



第3章 分析化学中的误差与 数据处理



内容提要

3.1 分析化学中的误差

3.2 有效数字及其运算规则

3.3 分析化学中的数据处理

3.4 显著性检验

3.5 可疑值取舍

3.6 回归分析法

3.7 提高分析结果准确度的方法



3.1 分析化学中的误差

3.1.1 误差与偏差

1 误差 (Error)

测量值与真值之间的差值称为**误差**。

绝对误差 (absolute error, E) $E = x - x_T$

相对误差 (relative error, E_r)

$$E_r = \frac{E}{x_T} \times 100\% = \frac{x - x_T}{x_T} \times 100\%$$



- ① 绝对误差有大、小，正、负之分；单位与测量值有相同。
- ② 相对误差有大、小，正、负之分，但量纲为1。
- ③ 分析结果的误差表示

n 次平行测定值为 x_1 、 x_2 、 x_3 、 \dots 、 x_n ，其算术平均值为 \bar{x}

分析结果的绝对误差 $E = \bar{x} - x_T$

分析结果的相对误差为 $E_r = \frac{\bar{x} - x_T}{x_T} \times 100\%$

- ④ 相对误差能反映误差在真值中所占的比例，对于比较在各种情况下测定结果的准确度更为合理。



例 用分析天平称取两物体的质量各为2.1750 g和0.2175 g，已分析天平的误差为±0.1mg，计算两次结果的相对误差？

解： $E = \pm 0.2 \text{ mg}$

$$E_{r1} = \frac{E}{m_1} \times 100\% = \frac{\pm 0.2 \text{ mg}}{2.1750 \text{ g}} \times 100\% = \pm 0.0092\%$$

$$E_{r2} = \frac{E}{m_2} \times 100\% = \frac{\pm 0.2 \text{ mg}}{0.2175 \text{ g}} \times 100\% = \pm 0.092\%$$

- ※ 测量绝对误差相同的情况下，测量值越大，相对误差越小；
- ※ 要求相对误差相同，测量值愈大，允许的绝对误差越大；



※ 每个测量值都按其测量的相对误差去影响分析结果，所以每个测量值的相对误差要小于分析结果的允许误差。

例如常量分析结果的允许相对误差为0.1%，则用万分之一分析天平称量时，被称物质的质量最少要多少？

$$E = \pm 0.2 \text{ mg}$$

$$E_r = \frac{E}{m_1} \times 100\% \leq \frac{0.2 \text{ mg}}{m} \times 100\% \quad m \geq 0.2 \text{ g}$$

例如常量分析结果的允许相对误差为0.1%，则滴定分析中每次消耗的滴定剂的体积最少要多少？

$$E = \pm 0.02 \text{ mL}$$

$$E_r = \frac{E}{V} \times 100\% \leq \frac{0.02 \text{ mL}}{V} \times 100\% \quad V \geq 20 \text{ mL}$$



⑤ 真值(True value)

某一物理量本身具有的客观存在的真实值。 x_T

真值：客观存在，但真值不可通过测量得到。

理论真值：化合物的理论含量

约定真值：国际计量学约定的真值

相对真值：标准样、管理样。



例 用重量分析法测定纯 $\text{BaCl}_2 \cdot 2\text{H}_2\text{O}$ 试剂中 Ba 的含量，四次平行测定的结果分别为：56.14%，56.16%，56.17%，56.13%。试计算测定结果的绝对误差和相对误差？

解：
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{56.14\% + 56.16\% + 56.17\% + 56.13\%}{4} = 56.15\%$$

$$x_T = \frac{M_{\text{Ba}}}{M_{\text{BaCl}_2 \cdot 2\text{H}_2\text{O}}} \times 100\% = 56.22\%$$

$$E_a = \bar{x} - x_T = 56.15\% - 56.22\% = -0.07\%$$

$$E_r = \frac{E_a}{x_T} \times 100\% = \frac{-0.07\%}{56.22\%} \times 100\% = -0.12\%$$



2 偏差 (Deviation)

① 偏差

设对一分析对象进行 n 次平行测定，个别测定值分别为 x_1 、 x_2 、 x_3 、 \dots 、 x_n ，这组数据的算术平均值为

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$d_1 = x_1 - \bar{x} \quad d_2 = x_2 - \bar{x} \quad d_n = x_n - \bar{x} \quad d_i = x_i - \bar{x}$$

A n 次测定得到 n 个偏差，偏差有大小、正负之分

B 相对偏差 $d_r = \frac{d}{\bar{x}}$



② 平均偏差, average deviation 用 \bar{d} 表示。

$$\bar{d} = \frac{|d_1| + |d_2| + |d_3| + \dots + |d_n|}{n} = \frac{1}{n} \sum_{i=1}^n |d_i|$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \Rightarrow \sum_{i=1}^n x_i = n\bar{x}$$

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

平均偏差反映了一组平行测定的精密度、无正负之分。



③ 相对平均偏差, 用 \bar{d}_r 表示

$$\bar{d}_r = \frac{\bar{d}}{\bar{x}} \times 100\%$$

④ 标准偏差, standard deviation 用 s 表示

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

⑤ 相对标准偏差 (relative standard deviation, *RSD*)

$$s_r = \frac{s}{\bar{x}} \times 100\% \quad RSD = \frac{s}{\bar{x}} \times 1000^0 /_{00}$$



★ 标准偏差讨论

※ s 能更客观的反映测量数据的精密度

A/%: 0.3, -0.2, -0.4, 0.2, 0.1, 0.4, 0.0, -0.3, 0.2, -0.3

B/%: 0.0, 0.1, -0.7, 0.2, -0.1, -0.2, 0.5, -0.2, 0.3, 0.1

$$\bar{x}_A = 0.0$$

$$\bar{x}_B = 0.0$$

$$\bar{d}_A = \frac{1}{n} \sum_{i=1}^n |d_{i,A}| = 0.24\%$$

$$\bar{d}_B = \frac{1}{n} \sum_{i=1}^n |d_{i,B}| = 0.24\%$$

$$s_A = 0.28\%$$

$$s_B = 0.33\%$$



※ s 的计算

★ 算法一：按定义式计算

★ 算法二：由测量值直接计算

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

★ 算法三：科学计算器计算，注意位数保留



⑥ 极差 (range, R)

一组平行测量数据中，最大值与最小值之差，又称全距。

$$R = x_{\max} - x_{\min} \quad \text{相对极差: } \frac{R}{\bar{x}} \times 100\%$$

使用R表示测定结果的精密度优点是直观简单。不足之处是没有充分利用数据，实际应用中并不多用。

⑦ 中位数 (median, x_M)

将一组测量数据从小到大排列，若测量值个数为奇数时，中位数即为中间的那个数；若测量值为偶数时，中位数为中间两个数的平均值。



例3：用分光光度法测定某试样中微量Cu的含量，六次测定结果分别为：0.21%，0.23%，0.24%，0.25%，0.24%，0.25%。求：平均偏差，相对平均偏差，标准偏差，相对标准偏差，极差，中位数？

$$\text{解： } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{0.21\% + 0.23\% + 0.24\% + 0.25\% + 0.24\% + 0.25\%}{6} = 0.24\%$$

$$d_1 = 0.21\% - 0.24\% = -0.03\%, d_2 = -0.01\%, d_3 = 0.0, d_4 = 0.01\%, d_5 = 0.0, d_6 = 0.01\%,$$

$$\bar{d} = \frac{1}{n} \sum_{i=1}^6 |d_i| = 0.01\% \quad \bar{d}_r = \frac{\bar{d}}{\bar{x}} \times 100\% = 4.2\%,$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = 0.024\%$$

$$s_r = \frac{s}{\bar{x}} \times 100\% = 10\%$$

$$R = x_{\max} - x_{\min} = 0.04\%$$

$$R_r = \frac{R}{\bar{x}} \times 100\% = 16.7\%$$

$$x_M = 0.24\%$$



3.1.2 准确度 (Accuracy) 与精密度 (Precision)

准确度: 测定结果与真值接近的程度, 用误差衡量。

准确度决定于测量中系统误差的大小。

精密度: 平行测定结果相互接近的程度, 用偏差衡量。

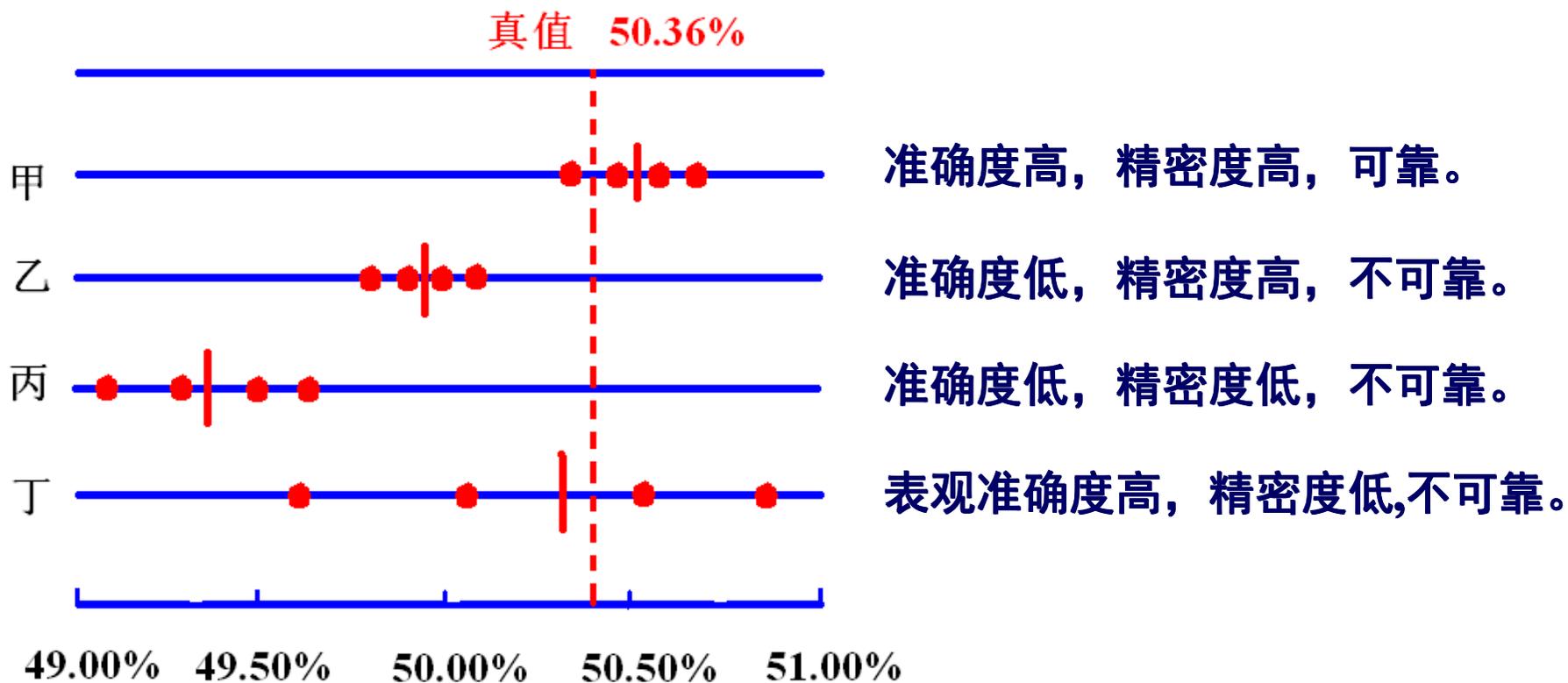
精密度决定于测量中随机误差的大小。

重复性 (repeatability)

再现性(reproducibility)



准确度与精密度的关系



不同分析的分析结果



结论

1. 精密度好是准确度好的前提;
2. 精密度好不一定准确度高。

系统误差!

