Chapter 2

误差及分析数据的统计处理

Errors and statistical Treatment of Analytical Data

主要内容

2.1 定量分析中的误差

2.2 分析结果的数据处理

2.3 误差的传递

2.4 有效数字及其运算规则

2.5 标准曲线的回归分析

2.1 定量分析中的误差

2.1.1 误差与准确度

准确度（accuracy)

表征测定值和真实值之间的符合程度

真实值

理论真值

计量学约定真值

相对真值

准确度的表征：误差

误差

绝对误差

$$E = x_i - \mu$$

相对误差

$$E_r = \frac{x_i - \mu}{\mu}$$

2.1.2 偏差与精密度

精密度（precision）

表示各次分析结果相互接近的程度

偏差 $d_i = x_i - \bar{x}$

相对偏差 $d_r = \dfrac{x_i - \bar{x}}{\bar{x}} \times 100\%$

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

精密度的表征：偏差

$$\text{平均偏差}\bar{d}=\frac{|d_1|+|d_2|+\cdots+|d_n|}{n}$$

$$\text{相对平均偏差}=\frac{\bar{d}}{\bar{x}}\times 100\%$$

例2.1：有两组测定值

甲组：2.9 2.9 3.0 3.1 3.1

乙组：2.8 3.0 3.0 3.0 3.2

判断精密度的差异。

解：

平均值	平均偏差	标准偏差
$\bar{x}_{甲}=3.0$	$d_{甲}=0.08$	$s_{甲}=0.10$
$\bar{x}_{乙}=3.0$	$d_{乙}=0.08$	$s_{乙}=0.14$

在偏差的表示中，用标准偏差更合理，因为将单次测定值的偏差平方后，能将较大的偏差显著地表现出来。

样本标准偏差(测量次数有限）

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

$n-1$:自由度，

具有独立偏差的数目

相对标准偏差，即变异系数

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

总体标准偏差，$n \to \infty$

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

μ：总体平均值

精密度

重复性(r)：同一操作者，在相同条件下，获得一系列结果之间的一致程度。又称室内精密度。

$$r = 2\sqrt{2}S_r$$

再现性(R)：不同操作者，在不同条件下，用相同方法获得的单个结果之间的一致程度。又称室间精密度。

$$R = 2\sqrt{2}S_R$$

$$S_R = \sqrt{\frac{\sum_{j=1}^{m}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{m(n-1)}}$$

m:参加测定的实验室数

n:每个实验室重复测定次数

验证情况

样品基质	添加水平（μg/kg）	测定次数										平均测定值（μg/kg）	标准偏差（μg/kg）	相对标准偏差（%）	平均回收率（%）
		单位1		单位2		单位3		单位4		单位5					
		1	2	3	4	5	6	7	8	9	10				
基质1	50	42.22	49.1	39.2	42.2	42.2	38.3	44.5	38.7	47.85	42.02	42.63	3.6	8.5	85.2
	100	92.1	105.2	82.4	90.4	96.2	98.7	86.4	78.6	98.1	92.2	92.03	8.0	8.7	92.0
	500	469.2	445.6	475.4	456.8	466.3	435.2	445.6	467.8	450	426.2	453.81	16.0	3.5	90.7
基质2	50	38.34	41.36	42.4	48.6	45.4	48.2	47.5	41.8	39.12	44.34	43.71	3.6	8.4	87.4
	100	86.2	94.2	104.2	108.6	86.2	81.9	105.6	96.8	106	102.3	97.20	9.6	9.9	97.2
	500	515.2	508.8	501.2	485.4	530.2	512.6	511.2	532.1	482.2	495.4	507.43	16.7	3.3	101.4
基质3	50	46.26	51.3	50.2	53.3	52.1	50.9	52.2	50.9	46.34	54.12	50.76	2.6	5.1	101
	100	83.2	95.3	86.2	93.4	98.1	108.6	85.8	93.6	85.2	94.2	92.36	7.6	8.2	92.3
	500	458.6	432.4	428.8	438.6	439.4	429.8	455.2	438.5	530.6	510.2	456.21	35.5	7.7	91.2

