

化学分析数据处理

一.误差

1 误差和偏差

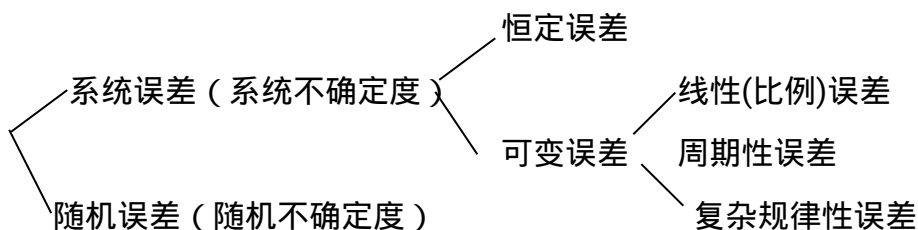
1.1 误差: 误差 = 真值 - 测定值

真值可以是已知标准值,在消除系统误差以后,可以认为是测定次数为 ∞ 的平均值。

1.2 偏差: 偏差 = 均值 - 测定值

2.实验误差的种类

2.1 种类



过失误差 (粗差)

2.2 系统误差和随机误差的异同点

2.2.1 相同点: 都是误差, 它们都存在于一切科学实验中, 只能减弱到一定程度, 而无法彻底消除。

2.2.2 不同点:

系统误差	随机误差
具有确定性。在相同条件下, 多次测定同一值时, 误差的绝对值和符号保持恒定。条件改变时, 误差也按确定规律变化。	具有随机性, 在相同条件下, 多次测定同一量时, 误差的绝对值和符号以不可预定的方式变化。
单项系统误差多与单个或少数几个因素有关。	多由大量均匀小的因素共同影响所造成。
无抵偿性	有抵偿性

影响系统误差的条件一经确定，误差亦随之确定，即使重复测定误差的绝对值和符号保持不变。	同条件重复测定，求均值，可减少随机误差。
--	----------------------

2.2.3 消除方法

- a.减小系统误差的方法：对照实验、空白实验、仪器校正、分析结果校正。
- B.减小随机误差的方法：利用随机误差的抵偿性，多次重复测定求均值，使之减弱到一定程度。
- C.过失误差：如写错报告，算错结果等。对分析人员要进行素质训练；分析结果要相互审查，以免出错。

二.精密度

表示分析数据的重复性，其大小由随机误差所决定。

1.随机误差的表示方法

1.1 全距 (R)：亦称极差。

$$R = \text{极大值} - \text{极小值}$$

1.2 标准差 (S)：它是估计测定值偏离最佳值(真值或均值)程度的一个常用方法。

$$S = \left[\frac{\sum (x - \bar{X})^2}{n - 1} \right]^{1/2} = \left[\frac{\sum x^2 - (\sum x)^2 / n}{n - 1} \right]^{1/2}$$

1.3 相对标准差(RSD%)：又称变异系数 (C.V %)。

$$RSD\% = S \times 100 / \bar{X} (\%)$$

2.种类

2.1 连续测定精密度：同一时间对同一样品进行多次重复测定的精密度。

例 2000年10月11日上午测定某种酒中甲醇的含量，共测定5次，结果(g/100ml)如下：

0.021 0.019 0.022 0.020 0.022

$$\bar{X} = 0.0208 \quad S = 0.0013 \quad CV\% = 6.2$$

2.2 重复性精密度: 在相同条件下, 不同时间内对同一样品进行测定的精密度。

例: 测定某种酒中的异丁醇含量。连续测定 7 天, 每天测定一次, 其结果 (g/100ml) 如下:

0.029 0.032 0.027 0.032 0.027 0.028 0.030

$$\bar{X} = 0.0293 \quad S = 0.0021 \quad CV\% = 7.2$$

2.3 再现性精密度: 用同一方法, 不同实验室对同一样品进行重复测定的精密度。

例: 把一个酒样分成 5 份, 分发给 5 个实验室用同一方法测定酒中异戊醇的含量, 结果如下:

0.040 0.035 0.042 0.037 0.047

$$\bar{X} = 0.402 \quad S = 0.0047 \quad CV\% = 12$$

三. 准确度

表示分析结果同真值的符合程度, 其大小由系统误差所决定。系统误差主要有两种表现形式: 一是恒定误差, 另一种是线性误差(比例误差)。主要来源于测定方法本身、仪器、试剂和操作者、样品的基底成分四个方面。它们是影响准确度的主要原因。确定分析中是否存在系统误差, 并采用相应的消除方法, 可以提高分析结果的准确度。

1. 绝对误差和相对误差

绝对误差 = 真实值(准确值) - 测定值

绝对误差

相对误差 = $\frac{\text{绝对误差}}{\text{真实值}} \times 100\% \text{ (RE\%)}$

真实值

例: 测定值	准确值	误差	相对误差%
0.095	0.097	+0.002	2

2. 回收率

分析样品时, 准确值是不能知道的。常用加标回收的方式来估计测定值的准

确程度。

本底加标后测定值 - 本底值

$$\text{回收率}\% = \frac{\text{本底加标后测定值} - \text{本底值}}{\text{加标量}} \times 100$$

须说明的是,加入标准回收的方法并不是一个十分可靠的方法,有时回收率 100%,测定值仍可存在系统误差。

3.用标准物质进行对照实验

3.1 选择标准物质的一种浓度与样品作平行测定,如果标准物质的分析结果 ($\bar{X} \pm t \times S / (n)^{1/2}$) 与证书上所给的保证值 ($A \pm U$) 一致。则表明分析测定过程中不存在明显的系统误差。如试剂百分含量的测定,药品纯度测定,质控方法等。

3.2 在方法的线形范围内选择三种以上不同浓度的标准物质,按分析样品的方法和程序去测定,然后用 a、b 的置信区间 $a \pm t_{0.05(n-2)} S_a$ 和 $b \pm t_{0.05(n-2)} S_b$ 来分析方法是否存在系统误差。

示例:重量法测定 Al 的数据

取 Al_2O_3 (g) X	测得 Al_2O_3 (g) Y	相差 (g)	RE%
0.1077	0.1288	0.021	195
0.2154	0.2384	0.023	107
0.3231	0.3489	0.026	80
0.4308	0.4588	0.028	65

$$r=0.999999 \quad a=0.0186 \quad b=1.022$$

$$L_{xx} = [\sum X^2 - (\sum X)^2 / n] \quad L_{xy} = \sum X \cdot \sum Y - (\sum X \cdot \sum Y)^2 / n$$

$$L_{yy} = [\sum Y^2 - (\sum Y)^2 / n]$$

$$S_{x,y} = \{[\sum (Y - \hat{Y})^2] / (n-2)\}^{1/2} = [(1-r^2)L_{yy} / (n-2)]^{1/2}$$

$$= [(L_{yy} - L_{xy} \cdot b) / (n-2)]^{1/2}$$

$$S_a = S_{x,y} (1/n + X^2 / L_{xx})^{1/2}$$

$$S_b = S_{x,y} (L_{xx})^{-1/2} \quad r = L_{x,y} / (L_{xx} \cdot L_{yy})^{1/2} \quad a = Y - bX$$

$$L_{xx} = 0.057996 \quad L_{yy} = 0.060555$$

$$S_{xy} = 0.00025981 \quad S_a = 0.0003 \quad S_b = 0.001$$

$$a \pm t_{0.05(2)} S_a \quad 0.0186 \pm 0.0013 (0.0173 - 0.0199)$$

$$b \pm t_{0.05(2)} S_b \quad 1.022 \pm 0.0044 (1.018 - 1.026)$$

a 的置信区间包括 0，b 的置信区间包括 1，才能肯定测定无系统误差。若 a 的置信区间不包括 0，说明测定有恒定误差，b 的置信区间不包括 1，说明测定有比例误差。从上例重量法测定 AI 的数据的计算结果可知，分析既有恒定误差又有比例误差。

4. 用标准方法(A)来确定分析方法(B)有无系统误差。

用标准方法或经典方法 (A) — 不含系统误差的方法和分析方法 (B) 同时测定多个不同浓度的样品，然后对方法 A 和 B 的测定结果进行回归分析。如果 a 的置信区间包括 0，b 的的置信区间包括 1，可在 95% 置信水平上断定，方法 A 和方法 B 具有相同的准确度。