3. 准确度及精密度都高一结果可靠。



3.1.3 误差的分类及性质

1 系统误差systematic error, bias

① 特点

是由测量中某种固定的原因引起的误差。具单向性、重现性、可校正特点，又叫可测误差。

② 系统误差的分类

方法误差: 分析方法本身不够完善—用其他方法校正;

仪器误差: 测量仪器不够精确—校准仪器;

试剂误差: 水、试剂不纯—空白实验;

操作误差: 掌握的分析方法不够正确;

主观误差: 个人主观原因—多次练习



2 随机误差 random error

① 特点

偶然误差，是由某些难以控制且无法避免的非人为的偶然因素引起的。不可测，无法避免，只能减少、不能消除。

② 随机误差处理方法

增加平行测定的次数，取多次平行测定的平均值作为分析结果。不存在系统误差的情况下，测定次数越多其平均值越接近真值。一般平行测定4 - 6次。

3 过失

本质是错误，由粗心大意引起，可以避免的。

重做，数据必须无条件舍弃。



4 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	固定因素，有时不存在	不定因素，总是存在
分类	方法误差、仪器与试剂误差、主观误差	环境的变化因素、主观的变化因素等
性质	单向性、重现性、可校正	服从概率统计规律、不可校正
影响	准确度	精密度
消除或减小的方法	<u>校正</u>	增加测定的次数 以平均值作为分析结果



例题：下列情况可引起何种误差，如何消除？

- 1、天平零点稍有变动
- 2、过滤时出现透滤现象但没有及时发现
- 3、读取滴定管读数时，最后一位估计不准
- 4、标准试样保存不当，失去部分结晶水
- 5、移液管转移溶液后残留量稍有不同
- 6、试剂中含有少量待测组分
- 7、重量法测定 SiO_2 时，试样中硅酸沉淀不完全
- 8、天平砝码腐蚀
- 9、 NaOH 滴定 HAc 时，选用酚酞做指示剂。



3.1.4 公差

公差:生产部门对分析结果允许误差的一种限量。

公差范围的确定

1 对分析任务准确度的要求

2 试样组成的复杂程度, 待测组分含量的高低

待测组分含量%	~100	~50	~10	~1	~0.1	~0.01-0.001
允许误差%	0.1~0.3	0.3	1	2~5	5~10	~10

3 与所用分析方法有关



3.1.5 误差的传递

$$\omega_B = \frac{\frac{b}{t} c_T V_T M_B}{m_s} \times 100\%$$

误差传递规律取决于误差的性质和测量值间的运算方式

1 系统误差的传递

① 加减法

$$R = mA + nB - pC \quad E_R = mE_A + nE_B - pE_C$$

② 乘除法

$$R = m \frac{AB}{C} \quad \frac{E_R}{R} = \frac{E_A}{A} + \frac{E_B}{B} - \frac{E_C}{C}$$

③ 指数运算

$$R = mA^n \quad \frac{E_R}{R} = n \frac{E_A}{A}$$



④ 对数运算

$$R = m \lg A \quad E_R = 0.434 m \frac{E_A}{A}$$

2 随机误差的传递

① 加减法

$$R = mA + nB - pC \quad s_R^2 = m^2 s_A^2 + n^2 s_B^2 + p^2 s_C^2$$

② 乘除法

$$R = m \frac{AB}{C} \quad \left(\frac{s_R}{R}\right)^2 = \left(\frac{s_A}{A}\right)^2 + \left(\frac{s_B}{B}\right)^2 + \left(\frac{s_C}{C}\right)^2$$

③ 指数运算

$$R = mA^n \quad \frac{s_R}{R} = n \frac{s_A}{A}$$

④ 对数运算

$$R = m \lg A \quad s_R = 0.434 m \frac{s_A}{A}$$



3 极值误差 最大可能误差

$$R = mA + nB - pC \quad \left| E_R \right|_{\max} = \left| mE_A \right| + \left| nE_B \right| + \left| pE_C \right|$$

$$R = m \frac{AB}{C} \quad \left| \frac{E_R}{R} \right|_{\max} = \left| \frac{E_A}{A} \right| + \left| \frac{E_B}{B} \right| + \left| \frac{E_C}{C} \right|$$



例1 欲配制 $0.05000 \text{ mol L}^{-1} \text{ Ca}^{2+}$ 标准溶液，称取 5.0045 g CaCO_3 基准试剂，用 HCl 溶解后，转移至 1000 mL 容量瓶中定溶至刻度。称取 CaCO_3 后，发现天平零点由原来的 0.0 mg 变至 -0.5 mg 。又已知容量瓶的体积校正值为 -0.2 mL 。求所配 Ca^{2+} 标准溶液浓度的相对误差、绝对误差及准确浓度。

解：天平零点的明显变动，容量器皿的体积误差属于系统误差，故要利用系统误差的传递规律计算。

先计算溶液的表现浓度：

$$c_{\text{Ca}^{2+}} = \frac{m_{\text{CaCO}_3}}{M_{\text{CaCO}_3} V} = \frac{5.0045}{100.09 \times 1000.0 \times 10^{-3}} = 0.050000 \text{ mol L}^{-1}$$



再分析误差：

称量质量时，天平的零点实际在 -0.5 mg处，即将试样质量少称了 0.5 mg，即 $E_m = -0.5$ mg。

容量瓶的体积校正值为 -0.2 mL，则容量瓶的真实体积为

$$V_T = 1000.0 - 0.20 = 999.80 \text{ mL,}$$

$$\text{故 } E_V = 1000.0 - 999.80 = 0.20 \text{ mL.}$$

$$\frac{E_{c_{\text{Ca}^{2+}}}}{c_{\text{Ca}^{2+}}} = \frac{E_m}{m} - \frac{E_V}{V} = \frac{-0.0005}{5.0045} - \frac{0.20}{1000.0} = -0.03\%$$

$$E_{c_{\text{Ca}^{2+}}} = c_{\text{Ca}^{2+}} \times \frac{E_{c_{\text{Ca}^{2+}}}}{c_{\text{Ca}^{2+}}} = 0.050000 \times (-0.03\%) = -0.000015 \text{ mol L}^{-1}$$

$$c_T = c - E_c = 0.050000 - (-0.000015) = 0.050015 \text{ mol L}^{-1}$$



例2 天平称量的标准偏差 $s = 0.1$ mg, 求称量试样时的标准偏差。

解：用天平称量时的读数误差是随机误差，随机误差用标准偏差来表示，意即每次读数会引起 ± 0.1 mg的误差。

用天平称量时，不论用何种称量方法，均需前后两次读天平的平衡点，故称取的质量实际上是两次质量的差值，即：

$m = m_2 - m_1$ ，由于每次都有0.1 mg的误差，又属于加减法运算，故：

$$s_m^2 = s_{m_2}^2 + s_{m_1}^2 = 2s^2$$

$$s_m = \sqrt{2s^2} = \sqrt{2 \times 0.1^2} = 0.14 \text{ mg}$$



例3 用 $0.1000 \text{ mol}\cdot\text{L}^{-1}(c_2)$ HCl标准溶液标定 $20.00 \text{ mL}(V_1)$ NaOH溶液，耗去HCl $25.00 \text{ mL}(V_2)$ ，已知用移液管量取溶液体积时的标准偏差为 $s_1 = 0.02 \text{ mL}$ ，每次读取滴定管读数时的标准偏差为 $s_2 = 0.01 \text{ mL}$ ，假定 c_2 是准确的，求NaOH溶液的浓度？

解：先计算溶液的表现浓度：

$$c_1 = \frac{c_2 V_2}{V_1} = \frac{0.1000 \times 25.00}{20.00} = 0.1250 \text{ mol L}^{-1}$$

移液管的标准偏差为 $s_{V_1} = s_1 = 0.02 \text{ mL}$

滴定管的标准偏差为 $s_{V_2} = \sqrt{2s_2^2} = \sqrt{2 \times 0.01^2}$



由随机误差在乘除法运算中的传递规律可知：

$$\left(\frac{s_{c_1}}{c_1}\right)^2 = \left(\frac{s_{V_2}}{V_2}\right)^2 + \left(\frac{s_{V_1}}{V_1}\right)^2 = \frac{2s_2^2}{V_2^2} + \frac{s_1^2}{V_1^2} = \frac{2 \times 0.01^2}{25.00^2} + \frac{0.02^2}{20.00^2} = 1.32 \times 10^{-6}$$

$$\frac{s_{c_1}}{c_1} = 1.15 \times 10^{-3}$$

$$s_{c_1} = c_1 \times 1.15 \times 10^{-3} = 0.1250 \times 1.15 \times 10^{-3} = 0.0001 \text{ mol L}^{-1}$$

故NaOH溶液的准确浓度为：

$$c_{\text{NaOH}} = c_1 \pm s_{c_1} = 0.1250 \pm 0.0001 \text{ mol L}^{-1}$$



例4 滴定管的初读数为 (0.05 ± 0.01) mL，末读数为 (22.10 ± 0.01) mL，问滴定剂的体积可能在多大范围内波动？

解：按极值误差处理：

$$|E_V| = |E_{V_2}| + |E_{V_1}| = |\pm 0.01| + |\pm 0.01| = 0.02 \text{ mL}$$

故滴定剂的体积为：

$$V = (V_2 - V_1) \pm |E_V| = (22.10 - 0.05) \pm 0.02 = 22.05 \pm 0.02 \text{ mL}$$



3.2 有效数字及运算规则

问题：重量法测定SiO₂的含量？

$$\omega_{\text{SiO}_2} = \frac{m_{\text{SiO}_2}}{m_s} = \frac{0.1474}{0.4538} \times 100\% = 30.277655354\%$$

为何小数点后保留四位数字？

应保留几位数字？

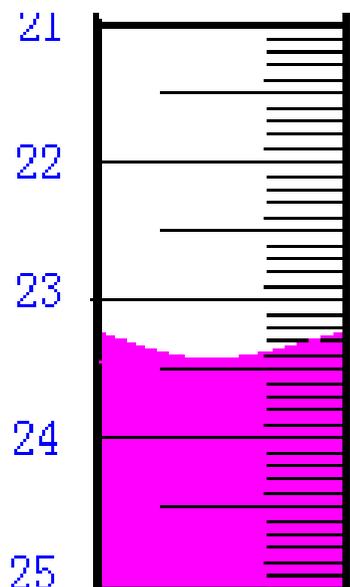
有效数字的问题



3.2.1 有效数字 (Significant figures)

1 有效数字

实际上能测量到的数字。定义为一个数据中全部可靠数字再加一位不确定数字。



23.43、23.42、23.44 mL

最后一位无刻度，估计的，不是很准确，但不是臆造的，称可疑数字或不确定数字，它也是客观存在的。



2 有效数字位数的确定:

1. 非0, \checkmark ; (54、43181、2954)

2. 0是普通数字 \checkmark ; (30.07);

0起小数点的定位作用, \times (0.000005), 15.00000 \checkmark

3. 改变单位不能改变有效数字位数.