2.1.3 准确度与精密度

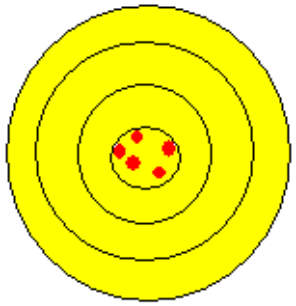
- **准确度 Accuracy**

  准确度表征测量值与真实值的符合程度。准确度用误差表示。

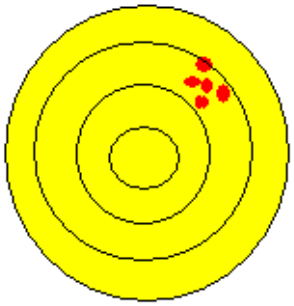
- **精密度 Precision**

  精密度表征平行测量值的相互符合程度。精密度用偏差表示。

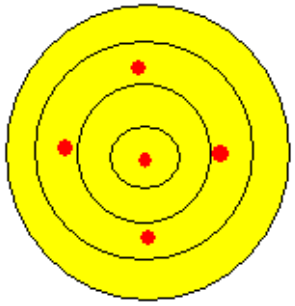
Illustration of the terms accuracy and precision



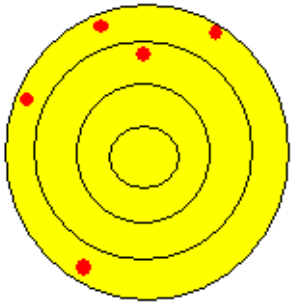
accurate and precise



inaccurate but precise



accurate but not precise



inaccurate and not precise

准确度与精密度的关系

- 结论：

1、精密度是保证准确度的前提。

2、精密度高，不一定准确度就高。

2.1.4 误差的分类及减免误差的方法

- 系统误差 （Systematic error)—某种固定的因素造成的误差
- 随机误差 （Random error)—不定的因素造成的误差
- 过失误差 （Gross error, mistake)

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	重现性、单向性（或周期性）、可测性	服从概率统计规律、不可测性
影响	准确度	精密度
消除或减小的方法	校正	增加测定的次数