例如，用标准方法 A 和分析方法 B 测定五个不同浓度的检品，结果如下：

A(X)	0.226	0.472	0.943	1.886	3.772	ppm
B(Y)	0.224	0.461	0.988	1.890	3.896	ppm
n=5	r=0.99983	a=-0.0170	b=1.034			
L _{xx} =8.29301		L _{yy} =8.86242				
S _{xy} =0.03199	S _a =0.0216	S _b =0.01111	t _{0.05(3)} =3.182			
a ± t _{0.05(3)} S _a	-0.0170 ± 0.0687 (-0.0857 - 0.0517)					
b ± t _{0.05(3)} S _b	1.034 ± 0.0353 (0.999 - 1.069)					

因 a 的置信区间包括 0，b 的的置信区间包括 1，可在 95% 置信水平上断定，方法 A 和方法 B 具有相同的准确度。由方法 A 推知方法 B 不存在系统误差。

5.t 检验方法

若两种方法对不同样本测定结果无显著差异，可以认为这两种方法具有相同的准确度。

四.有效数字的运算

1.有效数字的意义

有效数字反映测量的可靠程度，即所报告的各位数字除末位外，皆为准确知道的，只有最后一位数字是可疑数，末尾第二位是可靠数，可疑数以后是无意义数。分析报告的数字只能在方法的准确度以内。

如报告酒中己酸乙酯含量是 1.4g/L，表明分析者对 1 是肯定的，0.4 是不确定的，是 1.3 或 1.5 可能不好说。

2.有效数字的修约

当有效位数确定以后，多余的数字一律按“四舍六入五成双”的原则处理。

酒样号	甲醇 (g/100ml) 计算值	甲醇 (g/100ml) 报告值
1	0.02249	0.022
2	0.02251	0.023
3	0.02349	0.023
4	0.02350	0.024

对多位数不得进行连续修约，如上例 3：

$$0.02349=0.0235=0.024 \quad \times$$

0.02349 应一次修约成 0.023

3.有效数的一般运算规则

3.1 加减时：保留小数点后位数，应与各数中小数位数最少者相同。

例：

乙酸乙酯 (g/L) 1.37

甲酸乙酯 (以乙酸乙酯计, g/L) 0.0082

乳酸乙酯 (以乙酸乙酯计, g/L) 0.16

己酸乙酯 (以乙酸乙酯计, g/L) 0.92

总酯 (以乙酸乙酯计, g/L) 2.46(不得写成 2.4582)

3.2 乘除时：保留位数以有效数字位数最少者为准。

例 1：称甲醇 0.200g，用 60%乙醇稀释成 100.0mL，混匀。取 10.0mL 稀释 100 倍成标准应用液，求应用液甲醇含量(g/100mL)。

$$(0.200/100.0) \times 10.0 \times (1/1000) \times 100 = 0.0200\text{g}/100\text{mL}$$

例 2：用微量注射器取 2.0uL 正丙醇于 10.0mL 比色管（内装少量 60%乙醇）中，用 60%乙醇稀释至刻度，求正丙醇的含量（g/100mL）。

$$(2.0 \times 0.804 \times 99\%) / 10.0 \times (1/1000) \times 100 = 0.016\text{g}/100\text{mL}$$

3.3 在计算式中，常数 π 、e、 $n^{1/2}$ 、 $1/n$ 、n 等数，其有效数可以认为是无限制的，需要几位可写几位。

3.4 在计算平均数时，均数的有效数字可增加一位。

3.5 表示分析方法的精密度、准确度，大都取 1—2 位有效数字。

五.可疑数字的处理

1.求均值时可疑值的取舍

1.1 三倍标准法：若分析数据呈正态分布，可用下列公式确定极端值是否可以舍弃。

|极端值 - 不包括极端值在内的均值|

$$\frac{\quad}{\quad} > 3, \text{舍去};$$

$$\quad < 3, \text{保留}。$$

例:0.040 0.035 0.042 0.037 0.047

除去 0.035, $\bar{X} = 0.0415$, $S = 0.0042$

$$\frac{|0.035 - 0.0415|}{\quad}$$

0.0042 = 1.5，说明 0.035 应保留，并参加均值计算，

$\bar{X} = 0.0402$

1.2 偏差与极差比值法

1971 年由 E LBauer 提出，又称 Baure 检验法

求观察值舍弃的临界值，如计算 t_i 值大于表值，则舍弃。

$$t_i = |X - \bar{X}| / R$$

观察值舍弃的临界值 t_i (P=0.95)

