可能被读成“4”或读成“2”，因此最后这位数字“3”是一个可疑数字。由此可见，从一个正确表示的测量数据的位数，可以了解该测量的精度或用作测量该数据的仪表的精度。例如，“某试样重 1.256 g”，我们就知道称量该试样的精度为千分之一克，称量时用的天平的精度为千分之一克。如果某分析操作规程中写明称取某化学试剂 1.2 g，那么我们就只需用一般实验室用的台秤来称量，而不需要用分析天平去称量。

由上可知，在分析测试、检验，计量等工作中，正确表达测

量数据的位数，非常重要。不正确地多写了一位数字，则该数据不真实，因而也不可靠；少写了一位数字，则损失了测量精度，实质上对测量该数据所用的精度高的仪表和耗费的时间也是浪费。

用有效数字表示的测量数据，小数点的位置不影响测量数据有效数字的位数，例如，20.25 毫升和 0.020 25 升，它们的有效数字都是四位。

零是否算一个有效数字，要看它在测量数据中的位置。在数据前面的零都不是有效数字，在数据后面的零都是有效数字。因此，0.010 20 升，这个测量数据，它的有效数字是四位，而不是六位或三位。和算术中数字的表示方法不一样，测量数据中小数点后最后的数字零不能略去，如 1.000 0 克，不能简写成 1 克。同样，也不能在测量数据的最后随意加一个或几个零。

因此，一些在数字最前面或最后面的没有意义的零，也就是那些不是有效数字的零，在按照有效数字规则表达时，均应略去。为此，0.000 134，应写成 1.34×10^{-4} ；而对于“965 000”这个数字而言。如果测量精度为千分之一，则应写成 965×10^3 ，此时其中的“5”为可疑数字；如果测量精度为万分之一，则应写成 9650×10^2 ，此时其中的“0”为可疑数字；如果写成 965 000，那么它后面的三个“0”都是有效数字，仅最后一个“0”是可疑数字，此时它的测量精度为百万分之一。

值得指出的是人为指定的标准值，它的小数点后最后的零，可以根据需要进行增减，典型的例子是目前原子量的相对标准值为 $^{12}\text{C} = 12$ ，用它来确定其它元素的原子量时，它的有效数字的位数不是只有两位，在 12 加一小数点后可以增写零，增写零的个数视目前能获得的测定某元素相对质量的精度而定。例如：氟的原子量可测准至小数点后第六位（1975 年），因此可以把 ^{12}C 的原子量的相对标准看成是 12.000 000，一共是八位有效数字。同样，在分析测试的计算式中所用到的倍数数值，也和人为

指定的标准值一样，它的有效数字的位数也可以视计算时需要而设定。

§ 2 有效数字计算

在分析测试中往往要用各种单一测量值进行计算，以求得最后的分析测试结果。为了正确表示最后的结果和节约计算时间，在计算过程中应注意合理取用各数据的有效数字的位数。应用加、减和乘、除法进行计算时，各数据的有效数字的位数按以下方法确定：

一、加、减法计算

在加、减法计算中，各个数据取用的位数应以小数点后的有效数字位数最少的一个数为准，其它小数点后的有效数字的位数比它多的各数据，包括最后所计算得的结果均和它对齐。例如：
 $1.23 + 10.562 + 0.1486$ ，计算方法：

$$\begin{array}{r} 1.23 \\ 10.562 \\ + 0.1486 \\ \hline 11.9406 \end{array}$$

最后结果应为 11.94。

又如： $23 \times 10^4 + 12$ ，其结果应写为 23×10^4 。

减法也是一样。例如： $21.5780 - 3.142 - 0.56 = 17.88$ 。

二、乘、除法计算

在乘、除法计算中，各个数据取用的位数应以测量相对精度最差的一个数为准，最后的计算结果的位数也应根据该数据的测量相对精度来确定。例如： 6×32.47 ，其中 6 为倍数，32.47 这

个数据的相对精度约为三千分之一。因此，最后计算结果应为194.8，而不能写成194.82，后一数据的相对精度约为二万分之一。同理， $1.2030 \times 0.01963 \times 2.83$ ，计算结果为0.0668。这里也可以写成 $1.20 \times 0.0196 \times 2.83$ ，其计算结果为0.0666，两种计算结果之间的相对误差约为千分之三，而在乘数中2.83这一有效数字的相对误差也约为千分之三，两者的精度是一致的，在这种情况下，为减少手算的工作量，可以用后一式子计算。又如： 1256×12.2 ，其结果应写为 1.53×10^4 。