检验是否存在系统误差：回收试验

$$回收率=\frac{x_3-x_1}{x_2}\times 100\%$$

x_1: 原组分含量

x_2: 已知量

x_3: 原组分含量＋已知量

校正系统误差的方法：

方法误差：可以选择标准方法与所采用的方法作对照试验，或选择与试样组成接近的标准试样作对照试验，找出校正值。

仪器误差：校正仪器。

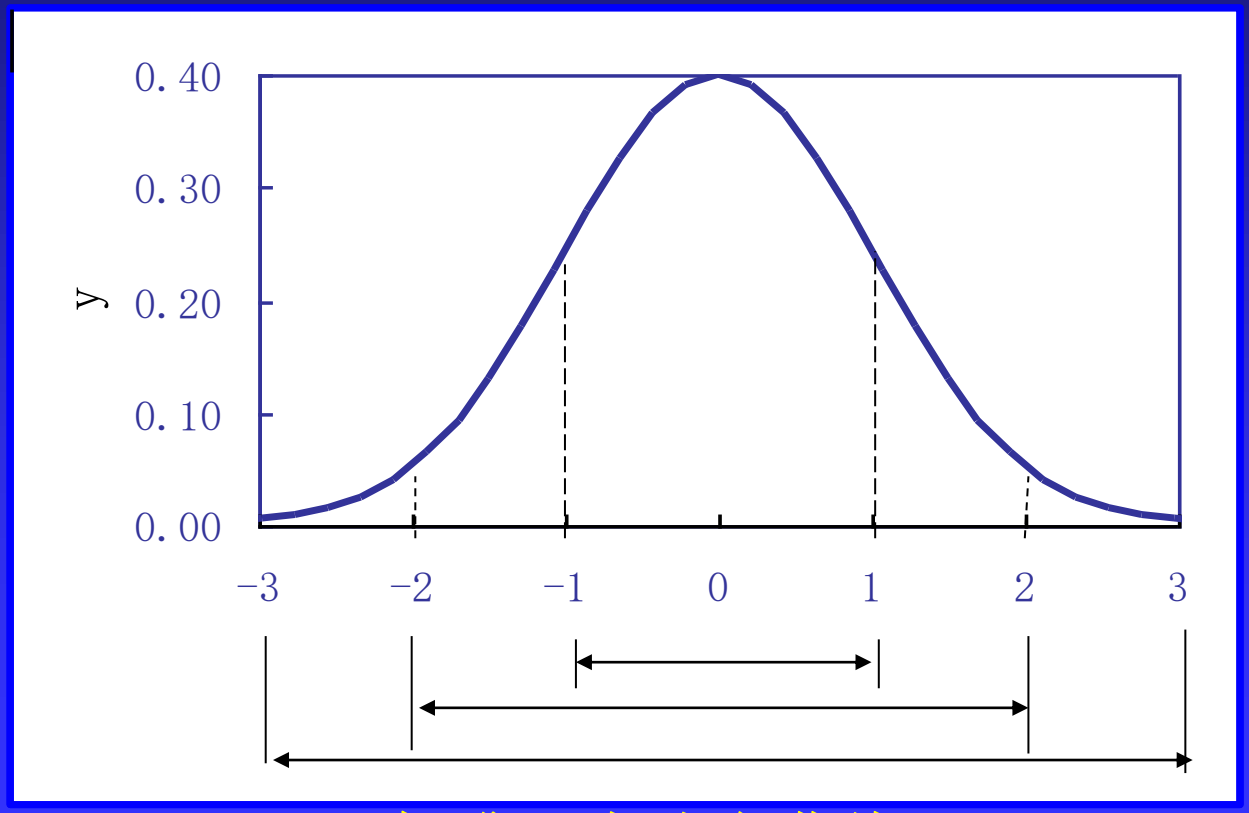
试剂误差：进行试剂的提纯，或通过空白试验扣除空白值加以校正。

空白试验：是指除了不加试样外，其他试验步骤与试样试验步骤完全一样的实验，所得结果称为空白值。

2.1.5 随机误差的分布服从正态分布

如果把曲线与横坐标从－∞至＋∞之间所包围的面积（代表所有随机误差出现的概率的总和）定为100%，通过计算发现误差范围与出现的概率有如下关系：

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



标准正态分布曲线

随机误差分布具有的性质：

1、对称性：正误差出现的概率与负误差出现的概率相等。误差分布曲线是对称的。

2、单峰性：小误差出现的概率大，大误差出现的概率小；特别大的误差出现的概率极小。误差分布曲线只有一个峰值，具有明显的集中趋势。

3、 有界性：大误差出现的概率很小，如果发现误差很大的测量值出现，往往是由于其他过失误差造成。

4、 抵偿性：误差的算术平均值的极限为零。

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{d_i}{n}=0$$

测定值或误差出现的概率称为置信度或置信水平，其意义可以理解为某一定范围的测定值（或误差值）出现的概率。

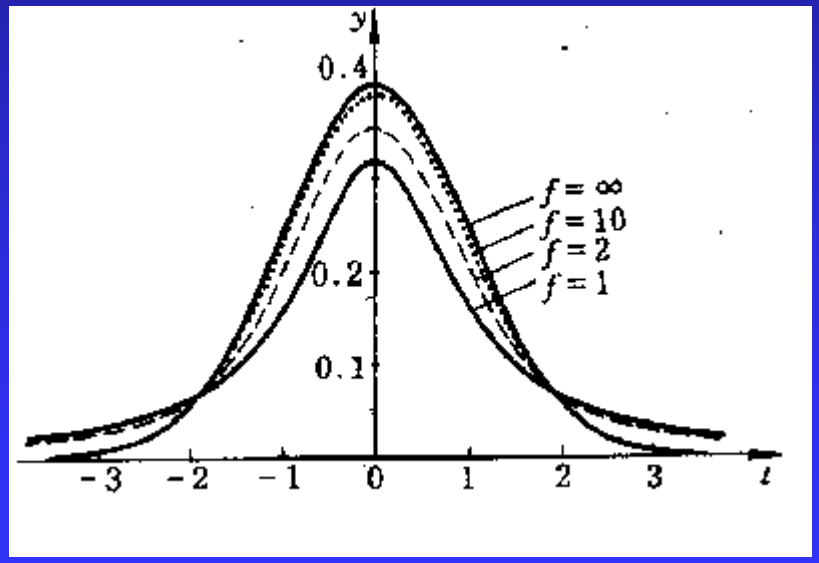
μ±σ， μ±2σ， μ±3σ 等称为置信区间，其意义为真实值在指定概率下，分布在某一个区间。置信度选得高，置信区间就宽。

