

-自动化学院学科核心课-

检测技术与自动化

测量误差与数据处理 (2)





本节内容:

1、随机误差的处理

表示方法、概率分布、真值、方差、置信度

2、系统误差的处理

分类、判断方法、消除方法

3、粗大误差的处理

判别方法、3 σ 准则、拉依达准则、格鲁布斯准则



直接测量结果的计算

重点

- 1) 等精度测量
- 2) 不等精度测量

理解

系统误差、粗大误差、随机误差的分类依据

1. 测量误差综述

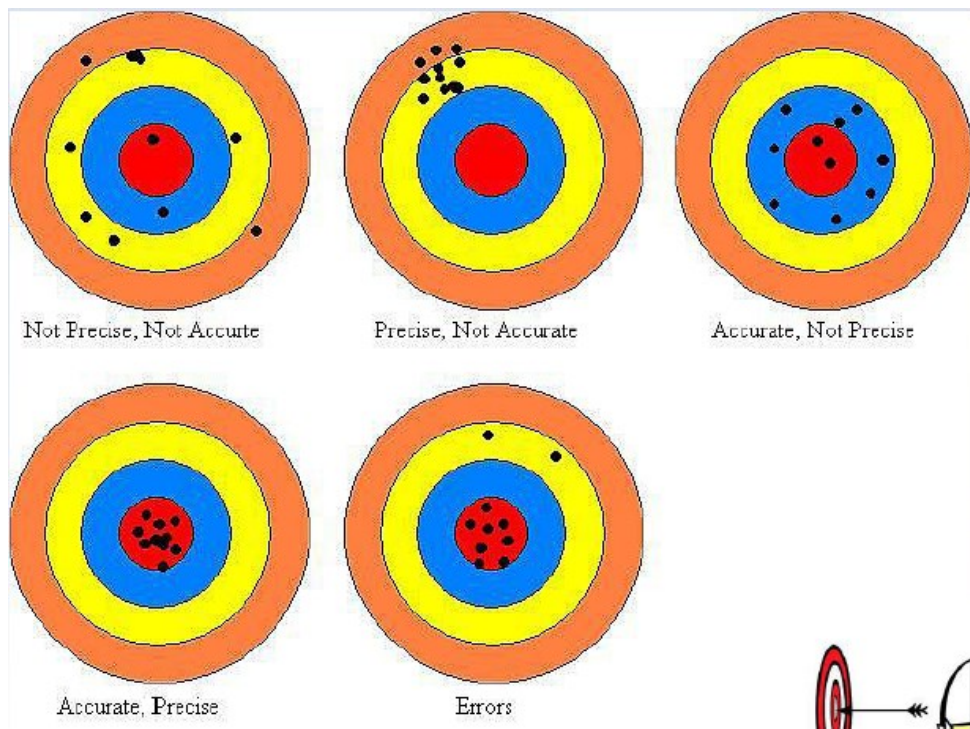
$$\begin{array}{ccccc} \text{测量} & & \text{测量} & & \text{系统} & & \text{随机} & & \text{粗大} \\ \text{结果} & & \text{真值} & & \text{误差} & & \text{误差} & & \text{误差} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ X_{meas} & = & X_{true} & + & e_s & + & e_r & + & e_A \end{array}$$

测量精度：反映测量结果与真值接近程度的量

(1) **精密度(precision)**：反映测量结果中**随机误差**的影响程度

(2) **准确度(accuracy)**：反映测量结果中**系统误差**的影响程度

(3) **精确度**：反映测量结果中**系统误差和随机误差综合**的影响程度，其定量特征可用测量的**不确定度**来表示





2. 随机误差的处理

名词区分

1	6.61	6	7.70	11	8.35	16	8.67	21	9.17	26	9.75
2	7.19	7	7.78	12	8.49	17	9.00	22	9.38	27	10.06
3	7.22	8	7.79	13	8.61	18	9.08	23	9.64	28	10.09
4	7.29	9	8.10	14	8.62	19	9.15	24	9.70	29	11.28
5	7.55	10	8.19	15	8.65	20	9.16	25	9.72	30	11.39

- 平均值 (average) 算术平均、几何平均…
- 平均均方差 (mean squared deviation from the average)
- 平均均方根 (root mean squared deviation from the average)
- 均值 (mean)
- 方差 (variance)
- 标准差 (standard deviation)