4. 计算公式中的自然数, 如倍数, 分数等, 其有效数字的位数没有限制, 需要几为就算几位。 Π, e

5. pH, pM, lgK等对数值, 有效数字位数仅决定于小数部分(尾数)的位数, pH = 10.28(两位), $[H^+] = 5.2 \times 10^{-11} \text{ mol}\cdot\text{L}^{-1}$



6. 34500, 1000 这样的大数, 有效数字位数不确定, 测量结果要求用科学技术法表示。

例: 1.0008, 43.181	5位
0.1000, 10.98%	4位
0.0382, 1.98×10^{-10}	3位
54, 0.0040	2位
0.05, 5×10^5	1位
pH = 11.20	2位



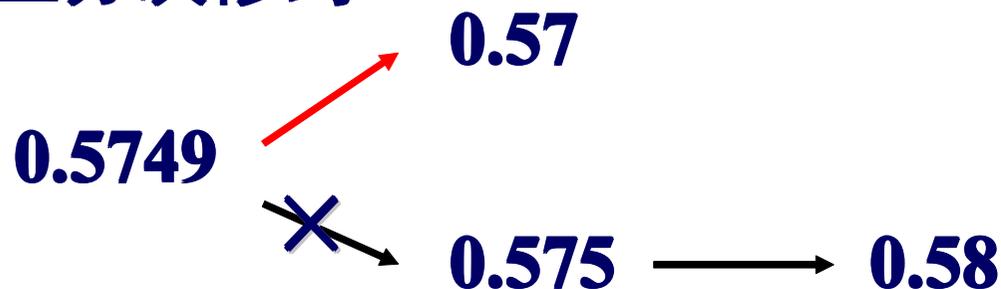
3.2.2 有效数字的修约规则

四舍六入五成双

例 下列值修约为四位有效数字

0.324 74	→	0.324 7
0.324 75	→	0.324 8
0.324 76	→	0.324 8
0.324 85	→	0.324 8
0.324 851	→	0.324 9

禁止分次修约





3.2.3 运算规则

1 加减法: 与小数点后位数最少的数一致。

$$50.1 + 1.45 + 0.5812 = 52.1$$

绝对误差 ± 0.1 ± 0.01 ± 0.0001

$$50.1 + 1.4 + 0.6 = 52.1$$



2 乘除法:与有效数字位数最少的一致。

$$0.0121 \times 25.64 \times 1.05782 = 0.328$$

相对误差 $\pm 8\%$ $\pm 0.4\%$ $\pm 0.009\%$

$$0.0121 \times 25.6 \times 1.06 = 0.328$$



3.2.4 几点补充说明

1 在大量数据运算时，为了避免误差的迅速累积，对参与运算的各数据可多保留1位有效数字，得到结果后，再修约。

$$5.2727 + 0.075 + 3.7 + 2.12 = 5.2_7 + 0.0_8 + 3.7 + 2.1_2 = 11.1_7 = 11.2$$

2 在乘除法运算中，当数据的首位为8或9，该数据有效数字位数应多记1位。 例 $9.0 \times 2.41 \times 28.36 = ?$

相对误差 $\pm 1\% \quad \pm 0.4\% \quad \pm 0.04\%$

$$9.0 \times 2.41 \times 28.4 = 616$$

3 正确报告分析结果

	四位	三位	二位
按组份含量	>10% (高含量)	1%~10% (中含量)	<1% (微量)



4 各类误差、偏差常取1-2位有效数字。

5 用计算器计算时，不必每步都修约，但要准确保留计算结果的位数。



3.7 提高分析结果准确度的方法

1 选择合适的分析方法

化学分析法准确度高，灵敏度低，适合高含量组分的测定

仪器分析法灵敏度高，准确度低，适合低含量组分的测定

2 减小测量误差

- ① 合理的控制被测量量值的大小，减小测量误差；
- ② 各测量值的误差应与分析方法的误差相适应。



4 减小随机误差

增加平行测定的次数，减小随机误差。

一般分析：3-4次

精密分析：5-9次

要求更高：10-12次

本章作业

P₇₄₋₇₅ 1, 2, 3, 4



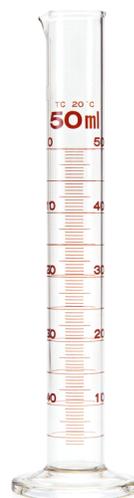
滴定分析常用仪器 分析天平

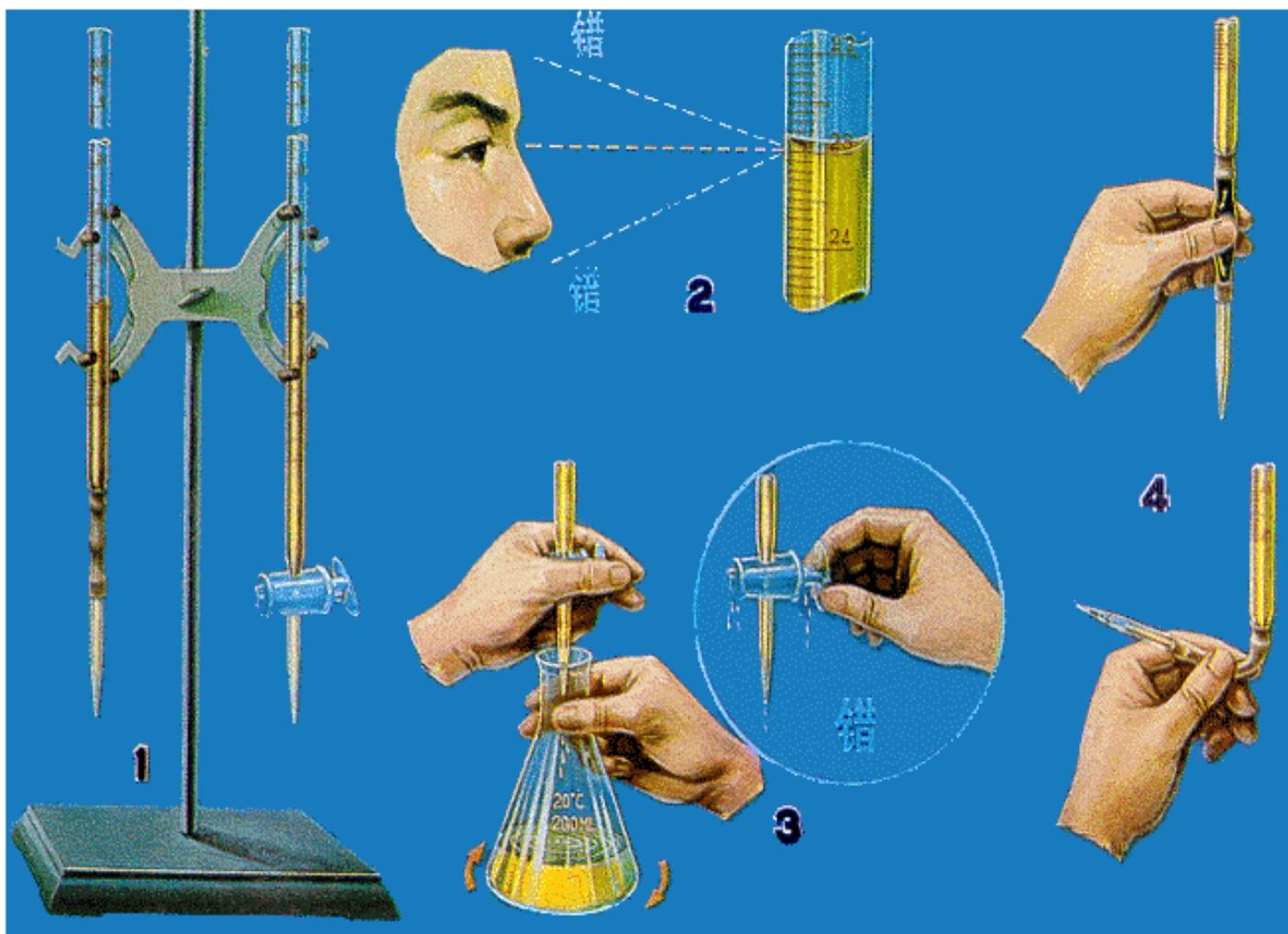




滴定分析常用仪器

容量仪器





滴定管的使用图示

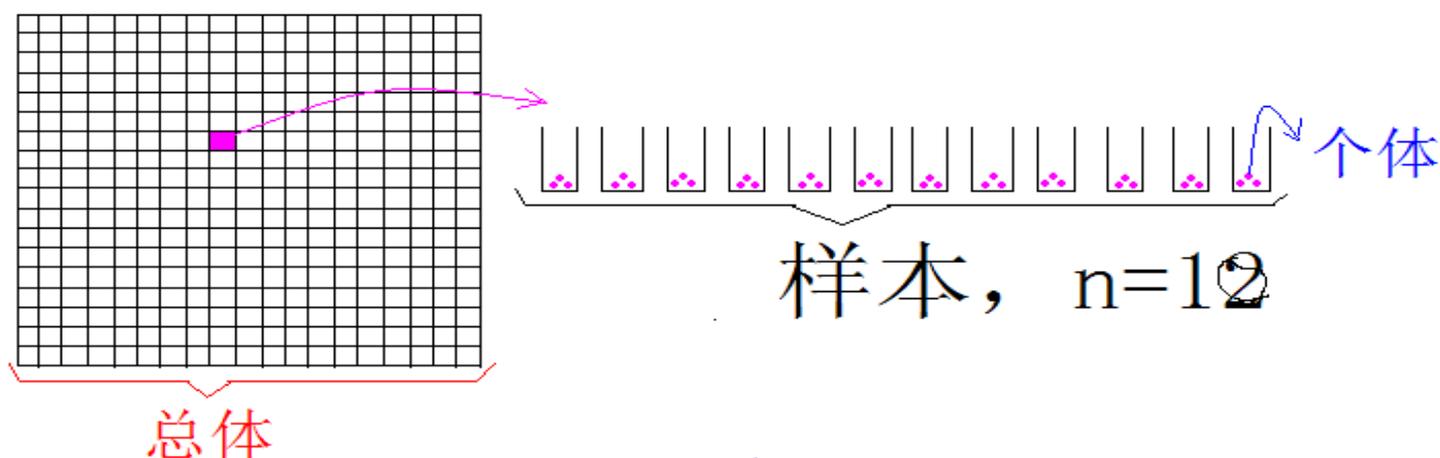
1. 滴定管架上的滴定管(左、碱式 右、酸式)。2. 观看管内液面的位置: 视线跟管内液体的凹液面的最低处保持水平。3. 酸式滴定管的使用: 右手拿住锥形瓶颈, 向同一方向转动。左手旋开(或关闭)活塞, 使滴定液逐滴加入。4. 碱式滴定管的使用: 左手捏挤玻璃球处的橡皮管, 使液体逐滴下降。如果管内有气泡, 要先赶走气泡。



3.3 分析化学中的数据处理

基本概念

- ① 总体（母体）—所考察对象的全体
- ② 个体—总体中的一个单位或某一次测定。
- ③ 样本—自总体中随机抽出的一组测量值。
- ④ 样本容量—样本中所含测量值的数目。