随机误差出现的区间u（x-μ, 以σ为单位）	测量值出现的区间（置信区间）	概率%（置信度）（置信水平）
（-1, +1)	(μ-1 σ, μ+1 σ)	68.3
(-1.96, +1.96)	(μ-1.96 σ, μ+1.96 σ)	95.0
(-2, +2)	(μ-2 σ, μ+2 σ)	95.5
(-2.58, 2.58)	(μ-2.58 σ, μ+2.58 σ)	99.0
(-3, +3)	(μ-3 σ, μ+3 σ)	99.7

2.1.6 有限次测定中随机误差服从t分布

在分析测试中，测定次数是有限的，其随机误差服从类似于正态分布的t分布。t分布曲线随自由度而变化。

置信因子 $t = \dfrac{\bar{x} - \mu}{s}\sqrt{n}$



f=n-1

t 分布曲线

t 值表

n	P=0.90 α=0.10	P=0.95 α=0.05	P=0.99 α=0.01
5	2.132	2.776	4.604
6	2.015	2.571	4.032
7	1.943	2.447	3.707
8	1.895	2.365	3.500
9	1.833	2.306	3.355
21	1.725	2.086	2.846
∞	1.645	1.960	2.576

t值与置信度(P)和测定值的次数(n)有关。

显著水平 α=1-P

置信区间（对μ存在区间的估计）

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

置信区间的宽窄与置信度、测量值的精密度和测定次数有关。

测定的精密度越高，测定次数越多时，置信区间越窄，即平均值越接近真值，平均值越可靠。

意义：在一定的置信度下(如95%)，真值（μ）将在测定平均值（$\bar{x}$）附近的一个区间（$\bar{x} - \frac{ts}{\sqrt{n}}, \bar{x} + \frac{ts}{\sqrt{n}}$)存在，把握程度为95%。

$\pm \frac{ts}{\sqrt{n}}$表示不确定度

置信度选择越高，置信区间越宽，其区间包括真值的可能性越大。

分析化学中，一般将置信度定在95%或90%

对于一样本分析，报告给出$\bar{x}, s, n$

则可以根据不同的置信度的要求P

找出相应的$t_{\alpha,n}$值，再求得μ的置信区间。

例题：2.2 分析铁矿中的铁的质量分数，得到如下数据：37.45，37.20，37.50，37.30，37.25（%）。

（1）计算此结果的平均值、平均偏差、标准偏差、变异系数。

（2）求置信度分别为95%和99%的置信区间。

解（1）：

$$\bar{x} = \frac{37.45 + 37.20 + 37.50 + 37.30 + 37.25}{5}\% = 37.34\%$$

$$\bar{d} = \frac{1}{n}\sum|d_i| = \frac{1}{n}\sum|x_i - \bar{x}|$$
$$= \frac{1}{5}(0.11 + 0.14 + 0.04 + 0.16 + 0.09)\% = 0.11\%$$

$$s=\sqrt{\frac{\sum d_i^2}{n-1}}=\sqrt{\frac{\sum (x_i-\bar{x})^2}{n-1}}$$
$$=\sqrt{\frac{(0.11)^2+(0.14)^2+(0.04)^2+(0.16)^2+(0.09)^2}{5-1}}$$
$$=0.13\%$$

$$CV=\frac{s}{\bar{x}}\times 100\%=\frac{0.13}{37.34}\times 100\%=0.35\%$$

分析结果：

$$n=5,\bar{x}=37.34\%,s=0.13\%$$

求置信度为95%的置信区间。

$$n=5,\bar{x}=37.34\%,s=0.13\%$$

置信度为95%，查表：$t = 2.776$

μ 的95%置信区间：

$$(\bar{x}-t\frac{s}{\sqrt{n}},\bar{x}+t\ \frac{s}{\sqrt{n}})$$

$$=(37.34\%-2.776\times\frac{0.13\%}{\sqrt{5}},37.34\%+2.776\times\frac{0.13\%}{\sqrt{5}})$$

$$=(37.18\%,37.50\%)$$

n	P=0.90 α=0.10	P=0.95 α=0.05	P=0.99 α=0.01
5	2.132	2.776	4.604
6	2.015	2.571	4.032
7	1.943	2.447	3.707

2.1.7 公差

公差是生产部门对于分析结果允许误差的一种表示方法

待测组分含量与公差范围关系

待测组分的质量分数/%	公差（相对误差）/%
90	0.3
40	0.6
20	1.0
5	1.6
1.0	5.0
0.1	20
0.01	50
0.001	100