除法的计算方法及结果表示方法也和乘法一样。例如： $20.35 \div 0.163$ ，其结果应写为125。

在计算过程中，为了避免损失精度，最好多保留一位甚至两位数字。但计算得的结果，其相对精度仍应与乘、除算式中相对精度最差的一个数据的相对精度相符。

由上可知，在有效数字计算过程中，经常会遇到需要舍去多余的数字。人们习惯用“四舍五入”法来舍去多余数字，但这一方法从统计学观点来看不合理，因在1到9的九个数字中有四个数字（1—4）属于舍去的，而有五个数字（5—9）属于进位的，即出现舍去的几率为九分之四，而出现进位的几率为九分之五。由于这原因，故现在规定末位为4和4以下的数字舍去，末位为6和6以上的数字进位。遇到末位为5这个数字，则需看它前面的一个数字，如果它前面的数字是单数，则进一位，如果是双数（包括零）则舍去，这样，数字进位或舍去的出现几率均相等。如果遇到5这个数字，它后面还有其它数字，则进一位，但应注意不能从最后一位数字开始连续进行取舍。

为了清楚起见，举例如下：对1.2534这个数字，如只需取四位，则为1.253；对1.2536，则写为1.254；1.2535写为1.254；1.2545写为1.254；1.25451写为1.255；1.25349应写为1.253，而不能先写成1.2535，再写成1.254。这一多余数字舍去的修约规则可简称为“四舍六入尾留双”数字修约规则。

§ 3 分析结果数据的位数

化学分析的结果往往是通过多次单独测量而求得。每次测量的数据，它的有效数字的位数由测量精度决定，但各次的测量精度可能不相同，因而它们的有效数字的位数也不等。此时就要按照上节所述的有效数字的计算方法进行计算，最后计算得的分析结果的位数应和各次测量中相对精度最差的一位数据的位数相符。

已知感度为万分之一的天平，如称 1 克以上的试样，最少可以得到五位有效数字。故用重量法测定试样中的某组份，如在操作步骤中仅经过两次测量，即用分析天平称取试样，最后又用分析天平称量所得的沉淀的重量，假如它们的重量均大于 1 克，那末，最后计算得的分析结果可以有五位有效数字。故经典的重量分析法，到目前为止仍被认为是准确度最高的一种化学分析方法。如果试样的重量在 1 克以上，但最后所得的沉淀的重量为 $0.0 \times \times \times$ 克，那末，所得的分析结果，它的有效数字只能写三位。

在容量分析中主要用到滴定管，它的读数的有效数字最多是四位，故即使用分析天平称取 1 克以上的试样，但最后得到的分析结果，它的有效数字只有四位。如果滴定时用去的标液不到 10 毫升，或操作过程中用到移液管，所量取的体积小于 10 毫升，那末，分析结果的有效数字只能写三位。

在仪器分析中，测量用的仪表可读得的有效数字，往往最多只有三位。故用仪器分析所得到的分析结果，它的有效数字一般都只有三位。即使称量试样时用分析天平可得五位有效数字的数据，但由于最后测量时，分析仪器仪表的读数只有三位，故分析结果的有效数字仍只有三位。

由上可知，在化学分析中，各次测量的精度应保持一致。如果在分析操作过程中，有一次操作，它的测量精度特别低，那末

不管其它各次的测量精度如何高，其最后所得的分析结果的精度只能是和测量精度最低的那次操作的精度相同，显然，此时其它各次高精度的测量就变得没有必要，对仪器、人力和时间都是浪费。

从多次分析结果求平均值 ($\bar{x}$) 时，通过用下式计算：

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

其中 n 为分析结果的个数， x_i 为第 i 次分析结果。显然，在各次分析结果相加时，应采用上节所述有效数字相加的计算方法，但当 n 较大时，平均值的有效位数可以增加一位。

第二章 可疑数据的取舍

§ 1 可疑数据及处理原则

在相同条件下进行多次重复分析测试时，可以得到一组平行数据。在这组数据中有时会发现，有个别的数据，其数值大小明显偏离其它大多数数据，但又找不到产生偏离的确切原因，这类数据就称为可疑数据。