数据集

总体



2. 随机误差的处理

一些概念

被测对象的测量值 x_i 来自总体 X ，该总体 X 具有finite期望和方差

$$\varepsilon_i = x_i - \mu = [x_i - E(X)] + [E(X) - \mu]$$



(1) 真值无偏估计的概念

构造统计量 $\hat{\mu} = g(x_1, x_2, \dots, x_n)$ s.t. $E(\hat{\mu}) = E(X)$

真值的无偏估计、最佳估计 (best estimate) : $\hat{\mu} \rightarrow E(\hat{\mu})$

(2) 误差与残差的定义

测量误差(error)为: $\varepsilon_i = x_i - E(X) = x_i - E(\hat{\mu})$

残差(residuals): $v_i = x_i - \hat{\mu}$

2. 随机误差的处理

(3) 无偏统计量 $\hat{\mu}$ 的计算

- 等精度测量情况下:

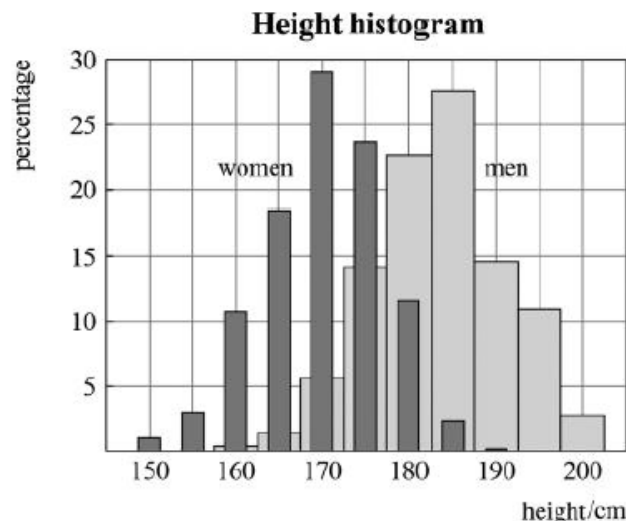
$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- 不等精度测量情况下:

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

$$\bar{x} = \frac{1}{p} \sum_{i=1}^n p_i x_i, \quad p = \sum_{i=1}^n p_i$$

真值估计



Histograms of the height distribution of men (light gray) and women (dark gray) in the age group 20-29 in the Netherlands, averaged over the years 1998, 1999 and 2000. The data have been gathered in bins of 5cm width.

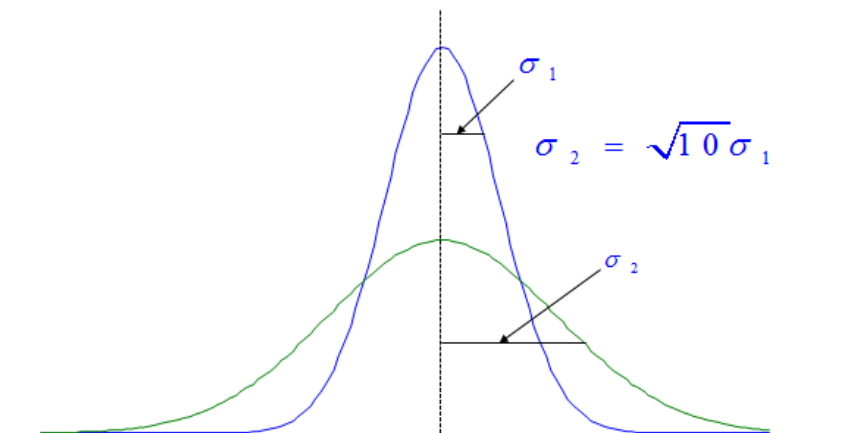
工作基准米尺在连续三天内与国家基准器比较, 得到工作基准米尺的平均长度的平均长度为999.9425mm (三次测量的)、999.9416mm (两次测量的)、999.9419mm (五次测量的), 每单次测量均为等精度测量, 求最后测量结果。

2. 随机误差的处理

随机误差的标准差估计

(4) 单次测量随机误差的无偏估计

$$\sigma = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{n}} \quad \xrightarrow{v_i = x_i - \hat{\mu}} \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n v_i^2}{n-1}}$$



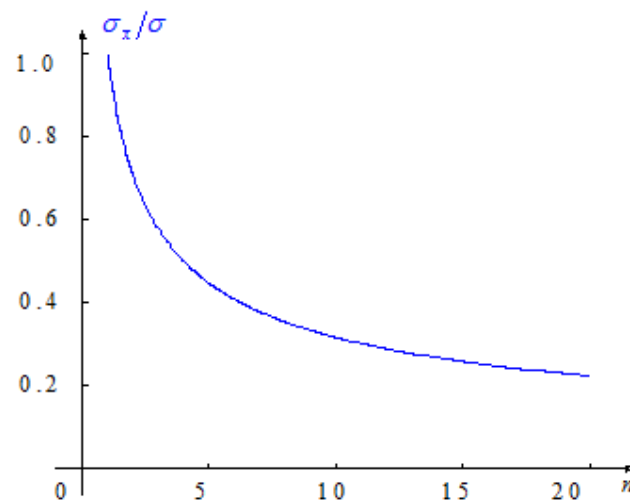
(5) 算术平均值的标准差

- 等精度测量情况下:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}$$

说明了什么?