2.2 分析结果的数据处理

2.2.1 可疑数据的取舍

1、格鲁布斯(Grubbs)法

（1）将测量的数据按大小顺序排列

$$x_1 \prec x_2 \prec \cdots \prec x_n$$

（2）设第一个数据可疑，计算 $G_{计算} = \dfrac{\bar{x} - x_1}{s}$

或 设第n 个数据可疑，计算 $G_{计算} = \dfrac{x_n - \bar{x}}{s}$

（3）查表： $G_{计算} > G_{表}$， 舍弃。

例题2.3　　已知一组测量值分别为1.25,1.27,1.31,1.40

采用*Grubbs*法判断1.40是否需要舍弃。

将数据从小到大排序：

1.25,1.27,1.31,1.40

其中$x_n = 1.40$

求得$\bar{x} = 1.31, s = 0.066$

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

查$G_{p,n}$表，$G_{0.95,4} = 1.46$，　则$G_{计算} \prec G_{0.95,4}$

故1.40这一数据应保留。

$G_{p,n}$值表

n	P=95%	P=97.5%	P=99%
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
10	2.18	2.29	2.41
20	2.56	2.71	2.88

2、Q 检验法 Dixon's Q-test

（1）将测量的数据按大小顺序排列。$x_1, x_2, x_3 \cdots\cdots x_n$

（2）计算测定值的极差R。 $R = x_{\max} - x_{\min}$

（3）计算可疑值与相邻值之差（应取绝对值）d。

（4）计算Q值： $Q_{计算} = \frac{d}{R}$

（5）比较： $Q_{计算} \geq Q_{表}$ 舍弃。

例题2.4：测定碱灰总碱量（%Na_2O)得到6个数据，按其大小顺序排列为40.02，40.12，40.16，40.18，40.18，40.20。第一个数据可疑，判断是否应舍弃？（置信度为90%）。

解：

$$Q_{计算}=\frac{40.12-40.02}{40.20-40.02}=0.56$$

查表：　n = 6，$Q_{表}$ = 0.56　舍弃。

舍弃商Q值

测定次数n	3	4	5	6	7	8	9	10
$Q_{0.90}$	0.94	0.76	0.64	0.56	0.51	0.47	0.44	0.41
$Q_{0.95}$	0.97	0.84	0.73	0.64	0.59	0.54	0.51	0.49

1、 Q值法由于不必计算 $\bar{x}$ 和s,使用方便。

2、Q值法在统计上有可能保留离群较远的值。置信度常选90%。

3、判断可疑值常用Grubbs法。

2.2.2 平均值与标准值的比较

为了检验一个分析方法是否可靠，是否有足够的准确度，常用已知含量的标准试样进行试验，用t检验法将测定的平均值与已知值（标准值）比较，按

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

计算t值。

若$t_{计算} \succ t_{表}$，则$\bar{x}$与已知值有显著差别，表明被检验的方法存在系统误差；

若$t_{计算} \leq t_{表}$，则$\bar{x}$与已知值的差异可以认为是偶然误差引起的正常差异。

例2.5：一种新方法用来测定试样含铜量，
用含量为11.7mg/kg标准试样，进行五次测定，
所得数据为10.9,11.8,10.9,10.3,10.0,
判断该方法是否可行？（是否存在系统误差？）