3.3.1 随机误差的正态分布

1 频数分布

用吸光光度法测定某合金样品中Fe的含量，进行了100次测定，获得了100个测量数据。

1.36	1.49	1.43	1.37	1.40	1.32	1.42	1.47	1.39	1.41
1.41	1.36	1.40	1.34	1.42	1.42	1.45	1.35	1.42	1.39
1.44	1.42	1.39	1.42	1.42	1.30	1.34	1.42	1.37	1.36
1.37	1.34	1.37	1.46	1.44	1.45	1.32	1.48	1.40	1.45
1.39	1.46	1.39	1.53	1.36	1.48	1.40	1.39	1.38	1.40
1.46	1.45	1.50	1.43	1.45	1.43	1.41	1.48	1.39	1.45
1.37	1.46	1.39	1.45	1.31	1.41	1.44	1.44	1.42	1.47
1.35	1.36	1.39	1.40	1.38	1.35	1.42	1.43	1.42	1.42
1.42	1.40	1.41	1.37	1.46	1.36	1.37	1.26	1.47	1.38
1.42	1.34	1.43	1.42	1.41	1.41	1.44	1.48	1.56	1.37

可绘制频数分布直方图来研究其分布的规律。



求极差 $R = x_{\max} - x_{\min} = 1.56 - 1.26 = 0.30$

分组，求组距 = $R / \text{组数} = 0.30 / 10 = 0.03$

统计频数，计算相对频数；

列表、绘图。

表 7-1 频数分布表

分 组	频 数	相对频数
1.265% ~ 1.295%	1	0.01
1.295% ~ 1.325%	4	0.04
1.325% ~ 1.355%	7	0.07
1.355% ~ 1.385%	17	0.17
1.385% ~ 1.415%	24	0.24
1.415% ~ 1.445%	24	0.24
1.445% ~ 1.475%	15	0.15
1.475% ~ 1.505%	6	0.06
1.505% ~ 1.535%	1	0.01
1.535% ~ 1.565%	1	0.01
Σ	100	1.00

80%

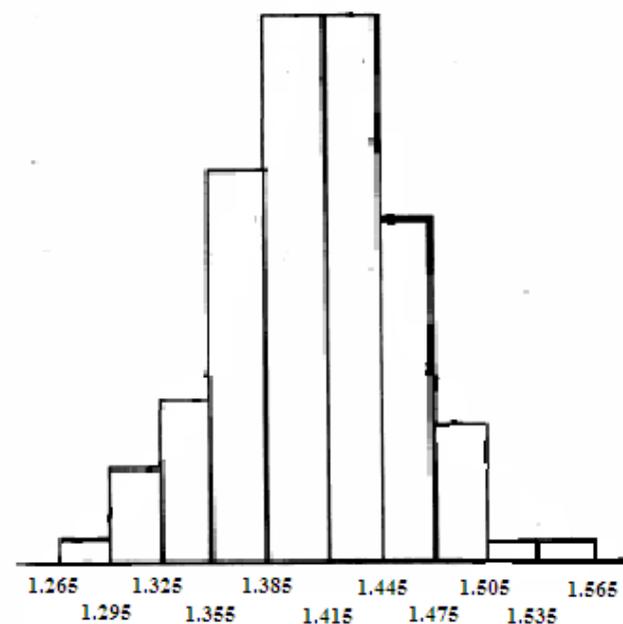


图 3-2 相对频数分布直方图



① 测量数据分散性的表达

总体标准偏差

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

样本标准偏差

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

$$n \rightarrow \infty, \bar{x} \rightarrow \mu \quad \lim_{n \rightarrow \infty} \frac{\sum (x - \bar{x})^2}{n-1} = \frac{\sum (x - \mu)^2}{n} \quad s \rightarrow \sigma$$

总体平均偏差

$$\delta = \frac{\sum |x_i - \mu|}{n}$$

样本平均偏差

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n}$$

$$\delta = 0.7979\sigma \approx 0.80\sigma$$

$$\bar{d} \neq 0.80s$$



② 测量数据集中趋势的表达

$$\mu = \lim_{n \rightarrow \infty} \frac{1}{N} \sum x_i \quad \text{若无系统误差, } \mu = x_T$$

$$\bar{x} = \frac{1}{n} \sum x_i$$

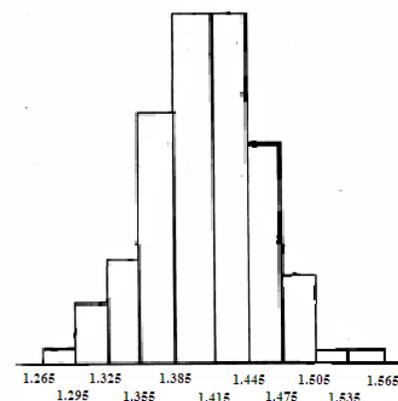


图 3-2 相对频数分布直方图

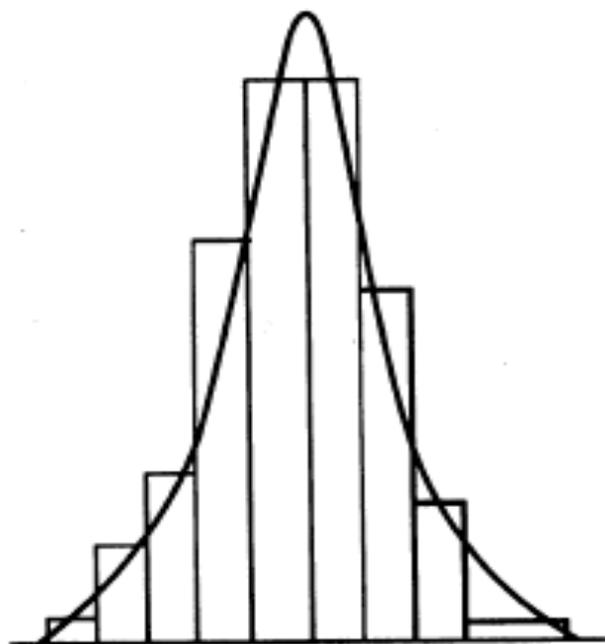


图 7-2 相对频数分布直方图



2 正态分布

① 正态分布曲线及随机误差的分布规律

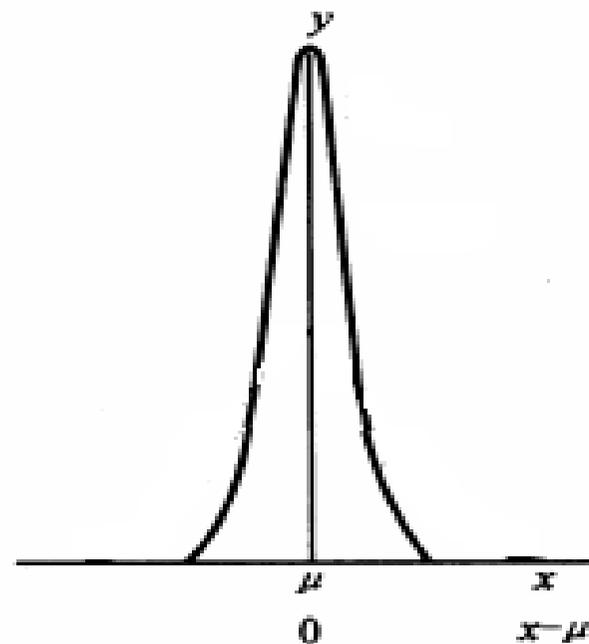
$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

y 概率密度

σ 总体标准偏差，表示测量数据分散性

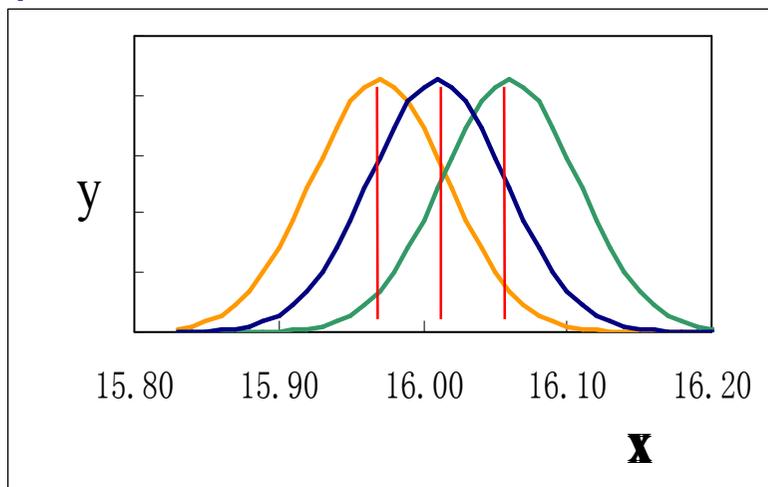
μ 总体平均值，表示测量数据值集中趋势

$x-\mu$ 随机误差





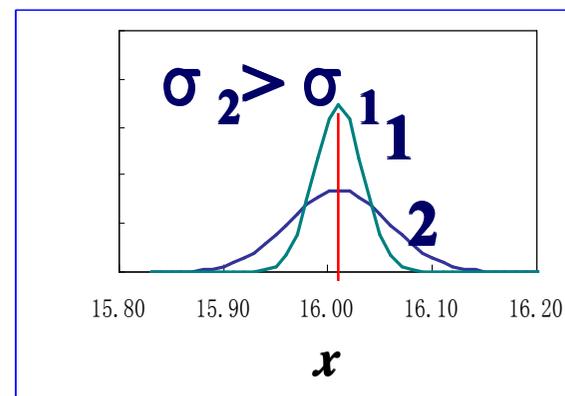
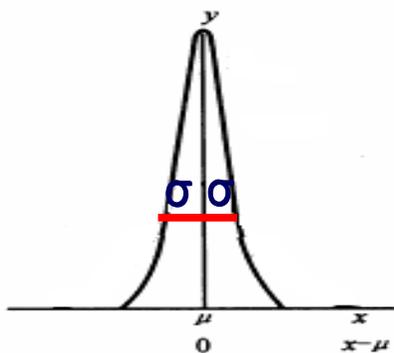
μ 总体平均值，表示测量数据值集中的趋势。