n	3	4	5	6	7	8	9	10	20
t_i	1.53	1.05	0.86	0.76	0.69	0.64	0.60	0.58	0.46

$$t_i = |0.035 - 0.0402| / (0.047 - 0.035)$$

$$= 0.433 \quad (n=5, t_i=0.433 < 0.86, P>0.95, 0.035 \text{ 应保留})$$

1.3 此外还有 Dixon 法 (1950)、David 法、Grubbs 法 (1959)

2. 标准曲线中异常值的取舍

在化学分析中, 根据浓度 X 和峰高 (或峰面积或吸收值) Y 的对应关系, 求出回归方程的剩余标准差 $S_{y,x}$, 若 Y 值在 $Y \pm t_{0.05(n-2)} S_{y,x}$ 范围内应保留, 未在此范围内应舍弃。舍弃异常值后再求回归方程, $Y = a + bX$, 并计算结果。

六. 平行样的允许误差

定量分析中, 人们通常用作平行样的方法测出样品中待测物的含量, 同一样品做多次测定, 可以参照资料舍弃异常值, 并按“1/4S”法确定均值的有效数字。《空气中有毒物质的测定方法》规定, 每一个采样点必须同时采集两份样品, 测定结果之差不得超过 20%, 国家地质总局对于一些测定项目平行样结果的误差范围作了规定。也有用 Rc 控制图或按标准曲线浓度范围来判断的。作者参考资料提出一种平行样允许误差评判新方法, 能避免人为因素的影响, 适宜所有采用标准曲线定量的分析方法。

我们知道, 两平行样的允许误差受方法、仪器设备、操作水平、样品均匀性等因素的影响。假定样品均匀, 只有随机误差影响测定结果, 用标准曲线上各浓度点的波动来确定平行样的最大理论相对误差并与之比较, 决定其是否求均值予以报告或舍弃重作。下面举一个例子说明。

二氮杂菲比色法测定水中总铁结果见表 1、表 2

表 1 铁含量与最大理论相对误差

Fe (mg/L)	0.050	0.10	0.20	0.40	0.60	0.80	1.00
-----------	-------	------	------	------	------	------	------

A	0.010	0.020	0.040	0.075	0.118	0.158	0.202
RE% (极值)	62	31	16	7.8	5.2	3.9	3.1

C 波动极值计算方法：

$$S_{C,A} = \{[\sum(A - \bar{A})^2]/(n-2)\}^{1/2} = [(1-r^2)L_{AA}/(n-2)]^{1/2}$$

$$= [(L_{AA} - L_{CA} \cdot b)/n-2]^{1/2} = 0.2004 ; t_{0.05(5)} S_{C,A} = 2.571 \times 0.0024 = 0.0062$$

$$L_{AA} = [\sum A^2 - (\sum A)^2/n] ; L_{CA} = \sum C \cdot \sum A - (\sum C \cdot \sum A)^2/n$$

$$C \text{ 波动极值} = \pm t_{0.05(5)} S_{C,A} / b = 0.0062 / 0.2004 = \pm 0.031 \text{ mg/L}$$

$$RE\% \text{ 极值} = (C \text{ 波动极值} / C) \times 100$$

表 2 水样中总铁测定结果

水样	平行样测定值		RE%	均值
1	0.060	0.081	-30%	0.070
2	0.060	0.097	-47%	--
3	0.176	0.206	-16%	0.19
4	0.630	0.600	+4.9%	0.62
5	0.312	0.471	-41%	--
6	0.754	0.704	+6.9%	--