解：计算$\bar{x}=10.8, s=0.7$

$$t=\frac{|\bar{x}-\mu|}{s}\sqrt{n}=\frac{|10.8-11.7|}{0.7}\times\sqrt{5}=2.87$$

查表得$t_{(0.95,\ n=5)}=2.776$，因此$t_{计算}\succ t_{表}$

说明该方法存在系统误差，结果偏低。

2.2.3 两个平均值的比较

当需要对两个分析人员测定相同试样所得结果进行评价，或需要对两种方法进行比较，检查两种方法是否存在显著性差异时，可选用t检验法进行判断。

判断两个平均值是否有显著性差异时，首先要求这两个平均值的精密度没有大的差异，为此可以采用F检验法进行判断。

1、*F* 检验法检验两组实验数据的精密度S_1和S_2之间有无显著差异：

$$F_{计算}=\frac{s_{大}^2}{s_{小}^2}$$

$$F_{计算}<F_{表}$$

精密度无显著差异。

2、*t* 检验确定两组平均值之间有无显著性差异

$$t_{计算}=\frac{|\bar{x}_1-\bar{x}_2|}{s_p}\sqrt{\frac{n_1\times n_2}{n_1+n_2}}$$

$$s_p=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

3、查t值表　　注意：$f = n_1 + n_2 - 2$

而t值表中$n = f + 1$

经典$t_{\alpha,f}$表更直观。

4、比较　　若$t_{计算} < t_{表}$

则表明两个平均值之间不存在显著差异，无系统误差

例2.6　　甲、乙二人对同一试样用不同方法进行测定，两组测定值如下：

甲：1.26　　1.25　　1.22

乙：1.35　　1.31　　1.33　　1.34

问两种方法间是否存在显著性差异？

解：$n_{甲}=3$　　$\bar{x}_{甲}=1.24$　　$s_{甲}=0.021$

$n_{乙}=4$　　$\bar{x}_{乙}=1.33$　　$s_{乙}=0.017$

$$F_{计算}=\frac{s_{大}^2}{s_{小}^2}=\frac{(0.021)^2}{(0.017)^2}=1.53$$

查表得$F_{表}=9.55$, 说明两组的方差无显著性差异。

置信度95%时的F值

$f_{s大}$ / $f_{s小}$	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

进一步用t公式进行计算。

$$t=\frac{\left|\bar{x}_1-\bar{x}_2\right|}{s_{合}}\sqrt{\frac{n_1n_2}{n_1+n_2}}$$

式中

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

$$=\sqrt{\frac{(3-1)\times0.021^2+(4-1)\times0.017^2}{3+4-2}}$$

$$=0.020$$

则

$$t=\frac{\left|\bar{x}_1-\bar{x}_2\right|}{s_{合}}\sqrt{\frac{n_1n_2}{n_1+n_2}}$$

$$=\frac{\left|1.24-1.33\right|}{0.020}\sqrt{\frac{3\times4}{3+4}}$$

$$=5.90$$

查表$f = n_1 + n_2 - 2 = 3 + 4 - 2 = 5$

则置信度95%时，$t_{表} = 2.571$

由于$t_{计算} \succ t_{表}$,表明甲乙两人采用的不同方法间存在着显著性差异。

本例中$\left|\bar{x}_1 - \bar{x}_2\right| = 0.09$，其中包含了系统误差和随机误差。根据$t$分布规律，随机误差允许最大值为

$$\left|\bar{x}_1 - \bar{x}_2\right| = ts_{合}\sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} = 2.57 \times 0.02 \times \sqrt{\frac{3+4}{3 \times 4}} \approx 0.04$$

说明可能有0.05的值由系统误差产生。

2.3 误差的传递

要点：误差传递的方式取决于误差的性质（系统误差或随机误差），取决于分析结果与测量值之间的化学计量关系（计算方式）。

2.3.1、系统误差的传递

设分析结果R 由测量值A、B、C计算获得，测量值的系统误差分别为ΔA、ΔB、ΔC，

(1) $R = A + B - C$,

$$(\Delta R)_{\max} = \Delta A + \Delta B + \Delta C$$

(2) $R = \dfrac{AB}{C}$,　　$\left(\dfrac{\Delta R}{R}\right)_{\max} = \dfrac{\Delta A}{A} + \dfrac{\Delta B}{B} + \dfrac{\Delta C}{C}$