对可疑数据的取舍一定要慎重，因为该可疑数据如不属于异常值，若将它舍去，则表观上提高了精度，而实质上降低了平均值的准确度。如该可疑数据本身就是异常值，但没有将它舍去，那末降低了测量精度，同时所求的结果也不可靠。

对明显偏离的数据，通常有以下一些处理原则：

1. 在分析测试过程中，尚未得出最后结果以前，如已发现操作上的过失或测量仪表的不正常，或者其它的明显错误，它们均将导致最后测试结果的不可靠。那末，一般应及时停止分析测试操作，没有必要再继续完成操作步骤与计算分析结果，更不能没有根据地对所得到的分析结果任意进行校正，使它接近其它大多数数据。

2. 当分析测试过程已结束，并计算得分 析 结果，但发现该数据与其它大多数数据偏离。此时，首先应尽量寻找产生该数据偏离的原因。如已找到它的确切原因，那末该数据是异常值应舍去。对一个有经验的分析工作者，或对某一分析测试方法很熟悉

的工作者来说，这种引起分析结果偏离的原因是不难找到的。

3. 如出现可疑数据，当分析测试次数较少，即数据个数不多时，则应重复再作一次或几次分析测试，取得较多的数据，以便最后对该数据进行统计处理，较正确地判断对它的取舍。如条件许可，数据个数又不多，则最好将存在可疑数据的这一组数据全部舍去，再重新作一组数据。

4. 对舍去的任何数据，在报告中仍应将该数据写上，同时应说明舍去的原因或者所选用的处理可疑数据的统计方法。

总之，对待可疑数据要慎重，不能随意抛弃。首先要寻找产生偏离的原因。往往通过对可疑数据的考察，可以发现引起误差的原因，因而改进该分析方法，有时甚至可通过它，得到新的分析测试方法的线索。

在分析测试中，将一组平行测得的数据从小到大按顺序排列，可以发现，可疑数据的出现通常有以下两种情况：一种是一个或极少数几个数据和其它大多数数据比较，明显偏高或偏低；另一种是和其它大多数数据比较，有两个数据，一个明显偏高，同时还有一个明显偏低。以下就这两种情况，分别介绍运用统计学进行处理的方法。

§ 2 偏大或偏小的可疑数据的处理方法

将测得的一组 n 个数据，按大小顺序排列： $x_1, x_2 \cdots x_n$ ，其中 x_1 为可能出现的偏小数据， x_n 为可能出现的偏大的数据。

一、“4 d” 检验法

假设 x_n 为可疑值

1. 除去可疑值后求平均值 $\bar{x}'$

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

其中 n 为测量次数，即总的的数据个数； x_i 为第 i 次的测得值

2. 求平均偏差 $\bar{d}$

$$\bar{d} = \frac{\sum_{i=1}^{n-1} |x_i - \bar{x}'|}{n-1} \quad (2-2)$$

3. 计算可疑值与 $\bar{x}'$ 的差的绝对值 D

$$D = |(\text{可疑值}) - \bar{x}'| \quad (2-3)$$

4. 将计算得的 D 值与 $\bar{d}$ 比较，如 D 大于 $\bar{d}$ 的四倍，即 $D > 4\bar{d}$ ，则该可疑数据弃去（例 2-1）。

该方法的优点是简单易记，不需要计算标准偏差，也不需查表。它是从下面拉依达（Райта）法简化而来的。本法只适用测量次数较大（ $n > 10$ ）的情况。如测量次数较少（ $n = 5 \sim 10$ ），可改为当 $D > 2.5\bar{d}$ 时，将可疑数据弃去。本法的缺点是当测量次数较少，如 $n < 10$ （使用 $4\bar{d}$ 检验法）或 $n < 5$ （使用 $2.5\bar{d}$ ）时，则即使存在误差大，应该剔除的数据也无法舍去。

例 2-1 今有以下 11 次平行测定的分析数据：30.18；30.23；30.21；30.15；30.28；30.31；30.56；30.32；30.38；30.35；30.19(%)。问其中 30.56% 这一数据是否应弃去？

解：1. 除去 30.56% 这一数据后，求其余 10 个数据的平均值，

$$\begin{aligned} \bar{x}' &= \frac{\sum_{i=1}^{n-1} x_i}{n-1} \\ &= \frac{30.18 + 30.23 + 30.21 + 30.15 + 30.28 \\ &\quad + 30.31 + 30.32 + 30.38 + 30.35 + 30.19}{11-1} \\ &\approx 30.26(\%) \end{aligned}$$