- 1 总体不同
- 2 同一总体，存在系统误差

$$N(\mu, \sigma^2)$$

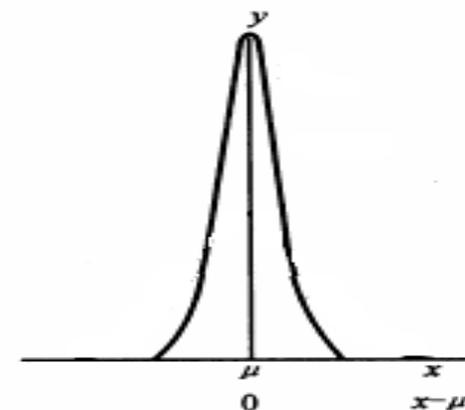
σ 总体标准偏差，表示测量数据的分散性。





随机误差的分布规律

A 当 $x = \mu$ 时， y 值最大，单峰性，测量数据有集中趋势；



B 曲线以 $x = \mu$ 为对称轴左右对称，对称性，说明绝对值相等的正负误差出现的概率相等，平均值是最可信赖值。

C 曲线以 x 轴为渐近线，有界性，说明小误差出现的概率大，大误差出现的概率小，极大误差出现的概率趋于零。



② 随机误差的区间概率

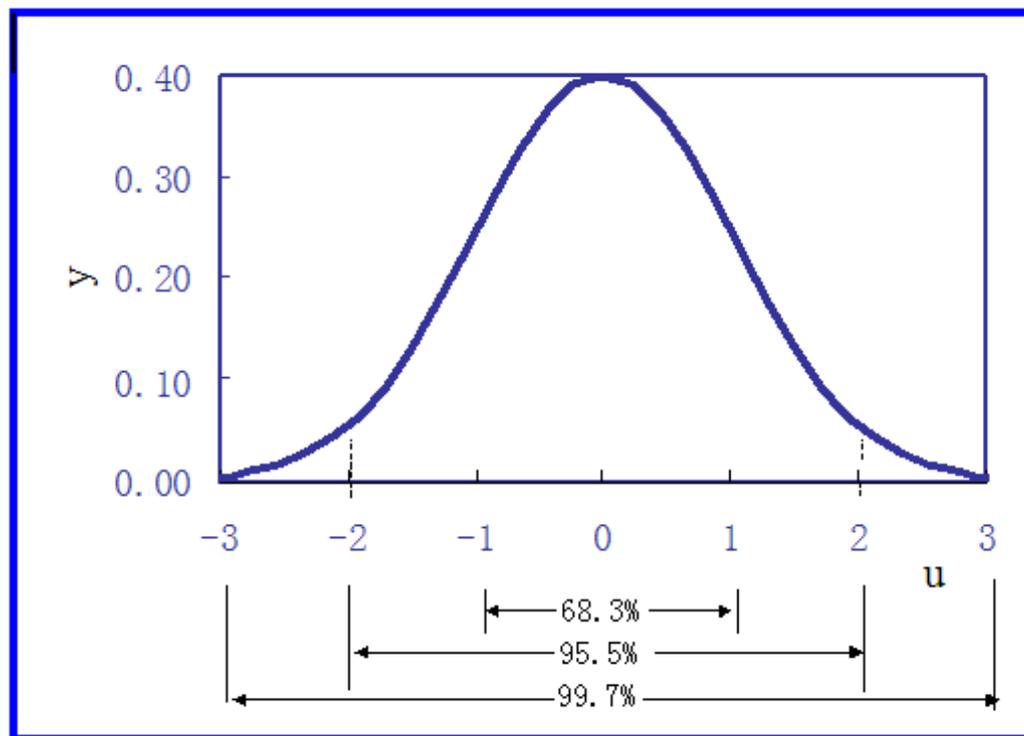
测量值（或随机误差）出现在任一区间的概率叫测量值（或随机误差）的区间概率。

$$P(-\infty < x < +\infty) = \frac{1}{\sigma \times \sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} dx = 1 \quad P(a < x < b) = \frac{1}{\sigma \times \sqrt{2\pi}} \int_a^b e^{-(x-\mu)^2/2\sigma^2} dx$$

$$u = \frac{X - \mu}{\sigma}, \quad \text{则 } du = \frac{dx}{\sigma}, \quad dx = \sigma du \quad y = f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

$$f(x) dx = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{u^2}{2}} \cdot \sigma du = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \cdot du = \Phi(u) du$$

$$\therefore y = \Phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$



标准正态分布曲线以 u 为横坐标，以概率密度为纵坐标，以 $N(0, 1)$ 表示。对不同总体或不同标准偏差的测量值，标准正态分布曲线均是适用的。



$$P = \int_{-u}^{+u} \phi(u) \cdot du = \int_{-u}^{+u} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

正态分布概率积分表

$$\text{概率} = \text{面积} = \int_0^u \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

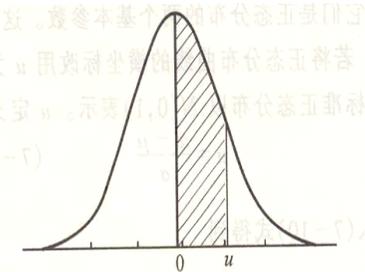


图 7-5 正态分布概率积分图

u	面积	u	面积	u	面积
0.0	0.000 0	1.0	0.341 3	2.0	0.477 3
0.1	0.039 8	1.1	0.364 3	2.1	0.482 1
0.2	0.079 3	1.2	0.384 9	2.2	0.486 1
0.3	0.117 9	1.3	0.403 2	2.3	0.489 3
0.4	0.155 4	1.4	0.419 2	2.4	0.491 8
0.5	0.191 5	1.5	0.433 2	2.5	0.493 8
0.6	0.225 8	1.6	0.445 2	2.6	0.495 3
0.7	0.258 0	1.7	0.455 4	2.7	0.496 5
0.8	0.288 1	1.8	0.464 1	2.8	0.497 4
0.9	0.315 9	1.9	0.471 3	3.0	0.498 7



随机误差出现的区间 ^a	测量值出现的区间	概率
$u = \pm 1$	$x = \mu \pm 1\sigma$	68.3%
$u = \pm 1.96$	$x = \mu \pm 1.96\sigma$	95.0%
$u = \pm 2$	$x = \mu \pm 2\sigma$	95.5%
$u = \pm 2.58$	$x = \mu \pm 2.58\sigma$	99.0%
$u = \pm 3$	$x = \mu \pm 3\sigma$	99.7%

例1 按照正态分布求测量值 x 在区间 $(\mu - 0.5\sigma, \mu + 1.5\sigma)$ 出现的概率。

$$u = \frac{x - \mu}{\sigma} \quad x = \mu \pm u\sigma \quad u (-0.5, 1.5)$$

$$|u| = 0.5, P_1 = 0.1915 \quad |u| = 1.5, P_2 = 0.4332$$

$$P = P_1 + P_2 = 0.1915 + 0.4332 = 0.6247$$

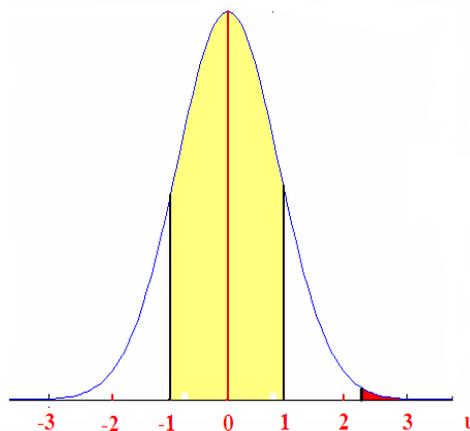


例2 已知某试样中Cu质量分数的标准值为1.48%， $\sigma = 0.10\%$ ，测量时没有系统误差。（1）求分析结果落在 $(1.48 \pm 0.10)\%$ 范围内的概率？（2）分析结果大于1.70%的概率？

解：

(1) 找u值 $x = 1.48\% \pm 0.10\%$ $u = \frac{x - \mu}{\sigma} = \frac{\pm 0.10\%}{0.10\%} = \pm 1$

$|u| = 1, P = 0.3413$ $u(-1, 1), P = 2 \times 0.3413 = 0.6826$



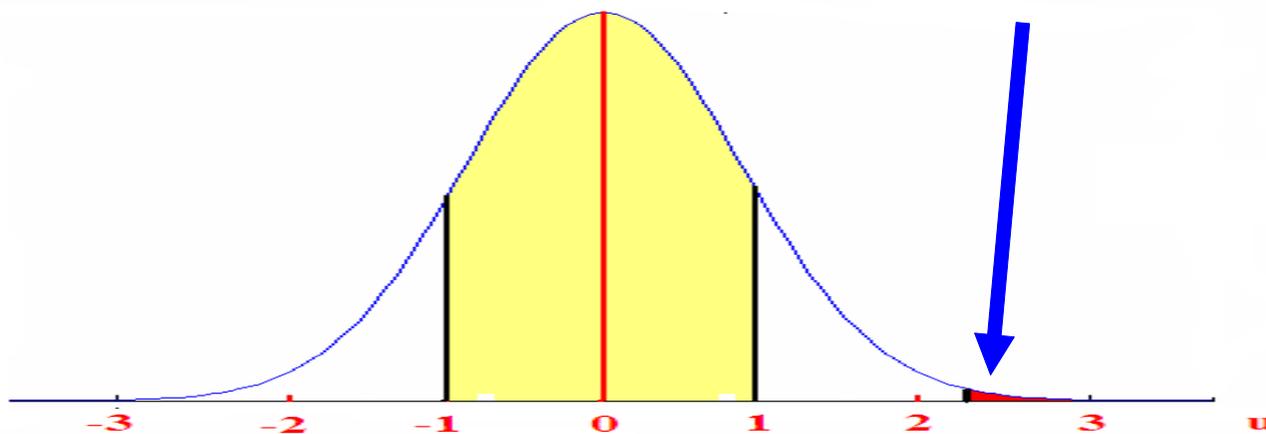


(2) 找u值

$$u = \frac{x - \mu}{\sigma} = \frac{1.70\% - 1.48\%}{0.10\%} = \frac{0.22\%}{0.10\%} = 2.2$$

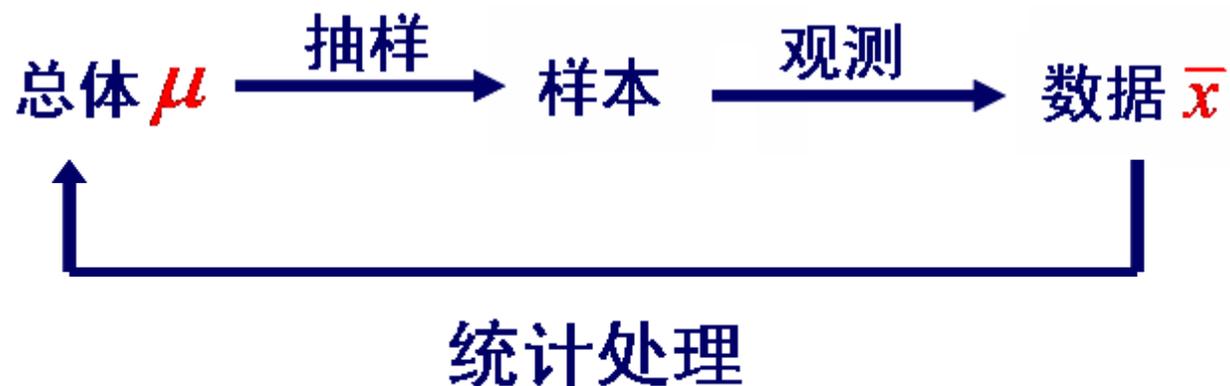
单边问题

$$|u| = 2.2, P = 0.4861 \quad P = 0.5000 - 0.4861 = 0.0139$$





3.3.2 总体平均值的估计





1 平均值的标准偏差 平均值的可靠性

试样总体	样本1	$x_{11}, x_{12}, x_{13}, \dots, x_{1n} \rightarrow \bar{x}_1$
	样本2	$x_{21}, x_{22}, x_{23}, \dots, x_{2n} \rightarrow \bar{x}_2$

	样本 m	$x_{m1}, x_{m2}, x_{m3}, \dots, x_{mn} \rightarrow \bar{x}_m$

平均值分别为 $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_m \rightarrow \bar{\bar{x}}$

这些平均值的精密度，可用样本平均值的标准偏差 $S_{\bar{x}}$ 来表示

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \bar{d}_x = \frac{\bar{d}}{\sqrt{n}} \quad \delta_x = \frac{\delta}{\sqrt{n}}$$