2.3.2、随机误差的传递

设分析结果R由测量值A、B、C计算获得，标准偏差分别为s_A、s_B、s_C。

(1) $R = A + B - C$,

$$s_R^2 = s_A^2 + s_B^2 + s_C^2$$

(2) $R = \dfrac{AB}{C}$,

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

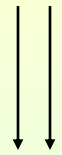
2.4 有效数字及其运算规则

有效数字及其运算规则

分析结果{表达了试样中待测组分的含量
反映了测量的准确程度

有效数字决定于{测量仪器
分析方法

有效数字



实际上能够测到的数字

示例2.7（0的作用）

数字	有效位数
0.00200	3
36000	不确定
3.6×10^4	2
3.60×10^4	3
3.600×10^4	4

示例2.8（倍数、分数、常数）

- 由于这些数据并不是测量所得到的，可视为无限多位有效数字。
- 常数的有效位数应取足够，在滴定分析中，常取4-5位有效数字。

元素	相对原子质量表
Ag	107.87
Ba	137.33
Ca	40.078
Fe	55.845
Pb	207.2

示例2.9（pH、pM、lgK）

方次

决定有效数字
2位

$$pH = 11.20$$

相当于[H^+]浓度 $6.3\times10^{-12}\,mol/L$

数字的修约规则

四舍六入五成双

一步修约到位

四舍六入

1、被修约的数字<=4，舍弃

3.148 ——→ 3.1

2、被修约的数字 >=6，进位

0.736 0.74

示例2.10（数字5的处理）

（1） 75.5 ⟶ 76

（2）2.451 ⟶ 2.5

例2.11 加减法的计算规则

	绝对误差
0.0121	0.0001
+25.64	0.01
+ 1.05782	0.00001
26.71	

以小数点后位数最少的数字为根据，

与绝对误差最大的数相对应。

例 2.12 乘除法的计算规则

	相对误差
0.0121	0.8%
× 25.64	0.4%
× 1.05782	0.009%
0.328	

根据有效数字最少的数来修约，

与相对误差最大的数相对应。

乘除法的计算规则

- 9.00，9.83等大数做4位有效数字处理
- 使用计算器进行连续运算时，过程中不必对每一步的计算结果进行修约，但应注意根据其准确度的要求，正确保留最后结果的有效数字位数。