表 2 中 2、5、6 号水样平行结果 RE% > RE% 理论极值，不能计算均值，应重作实验；其余均值按 " 1/4 S " 法求得。

用标准曲线上 A 的波动范围所对应的 C 的上、下限范围来限制平行样的允许误差有统计意义。样品操作环节多，影响因素多，其波动一般比标准品波动大，用 $S_{C,A}$ 代 S_A ($S_{C,A} < S_A$, $S_A = [1 + 1/n + (C - \bar{C})^2 / L_{CC}]^{1/2} \times S_{C,A}$ 来判断，对标准曲线低浓度及高浓度部分测定结果的相对误差要求更严。从表 1 低浓度看，0.1 C (将标准系列中最高浓度定为 C) 理论上最大允许误差已达 31%；显然按方法不分测定浓度大小均控制相对误差小于 20% 要求太高。两平行样的允许误差应根据自己使用的仪器，设备，操作条件，按统计学计算出 C 的波动值求得理论上的允许误差作为两平行样的最大允许误差。

有资料证明：在不考虑系统误差的前提下，真值在 C_1 和 C_2 之间的置信度为 50%，真值在 C_1+R 与 C_2-R 之间 ($C_1 > C_2$) 的置信度为 80%，因此，两平行样的允许误差小于由标准曲线计算出来的理论误差时，计算的样本均值作为微量分析、痕量分析是可以接受的。

七.方法的检出限和最低检出量

1.检出限 X_{dl}

根据 INPAC (国际纯理论化学和应用化学联合会) 的建议：

$$X_{dl} = \bar{X}_0 + 3S_0$$

X_{dl} 某方法的检出限； \bar{X}_0 某方法的空白均值 (>20 次)； S_0 某方法的空白标准差。

2.最低检出量 (浓度) $3S_0$

3.未检出量 (ND)：可以理解为最低检出量的半量或更小量，常用 $3S_0/2$ 表示。

例：已知某方法的回归方程为 $Y=0.029X+0.005$ ，今测得该方法的 21 次空白吸光度值(Y)为：

0.115 0.125 0.095 0.085 0.098 0.135 0.125
0.125 0.119 0.110 0.102 0.108 0.132 0.127
0.092 0.099 0.112 0.124 0.127 0.109 0.102

计算该方法的检出限,最低检出量,未检出量各为多少微克?

求得 $\bar{Y}=0.113$ $S=0.0143$ 代入回归方程

$$X=(Y-0.005)/0.029 \quad \bar{X}_0=3.72\mu\text{g}, S_0=0.32\mu\text{g}$$

因此: $X_{dl}=3.72+3\times 0.32=4.7\mu\text{g}$

$$3S_0=3\times 0.32=0.96=1.0\mu\text{g}$$

$$ND=3S_0/2=0.48=0.5\mu\text{g}$$

八.回归方程

1.直线相关与回归

比色分析中,线形范围内浓度 X 和吸光度值 Y 之间的关系可用 $Y=a+bX$ 表示.这个方程称直线回归方程。X 和 Y 两个变量间直线关系的密切程度和相关方向用相关函数(r)来表示。

$1>r>0$ 正相关; $r=1$ 完全正相关; $0>r>-1$ 负相关; $r=-1$ 完全负相关; $r=0$ 无相关关系.

2.回归方程的截距 ,斜率 b 及相关系数 r 的计算方法

$$r=(\sum xy-\sum x\sum y/n)/\{[\sum x^2-(\sum x)^2/n]\times[\sum y^2-(\sum y)^2/n]\}^{1/2}$$

$$b=(n\sum XY-\sum X\sum Y)/[n\sum X^2-(\sum x)^2]$$

$$a=Y-bX$$

$$S_{y,x}=\{[\sum (y-y)^2]/(n-2)\}^{1/2}(\text{回归方程的剩余标准差,表示 Y 的变异程度})$$

有些资料上用 b 表示方法的灵敏度。ISO (国际标准化组织) 定义 $b/S_{y,x}$ 为方法灵敏度。由此可以看出方法灵敏度随着工作曲线斜率的增大而提高 , 也随着 Y 的分散性减少而提高。