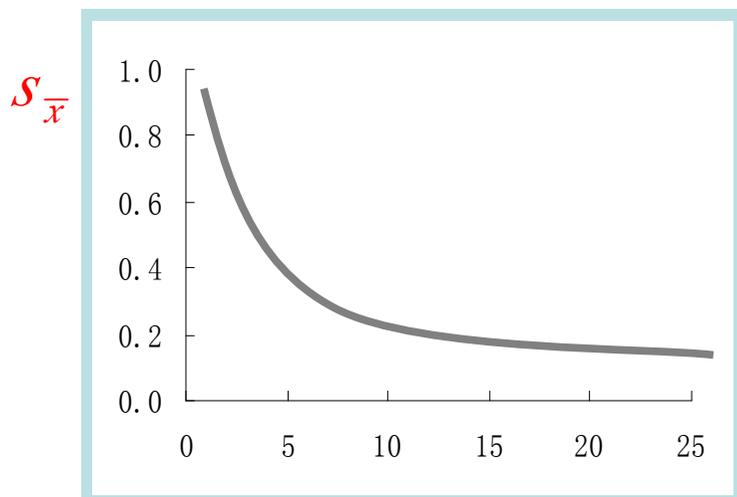


平均值的标准偏差在分析化学中的意义

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

① 增加测量次数可以提高精密度。

但过多增加测量次数是没必要的。



n

一般分析：3-4次

精密分析：5-9次

要求更高：10-12次

② 计算平均值的置信区间。



2 平均值的置信区间 ——对总体平均值 μ 的估计

在一定的置信度下，以单次测量结果或样本平均值为中心的包括总体平均值 μ 在内的可靠性范围。

随机误差出现的区间 ^a	测量值出现的区间	概率
$u = \pm 1$	$x = \mu \pm 1\sigma$	68.3%
$u = \pm 1.96$	$x = \mu \pm 1.96\sigma$	95.0%
$u = \pm 2$	$x = \mu \pm 2\sigma$	95.5%
$u = \pm 2.58$	$x = \mu \pm 2.58\sigma$	99.0%
$u = \pm 3$	$x = \mu \pm 3\sigma$	99.7%



$$u = \frac{x - \mu}{\sigma}$$

$$\mu = x \pm u\sigma$$

对上面的结果也可以倒过来说：有95%把握认为在 $x \pm 1.96 \sigma$ 区间内包括真值 μ 。

此处，95%概率就叫置信度P； $1 - P = \alpha$ 叫危险率（显著性水准）。

知道 σ ，以单次测量结果计算平均值的置信区间的公式：

$$\mu = x \pm u\sigma$$

指定置信度（概率），查u值表知道对应的u值，代入计算。



用样本平均值 估计平均值的置信区间

$$u = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$\mu = \bar{x} \pm \frac{u\sigma}{\sqrt{n}}$$

例：用标准方法分析钢样中P%，平行测定4次，得样本平均值为0.087%，且 $\sigma=0.002\%$ ，求置信度为95%时平均值的置信区间。

解：因为 $\sigma=0.002\%$ 已知，故用正态分布规律处理。

查正态分布概率积分表， $P=95\%$ 时， $u = 1.96$

$$\mu = \bar{x} \pm \frac{u\sigma}{\sqrt{n}} = 0.087\% \pm \frac{1.96 \times 0.002\%}{\sqrt{4}} = (0.087 \pm 0.002)\%$$

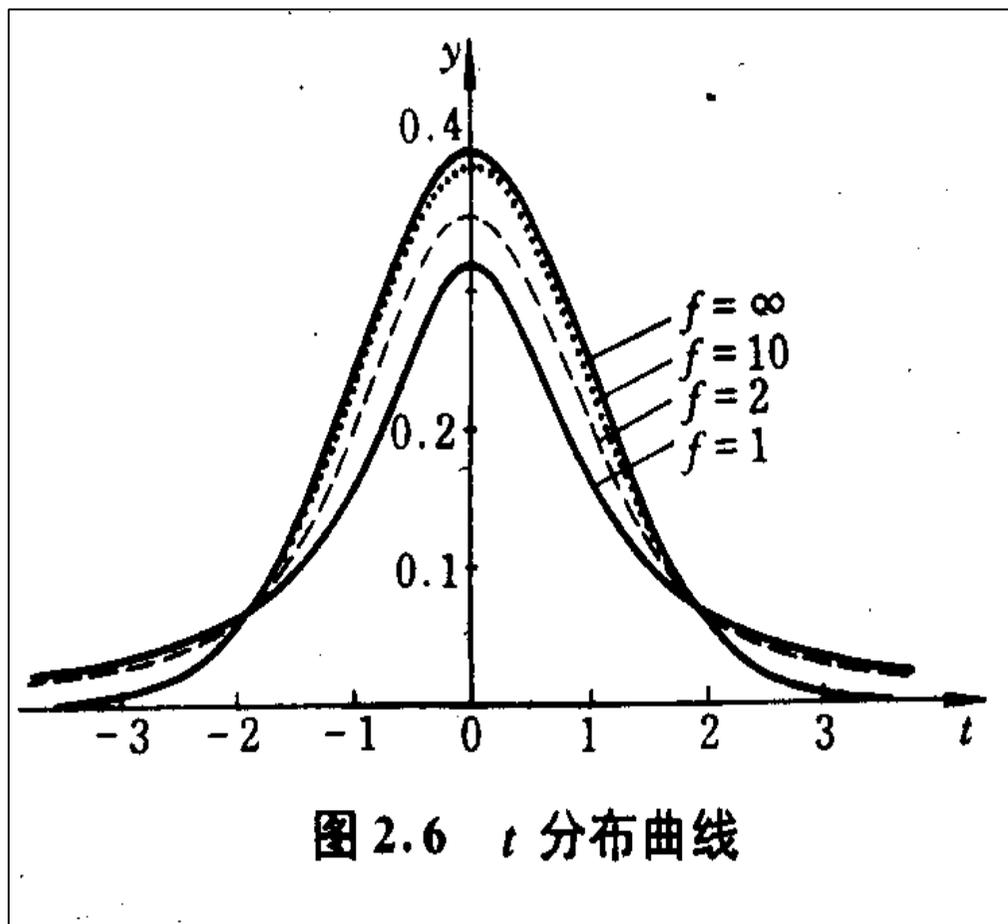
0.085% ~ 0.089%



2 少量实验数据的统计处理

① t 分布曲线

$$t = \frac{x - \mu}{s}$$





$t_{\alpha, f}$ 值表

自由度 $f=(n-1)$	置信度P			
	0.50	0.90	0.95	0.99
1	1.00	6.31	12.71	63.66
2	0.82	2.92	4.30	9.92
3	0.76	2.35	3.18	5.84
4	0.74	2.13	2.78	4.60
5	0.73	2.02	2.57	4.03
6	0.72	1.94	2.45	3.71
7	0.71	1.90	2.36	3.50
8	0.71	1.86	2.31	3.36
9	0.70	1.83	2.26	3.25
10	0.70	1.81	2.23	3.17
20	0.69	1.72	2.09	2.84
∞	0.67	1.64	1.96	2.58



★ 已知 s 时平均值的置信区间的计算

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} \quad t = \frac{\bar{x} - \mu}{S_{\bar{x}}} \quad \mu = \bar{x} \pm t_{\alpha, f} S_{\bar{x}} = \bar{x} \pm \frac{t_{\alpha, f} S}{\sqrt{n}}$$

平均值置信区间计算的思路：

- 1) 计算若干次测定结果的算术平均值及标准偏差；
- 2) 指定置信度，按指定的置信度及自由度查 t 值；
- 3) 代入公式进行计算。



例：分析某铁矿石中铁的含量。在一定条件下，平行测定5次，其结果为：39.10、39.12、39.19、39.17、39.22（%）。求置信度95%时，平均值的置信区间？

已知： $\bar{x} = 39.16\%$, $s = 0.05\%$, $n = 5$

查表知： $P = 95\%$, $f = 4$, $t_{0.05,4} = 2.78$

$$\mu = \bar{x} \pm \frac{t_{0.05,4}s}{\sqrt{n}} = 39.16\% \pm \frac{2.78 \times 0.05\%}{\sqrt{5}} = (39.16 \pm 0.06)\%$$



② 讨论

1) 平均值置信区间的意义

在一定的置信度下，以单次测定值 x 或样本平均值为中心的包括总体平均值 μ 在内的可靠性范围。

$$\mu = (39.16 \pm 0.06)\% (P = 95\%)$$

理解：在该区间内包括总体平均值的概率为95%；
有95%的把握认为该区间包括了总体平均值。

不能说：总体平均值落在该区间的概率为95%。



2) 置信度与置信区间的关系

置信度

置信区间

估计的把握程度，可靠性

估计的精度

i) 置信度高，估计的把握程度大，失误机会少，做出判断的可靠性高，置信区间宽，估计的精度差。P62-例10

ii) 置信度低，估计的把握程度小，失误机会大，做出判断的可靠性差，置信区间窄。

在分析化学中， $P = 95\%$ ， 90%



3) 置信区间与精密度和测定次数的关系

$$\mu = \bar{x} \pm t_{\alpha, f} S_x = \bar{x} \pm \frac{t_{\alpha, f} S}{\sqrt{n}}$$

置信区间还与测定的精密度和测定次数有关，一定的置信度下，测定次数越多，测定精密度越好，置信区间越窄，平均值就越接近总体平均值，没有系统误差时，测定结果越准确。



分析结果的表达

\bar{x} , s , n

$$\mu = \bar{x} \pm t_{\alpha, f} S_{\bar{x}} = \bar{x} \pm \frac{t_{\alpha, f} S}{\sqrt{n}} (P)$$



3.4 显著性检验

①某分析人员，用某种分析方法对标准试样进行分析，所得数据平均值与标准值不一致；

$$\bar{x} - \mu \neq 0$$

②某分析人员用两种不同的分析方法对某试样进行分析，所得数据平均值不一致；

$$\bar{x}_1 - \bar{x}_2 \neq 0$$

③不同的分析人员、不同的实验室对同一试样进行分析，所得数据平均值不一致；

显著性检验

$$\bar{x} - \mu \neq 0$$

显著性
检验

显著性差异 $\xrightarrow{\text{系统误差}}$ 校正

$$\bar{x}_1 - \bar{x}_2 \neq 0$$

非显著性差异 $\xrightarrow{\text{随机误差}}$ 正常



3.4.1 t 检验法

1 平均值与标准值的比较

① 检验对象：某分析人员、某分析方法是否可靠； $\bar{x} - \mu \neq 0$

② 检验原理：置信区间检验；

③ 检验步骤：

1) 进行分析测定，得到分析结果；

$$\bar{x}, s, n$$

2) 计算统计量 t ；

$$t = \frac{|\bar{x} - \mu|}{s} \sqrt{n}$$

3) 指定置信度 P ，根据 P 、 f 查表； $t_{\alpha, f}$

D 判断： $t > t_{\alpha, f}$ 表明有系统误差存在。否则不存在。

置信度一般为95%，不能过低、过高。

例1: 采用一种新方法测定基准明矾中铝的质量分数，9次测定结果分别为10.74%，10.77%，10.77%，10.77%，10.81%，10.82%，10.73%，10.86%，10.81%，已知明矾中铝含量的标准值为10.77%（以理论值代替）。试问采用新方法后是否引起系统误差？（P = 95%）

解: 1) 计算平均值、标准偏差:

$$\bar{x} = \frac{\sum x_i}{n} = 10.79\% \quad s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = 0.042\% \quad n = 9, f = 9 - 1 = 8$$