示例2.13 （混合计算）

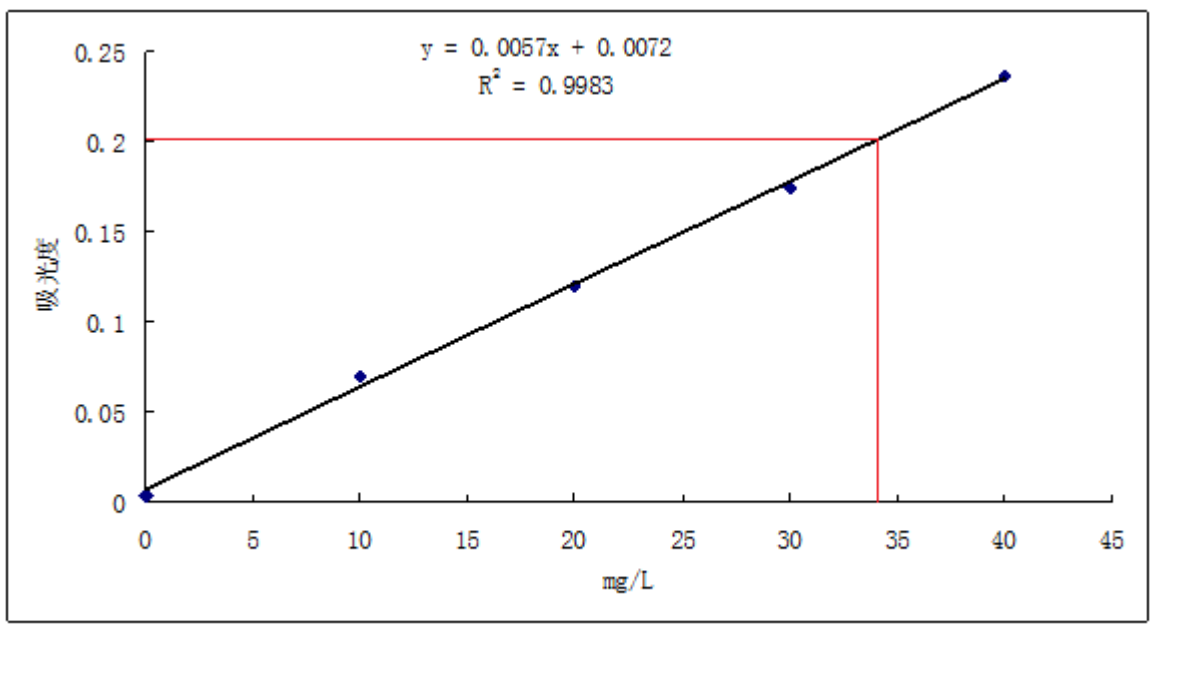
$$0.1239\times(40.00-39.10)$$

$$=0.1239\times0.90$$

$$=0.112$$

2.5 标准曲线的回归分析

2.5.1 标准曲线及线性回归



标样浓度 mg/L	0.0	10.0	20.0	30.0	40.0	试样
吸收值	0.004	0.070	0.120	0.175	0.236	0.200

线性回归 **Linear regression**

1、每个测量值都有误差，标准曲线应怎样作才合理？

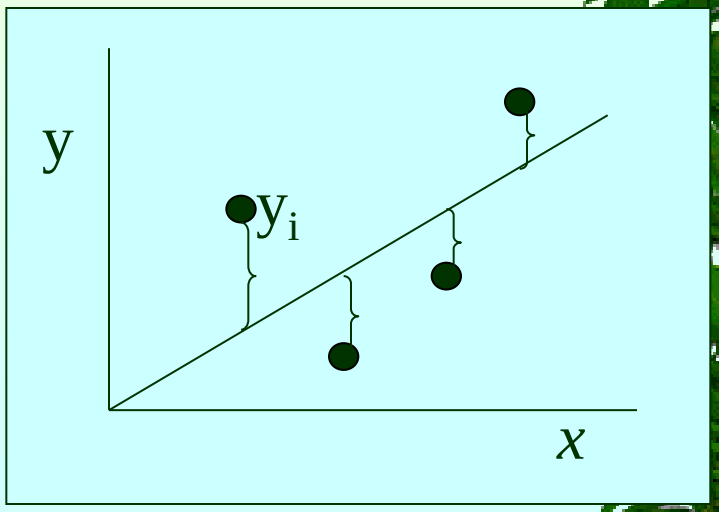
——线性回归问题

最小二乘法 method of least squares

设对y作n次独立的观测，得到一系列观测值。

一元线性回归方程表示为：

根据最小二乘法的原理，最佳的回归线应是各观测值y_i与相对应的落在回归线上的值之差的平方和（Q）为最小。



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

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

令：

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

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

解得：

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

其中：

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

相关系数 Correlation coefficient

2、应怎样估计线性的好坏？——相关系数的问题

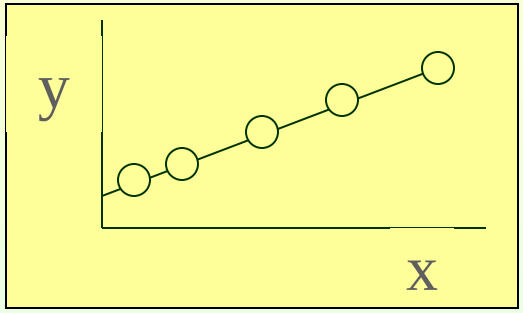
判断一元回归线是否有意义，可用相关系数来检验。

相关系数的定义为：

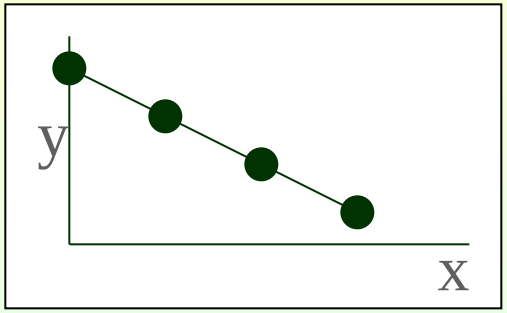
$$R = b\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

相关系数的物理意义

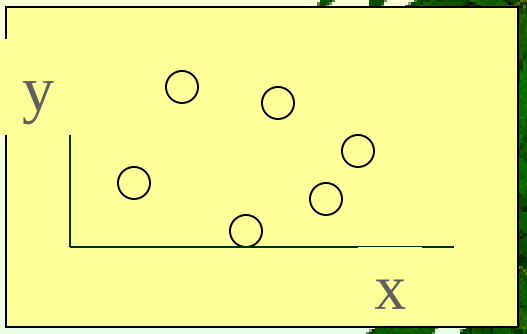
1. 当所有的y_i值都在回归线上时，$R=\pm1$。



$R = 1$



$R = -1$



$R = 0$

2. 当y与x之间不存在直线关系时，$R=0$。

3. 当R的值在0与1之间时，可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的。

相关系数的临界值表（部分）

f=n-2 / P / r	1	2	3	4	5
90%	0.988	0.900	0.805	0.729	0.669
95%	0.997	0.950	0.878	0.811	0.755
99%	0.999	0.990	0.959	0.917	0.875