3.作图方法及应用

例：用品红亚硫酸比色法测定酒中甲醇

标准系列(mg)	0	0.20	0.40	0.60	0.80	1.0
A	0.005	0.035	0.072	0.108	0.154	0.202

54° 某酒样 , 取样 0.60ml , 按方法测得 $A=0.025$, 求甲醇含量。

3.1 直接在坐标纸上描点:将标准系列各对应点在坐标纸上描点,用目测法画标准曲线 , 然后根据样品吸光值 , 查得对应含量。

3.2 用计算器求得回归方程 $a=-0.0024,b=0.197$, 利用回归方程

$$Y=-0.0024+0.197x$$

$$\text{求得:}X=0,Y=-0.0024$$

$$X=1,Y=0.195$$

直接在坐标纸上连接 (0,-0.0024) 和 (1.0,0.195) 两点 , 即得此标准曲线。计算样品含量同 (1)。

3.3 利用回归方程直接计算样品含量。

$$Y=-0.0024+0.197X$$

得 $X=(Y+0.0024)/0.197=(0.025+0.0024)/0.197=0.14\text{mg}$

即酒中甲醇含量为:

$$(0.14\text{mg}/0.60\text{ml})\times(1/1000\text{mg/g})\times 100\text{mL}\times(60/54)=0.026\text{g}/100\text{mL}$$

4.回归计算要正确扣除空白值

在用回归方程进行定量计算的分析中,常涉及到扣除空白值计算样品含量的问题。一个被分析的样品,如计算中扣除空白的方法不当,就会出现错误的结果。所以,正确扣除空白值是一个人们应该普遍注意的不容忽视的问题。因此,有必要搞清它们之间的关系,以避免由于计算方法错误造成的偏差。

4.1 标准未经消化、萃取等前处理的方法:不论是用标准的零浓度调节吸光值 $y=0$,还是用水或纯溶剂调节吸光值 $y=0$ 来测定标准、样品、空白的吸光值,在用回归方程计算样品含量时,均应以 $y-y_0+a$ 代入回归方程求值或用 $(y-y_0)/b$ 求值。不能用 $y-y_0$ 代入回归方程求值。

4.2 标准与样品都经前处理或都未经前处理的方法,不论是用空白或标准的零浓度调节吸光值 $y=0$,还是用水或纯溶剂调节吸光值 $y=0$ 来测定标准、样品的吸光值,在用回归方程计算样品含量时,均应以 y 代入回归方程求值,即 $x=(y-a)/b$ 。

4.3 以上结论同时适用于其它类似的分析方法,如比色分析法、原子吸收分光光度法、极谱分析法、气相色谱法、液相色谱法等。

九.质量管理控制图

1.概述

在检测过程中,影响测定结果准确度的主要原因有:人、仪器设备、样品基底成分、测定方法、测试环境等方面。对测定过程进行质量控制,就是控制这些因素造成的异常波动,使测定处于稳定良好的状态,从而保证检测数据准确可靠。

1.1 控制图 控制上限 (UCL) 中心线 (CL) 控制下限 (LCL)

1.2 控制图的分类

a 按用途分:分析用控制图、管理用控制图.

b 按统计量分:计量值控制图(单值控制图 - \bar{X} 控制图;均值与极差控制图 - \bar{X} -R,中位数与极差控制图 - \bar{X} -R);计数值控制图(不合格品率控制图 - P 控制图,不合格件数控制图 - P_n 控制图,缺陷数控制图 - C 控制图,单位缺陷数控制图 - U 控制图)。

2.单值控制图 (\bar{X} 控制图)

$CL = \bar{X}$, $UCL = \bar{X} + 3S$, $LCL = \bar{X} - 3S$, (LCL 如为负数,无意义,可以不考虑)

3.控制图的观察与分析

3.1 点在中心线一侧连续出现 7 次以上,测定有异常 ($P=1.6\%$)

3.2 点在中心线一侧多次出现,如连续 11 点中有 10 点,连续 14 点中有 12 点,连续 20 点中有 16 点处于同一侧,测定有异常。

3.3 连续有 7 点以上上升或下降,测定有异常。

3.4 连续 3 点中至少有 2 点偏离中心线的距离在中心线与管理界限间距离的 $2/3$ 以上,测定有异常。

3.5 点子保持周期性变化,一般认为测定有异常。