2) 计算t值:
$$t = \frac{|\bar{x} - \mu|}{s} \sqrt{n} = \frac{|10.79\% - 10.77\%|}{0.042} \sqrt{9} = 1.43$$

3) 由置信度及自由度查t值表, 得到 $t_{\alpha, f}$

已知P = 95%, f = 8, 则 $t_{\alpha, f} = t_{0.05, 8} = 2.31$

4) 判断: $t < t_{0.05, 8}$, 故 \bar{x} 与 μ 之间不存在显著性差异, 即采用新方法后没有引起明显的系统误差, 新方法准确可靠。



2 两组平均值的比较 ($F+t$)

① 检验对象：不同分析人员、不同实验室、不同分析方法；

② 检验原理：置信区间检验；

$$\bar{x}_1 - \bar{x}_2 \neq 0$$

③ 检验步骤：

1) 进行分析测定，得到分析结果； \bar{x}_1, n_1, s_1 \bar{x}_2, n_2, s_2

2) 进行F检验，判断精密度间有无显著性差异；

3) 计算总自由度及合并标准偏差s；

$$\text{总自由度} = n_1 + n_2 - 2 \quad s = \sqrt{\frac{\text{偏差平方和}}{\text{总自由度}}} = \sqrt{\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}}$$



4) 计算t值;

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

5) 指定置信度P, 根据P查 $f = n_1 + n_2 - 2$ 表 $t_{\alpha, f}$;

6) 判断: $t > t_{\alpha, f}$ 表明有系统误差存在。否则不存在。

3.4.2 F 检验法

通过比较两组数据的方差, 以判断两组数据的精密度间是否有显著性差异。

$$\bar{x}_1, n_1, s_1 \quad \bar{x}_2, n_2, s_2$$

1) 分别计算两组数据的方差;

2) 计算统计量F;

$$F = \frac{s_{\text{大}}^2}{s_{\text{小}}^2}$$

3) 指定置信度, 查F表;

4) 判断 $F > F_{\text{表}}$, 精密度有显著性差异; 否则, 无。



F检验时应注意的两个问题

- ① $f_{\text{大}}$ 指偏差大的那组数据的自由度；
 $f_{\text{小}}$ 指偏差小的那组数据的自由度。
- ② 确定待检验的问题属于双边检验还是单边检验。

置信度95%时部分F值（单边）

置信度90%时部分F值（双边）

$f_{\text{小}} \backslash f_{\text{大}}$	2	3	4	5	6
2	19.00	19.16	19.25	19.30	19.33
3	9.55	9.28	9.12	9.01	8.94
4	6.94	6.59	6.39	6.16	6.09
5	5.79	5.41	5.19	5.05	4.95
6	5.14	4.76	4.53	4.39	4.28



例3 用两种不同的方法测定合金中钼的质量分数，结果如下：

$$\text{No.1 } \bar{x}_1 = 1.24\%, s_1 = 0.021\%, n_1 = 3$$

$$\text{No.2 } \bar{x}_2 = 1.33\%, s_2 = 0.017\%, n_2 = 4$$

试问这两种方法间有无显著性差异（ $P = 90\%$ ）？

解： $F + t$

$$F_{\text{计}} = \frac{S_{\text{大}}^2}{S_{\text{小}}^2} = \frac{(0.021)^2}{(0.017)^2} = 1.53$$

查表得 $f_{\text{大}} = 2$ ， $f_{\text{小}} = 3$ 时 $F_{\text{表}} = 9.55$ ，估 $F_{\text{计}} < F_{\text{表}}$ ，这两组数据的精密度间无显著性差异（ $P = 90\%$ ）。



$$s_{\text{合}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 0.019\%$$

$$t_{\text{计}} = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{\text{合}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = 6.21$$

P = 90%, $f = n_1 + n_2 - 2 = 5$ 时, $t_{0.10,5} = 2.02$, $t_{\text{计}} > t_{\text{表}}$,

两组数据之间有显著性差异, 有系统误差。



例：在吸光光度分析中，用一台旧仪器测定溶液6次，得标准偏差 $S_1=0.055$ ，再用一台性能稍好的新仪器测定4次，得标准偏差 $S_2=0.022$ 。问新仪器的精密度是否显著地优于旧仪器的精密度？

解： 单边检验。

$$(1) \text{旧仪器: } n_1 = 6, s_1 = 0.055, s_{\text{大}}^2 = 0.055^2 = 0.003$$

$$\text{新仪器: } n_2 = 4, s_2 = 0.022, s_{\text{小}}^2 = 0.022^2 = 0.00048$$

$$(2) F = \frac{s_{\text{大}}^2}{s_{\text{小}}^2} = \frac{0.003}{0.00048} = 6.25 \quad (3) \text{查表: } F_{\text{表}} = 9.01$$

$$(4) F < F_{\text{表}}, \text{ 无显著性差异}$$

即新仪器并不明显的优于旧仪器， $P=95\%$ 。



3.5 可疑值取舍

过失误差的判断

3.5.1 $4\bar{d}$ 法 $\delta = 0.80\sigma \approx \frac{3}{4}\sigma$ $4\delta \approx 3\sigma$ $4\bar{d} \approx 3s$

步骤

A 求出除异常值外的其余数据的平均值 \bar{x} 和平均偏差 \bar{d} 。

B 计算可疑值偏差的绝对值并与 $4\bar{d}$ 比较。

若： $|x - \bar{x}| > 4\bar{d}$ 则舍，否则留。

例：测定药物中的钴含量($\mu\text{g/g}$): 1.25, 1.27, 1.31, 1.40, 试问1.40这个数应保留还是舍去?

$$\bar{x} = \frac{1.25 + 1.27 + 1.31}{3} = 1.28, \bar{d} = 0.023$$

$$|x - \bar{x}| = |1.40 - 1.28| = 0.12 > 4\bar{d}$$

舍



3.5.2 格鲁布斯(Grubbs)检验法 G检验法

1、基本步骤（只有1个可疑值）

(1) 排序： $x_1, x_2, x_3, x_4, \dots, x_n$

(2) 求 \bar{x} 和标准偏差 s

(3) 计算 T 值：
$$T = \frac{\bar{x} - x_1}{s} \quad T = \frac{x_n - \bar{x}}{s}$$

(4) 由测定次数和要求的置信度，查表得 $T_{\alpha, n}$

(5) 判断：若 $T > T_{\alpha, n}$ ，弃去可疑值，反之保留。



2、如果可疑值有两个或两个以上，且分布在数据的同一侧，例如 x_1 ， x_2 都是可疑值，那么检验时先检验内侧数据 x_2 是否应该舍弃，若 x_2 应该舍弃，则没有必要检验 x_1 ， x_1 也要舍弃。若 x_2 不应舍弃，则还要检验 x_1 。在检验 x_2 时，不要 x_1 ，这时测定次数也应减少1次。

3、若果可疑值有两个或两个以上，且分布在平均值的两侧，如 x_1 和 x_n 都是可疑数据，那么应分别先后检验 x_1 和 x_n 是否应该舍弃，如果确定有1个要舍弃，那么检验第二个时，就不要已舍弃的数据，测定次数应按少1次来处理，且应选择99%的置信度。



例题：用G法检验上题中1.40这个数据是否应舍弃？（P=95%）

解：排序： 1.25, 1.27, 1.31, 1.40。

计算平均值及标准偏差：

$$\bar{x} = \frac{1.25 + 1.27 + 1.31 + 1.40}{4} = 1.31, s = 0.066$$

计算T值： $T = \frac{x_n - \bar{x}}{s} = \frac{1.40 - 1.31}{0.066} = 1.36$

查表：已知P=95%时， $T_{0.05,4} = 1.46$ ，

判断，因为 $T < T_{0.05,4}$ ，留。



3.5.3 Q 检验法

(1) 排序: $x_1, x_2, x_3, x_4, \dots, x_n$

(2) 计算 Q 值: $Q = \frac{x_2 - x_1}{x_n - x_1} \quad Q = \frac{x_n - x_{n-1}}{x_n - x_1}$

(3) 由测定次数和要求的置信度, 查表得 $Q_{表}$

(4) 判断, 若 $Q > Q_{表}$ 舍; 否则, 留。



例题：用Q法检验上题中1.40这个数据是否应舍弃？（P=90%）

解：（1）排序：1.25, 1.27, 1.31, 1.40。

（2）计算Q值：
$$Q = \frac{x_n - x_{n-1}}{x_n - x_1} = \frac{1.40 - 1.31}{1.40 - 1.25} = 0.60$$

（3）查表：已知P=90%，n=4时，Q表=0.76，

（4）判断，因为Q < Q表，留。



3.5.4 几点注意

(1) $4\bar{d}$ 法较简单，不需表值，但此法数据上不严格，目前使用不多，仅在要求不高，测定次数少时使用。

(2) Q 法符合统计原理，计算方便，所以常采用，但此法将可疑限订得太低，所以有时会过多保留异常值。适合于只有1个可疑值的检验。

(3) G 法是最合理，舍取效果最好，使用最普遍的方法，但计算麻烦，当其他方法与 G 法发生矛盾时，以后者为主。

(4) 当测量次数过少， Q ， T 与表中的临界值接近时，应补测1~2个数据；如无条件补测，数据又较分散时，可以中位数报告结果。



统计检验的正确顺序

可疑数据取舍



F 检验



t 检验



例题：用 $\text{K}_2\text{Cr}_2\text{O}_7$ 基准试剂标定 $\text{Na}_2\text{S}_2\text{O}_3$ 溶液的浓度 ($\text{mol}\cdot\text{L}^{-1}$)，4 次结果分别为：0.1029，0.1026，0.1032 和0.1034。

(1) 用格鲁布斯法检验上述测定值中有无可疑值 ($\alpha=0.05$)；(2) 计算置信度为95% 时的置信区间？

解：(1) 排序：0.1026，0.1029，0.1032，0.1034。

计算平均值及标准偏差：

$$\bar{x} = \frac{0.1026 + 0.1029 + 0.1032 + 0.1034}{4} = 0.1030, s = 0.0004$$

$$\text{先检验 } x_1, T = \frac{\bar{x} - x_1}{s} = \frac{0.1030 - 0.1026}{0.0004} = 1.00$$

查T值表可知： $T_{0.05,4} = 1.46$ $T < T_{0.05,4}$ 留



检验 x_n ,
$$T = \frac{x_n - \bar{x}}{s} = \frac{0.1034 - 0.1030}{0.0004} = 1.00$$

查T值表可知: $T_{0.05,4} = 1.46$ $T < T_{0.05,4}$ 留

(2) 计算置信区间

分析结果为:
$$\bar{x} = \frac{0.1026 + 0.1029 + 0.1032 + 0.1034}{4} = 0.1030, s = 0.0004, n = 4$$

置信度为95%, 自由度为 $f = n - 1 = 4 - 1 = 3$ $t_{0.05,3} = 3.18$

$$\mu = \bar{x} \pm \frac{t_{0.05,3} s}{\sqrt{n}} = 0.1030 \pm \frac{3.18 \times 0.0004}{\sqrt{4}} = 0.1030 \pm 0.0006 (P = 95\%)$$



例题：某分析人员提出了一新的分析方法，并用此方法测定了一个标准试样，得如下数据(%)；40.15，40.00，40.16，40.20，40.18。已知该试样的标准值为40.19%，(1)用 Q 检验法判断40.00%是否应该舍弃？(P=90%) (2)试用 t 检验法对新分析方法作出评价。(α=0.05)

解：(1) 排序：40.00，40.15，40.16，40.18，40.20，

$$Q = \frac{x_2 - x_1}{x_n - x_1} = \frac{40.15 - 40.00}{40.20 - 40.00} = 0.75$$

查 Q 值表可知：P=90%， $n = 5$ 时， $Q_{表} = 0.64$

因为 $Q > Q_{表}$ ，舍。



分析结果为: $\bar{x} = \frac{40.15 + 40.16 + 40.18 + 40.20}{4} = 40.17$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = 0.022\% \quad n = 4$$

$$\mu = \bar{x} \pm \frac{t_{0.05,3} s}{\sqrt{n}} = 40.17\% \pm \frac{3.18 \times 0.022\%}{\sqrt{4}} = (40.17 \pm 0.03)\% (P = 95\%)$$

(2) 准确度检验, 与标准值比较。

$$\bar{x} = 40.17\%, s = 0.022\%, n = 4 \quad \mu = 40.19\%$$

$$t = \frac{|\bar{x} - \mu|}{s} \times \sqrt{n} = \frac{|40.17\% - 40.19\%|}{0.022\%} \times \sqrt{4} = 1.82$$

$$t_{0.05,3} = 3.18 \quad t < t_{0.05,3}$$

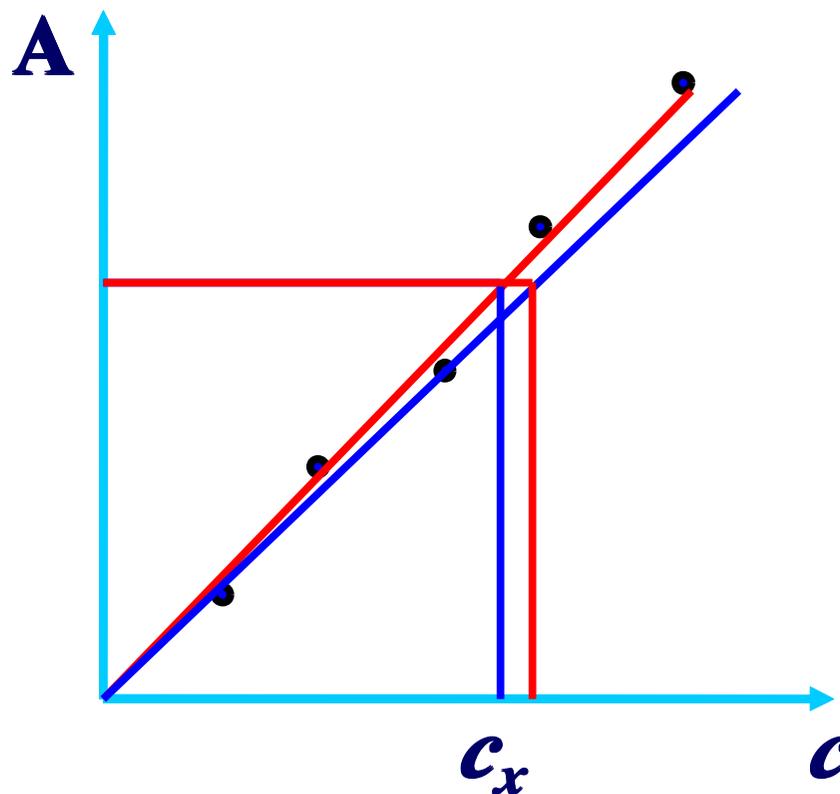
所以新方法没有系统误差, 新方法是准确可靠的。



3.6 回归分析法

$$A = \varepsilon bc = kc$$

No.	标样浓度 $\mu\text{g} / \text{L}$	吸收值
1	5.00	0.045
2	10.0	0.093
3	20.0	0.140
4	30.0	0.175
5	40.0	0.236
6	试样	0.155



每个测量值都有误差，标准曲线应怎样作才合理？



3.6.1 一元线性回归方程及回归直线

$$y = a + bx$$

$$(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$$

$$Q_i = [y_i - (a + bx_i)]^2$$

$$Q = \sum_{i=1}^n [y_i - (a + bx_i)]^2$$

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0$$

$$\frac{\partial Q}{\partial b} = -2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0$$

$$a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} = \bar{y} - b\bar{x}$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$



例1：用吸光光度法测定合金钢中Mn的含量，结果如下：

Mn/ μg	0.00	0.02	0.04	0.06	0.08	0.10	0.12	未知样
A	0.032	0.135	0.187	0.268	0.359	0.435	0.511	0.242

列出标准曲线的回归方程，计算未知样中Mn的含量？

解：此例中Mn的含量为自变量 x ，吸光度为因变量 y 。

$$\bar{x} = 0.06 \quad \bar{y} = 0.275 \quad n = 7$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 3.95 \quad a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} = \bar{y} - b\bar{x} = 0.038$$

$$y = a + bx = 0.038 + 3.95x \quad m_{\text{Mn}} = \frac{0.242 - 0.038}{3.95} = 0.052 \mu\text{g}$$



回归直线的特点

- ① 它必定通过 (\bar{x}, \bar{y}) 点；
- ② 对所有实验点而言，此线的误差最小；
- ③ 它也许不过任何一个实验点（与直尺作图习惯不同）。

求回归方程时应注意的几个问题

- ① 不要过早修约数字，应在获得a、b具体值后再合理修约；
- ② b的有效数字与x相等，a与y相等（最多多一位）；
- ③ 回归计算较烦易错，所以最好验算。

$$\sum_{i=1}^n y_i = na + b \sum_{i=1}^n x_i$$



3.6.2 线性相关系数 Correlation coefficient

$$r = b \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

相关系数的意义

- ① $r = \pm 1$ ，所有点都落在回归线上，完全相关。
- ② $r = 0$ ， y 与 x 之间无线性关系，回归线无意义。
- ③ $0 < r < 1$ ，有线性关系， r 越大，线性越好。要根据测量的次数及置信水平与相应的相关系数临界值比较，绝对值大于临界值时，则可认为这种线性关系是有意义。



判断上例中回归线的线性关系如何 (P=99%) ?

解: $n = 7$

$$r = b \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = 0.9993$$

P = 99%时, $r_{99\%, 5} = 0.875 < r_{\text{计}}$

标准曲线具有良好的线性关系。



3.7 提高分析结果准确度方法

1 选择恰当分析方法

化学分析法准确度高，灵敏度低，适合高含量组分的测定
仪器分析法灵敏度高，准确度高，适合低含量组分的测定

分析方法	组分相对含量
常量组分分析 (major)	> 1%
微量组分分析 (micro)	0.01% ~ 1%
痕量组分分析 (trace)	< 0.01%
超痕量组分分析 (ultratrace)	~ 0.0001%



2 减小测量误差

- ① 合理的控制被测量量值的大小，减小测量误差；
- ② 各测量值的误差应与分析方法的误差相适应。

3 消除系统误差

① 系统误差的检验

A 对照试验

与标准试样对照

与标准方法对照

B 加标回收试验

$$R = \frac{B-x}{A}$$

常量 > 99%

微量 90% - 110%



② 消除系统误差

A 空白试验

B 校正仪器

C 校正分析结果

4 减小随机误差

增加平行测定的次数，减小随机误差。

一般分析：3-4次

精密分析：5-9次

要求更高：10-12次



本次作业

P₇₅ : 5, 6, 8, 9, 11, 12, 15, 19, 20, 21



例题：

1 当测定次数很多时，下列关于标准偏差和平均偏差间的关系，那一个正确？

A $\sigma > \delta$, B $4\sigma = 3\delta$, C $\sigma = 0.8\delta$, D $\delta = 0.8\sigma$

2 下列关于随机误差正态分布曲线特点的论述中，错误的是：

A 横坐标 x 等于总体平均值 μ 时，曲线有极大；

B 曲线与横坐标所夹面积的总和，等于所有测量值出现的概率，其值为1

C 分布曲线以 μ 值的横坐标为中心左右对称，说明正负误差出现的概率相等；

D 纵坐标 y 值代表概率，它与标准偏差 σ 成正比， σ 越小，曲线越平坦。



3 下列关于n次测定结果的平均值的标准偏差与单次测定结果的标准偏差间的关系，正确的是？

$$A. s_{\bar{x}} = \frac{s}{\sqrt{n}}; \quad B. s_{\bar{x}} = \frac{s}{n}; \quad C. s = \frac{s_{\bar{x}}}{\sqrt{n}}; \quad D. s_{\bar{x}} > s$$

4 下列关于平均值置信区间的定义中，正确的是？

A 以总体平均值为中心的某一区间包括测定结果的平均值的概率；

B 在一定置信度下，以测定结果的平均值为中心的包括总体平均值的可靠性范围；

C 总体平均值落在某一区间的概率；

D 在一定置信度下，以总体平均值为中心的包括测定结果的可靠性范围；



5 下列关于平均值置信区间的论述中，错误的是？

A 在一定的置信度和标准偏差时，测定次数越多，平均值的置信区间窄，平均值就越准确；

B 其他条件不变时，给定的置信度越高，平均值的置信区间越宽；

C 平均值的数值越大，置信区间越宽；

D 当置信度与测定次数一定时，测定结果的精密度越高，则平均值的置信区间就越窄。

6 有一组测量值，总体标准偏差未知，要判断得到这组数据的分析方法是否可靠，应用哪一种方法？

A 4d； B Glubbus法； C F检验法； D t检验法



7 两个分析人员分析同一试样，得到两组数据，要判断两人的精密度有无显著性差异，应用哪种方法？

A Q检验法； B u检验法； C F检验法； D t检验法

8 对上题，若要判断有无系统误差，应用哪种方法？

A u检验法； B F检验法； C F+t 检验法； D t检验法

9 t分布曲线和u分布曲线的区别？

u分布——描述无限次测量数据

t 分布——描述有限次测量数据

u分布——横坐标为 u ，

t 分布——横坐标为 t

$$u = \frac{x - \mu}{\sigma} \quad t = \frac{x - \mu}{s}$$

两者所包含面积均是一定范围内测量值出现的概率；

u分布：概率随u 变化； u 一定，概率一定；

t分布：概率随 t 和 f 变化； t 一定，概率与 f 有关

$f \rightarrow \infty$ 时， t 分布 \rightarrow u 分布



- 10 写出一元线性回归方程的a, b及r的计算公式?
- 11 已知某试样中含Co的标准值为1.75%, 标准偏差 $\sigma=0.10\%$, 求分析结果落在 $1.75\% \pm 0.15\%$ 范围内的概率?
- 12 用 $K_2Cr_2O_7$ 基准试剂标定 $Na_2S_2O_3$ 溶液的浓度 ($mol \cdot L^{-1}$), 4次结果分别为: 0.1029, 0.1010, 0.1032 和0.1034。(1) 用格鲁布斯法检验上述测定值中是否有可疑值 ($\alpha=0.05$); (2) 计算置信度为95%时的置信区间?
- 13 某分析人员提出了一新的分析方法, 并用此方法测定了一个标准试样, 得如下数据(%); 40.15, 40.00, 40.16, 40.20, 40.18。已知该试样的标准值为40.19%, (1) 用Q检验法判断极端值是否应该舍弃? (2) 试用t检验法对新分析方法作出评价。 ($\alpha=0.05$)