测量次数的最佳选择
 $n \leq 10$





2. 随机误差的处理 计算实例1

802.40, 802.50, 802.38, 802.48, 802.42,
802.46, 802.39, 802.47, 802.43, 802.44

求算术平均值及其测量标准差？

算术平均值为： $\bar{X} = \frac{\sum_{i=1}^{10} x_i}{10} = 802.437 \approx 802.44$

测量结果的标准差： $\sigma = \sqrt{\sum_{i=1}^{10} (x_i - \bar{x})^2 / (10-1)} = 0.04$

算术平均值的标准差： $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 0.013$

测量结果： 802.44 ± 0.01

若已知仪器标准差为0.05？

802.40, 802.50, 802.38, 802.48, 802.42,
802.46, 802.39, 802.47, 802.43, 802.44

求算术平均值及其测量标准差？

算术平均值为： $\bar{X} = \frac{\sum_{i=1}^{10} x_i}{10} = 802.437 \approx 802.44$

测量结果的标准差： $\sigma = \sqrt{\sum_{i=1}^{10} (x_i - \bar{x})^2 / (10-1)} = 0.04$

算术平均值的标准差： $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = 0.013$

测量结果： 802.44 ± 0.01

若已知仪器标准差为0.05，则测量结果标准差为
[填空1]

作答



2. 随机误差的处理

等精度测量与不等精度测量

(1) 等精度的含义

(2) 实现等精度的条件?

测量环境、仪器、方法、人员水平及测量次数都相同

(3) 不等精度的含义

(4) 常见的不等精度测量

- 1) 相同条件下的不同次数测量;
- 2) 使用不同的仪器或不同的方法, 由不同的人员进行对比测量;
- 3) 对不同时期测得的数据加以综合, 给出最可信赖的测量结果。



2. 随机误差的处理

随机误差的标准差估计

- 不等精度测量情况下

例：用A、B两种仪器对5V稳压芯片的输出电压进行两次测量，测量结果分别是5.005V（标准差为0.006V）、5.002V（标准差为0.008V），求该输出电压的最佳估计值及其标准差。

例：工作基准米尺在连续三天内与国家基准器比较，得到工作基准米尺的平均长度的平均长度为999.9425mm（三次测量的）、999.9416mm（两次测量的）、999.9419mm（五次测量的），每单次测量均为等精度测量，求最后测量结果。



2. 随机误差的处理

随机误差的标准差估计

- 不等精度测量情况下

各组测量值的算数平均值标准差已知

$$n_1 \sigma_{\bar{x}_1}^2 = n_2 \sigma_{\bar{x}_2}^2 = \cdots n_m \sigma_{\bar{x}_m}^2 = \sigma^2$$

$$\sigma_{\bar{x}_i} = \frac{\sigma}{\sqrt{n_i}}$$

fai 为全部值
减掉均值除以
均值对应的数
量。

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{\sum_{i=1}^m n_i}}$$

$$\therefore \sigma_{\bar{x}} = \sigma_{\bar{x}_i} \sqrt{\frac{n_i}{\sum_{i=1}^m n_i}}$$

对于已知 σ_i 的情况

$$p_1 \sigma_{\bar{x}_1}^2 = p_2 \sigma_{\bar{x}_2}^2 = \cdots p_m \sigma_{\bar{x}_m}^2 = \sigma^2$$

$$p_1 : p_2 : \cdots : p_m = \frac{1}{\sigma_{\bar{x}_1}^2} : \frac{1}{\sigma_{\bar{x}_2}^2} : \cdots : \frac{1}{\sigma_{\bar{x}_m}^2}$$

$$\sigma_{\bar{x}} = \sigma_{\bar{x}_i} \sqrt{\frac{p_i}{\sum_{i=1}^m p_i}} = \sigma_{\bar{x}_i} \sqrt{\frac{\frac{1}{\sigma_{\bar{x}_i}^2}}{\sum_{i=1}^m \frac{1}{\sigma_{\bar{x}_i}^2}}}$$



2. 随机误差的处理

随机误差的标准差估计

- 不等精度测量情况下（续） 各组测量值的算术平均值标准差未知

对同一被测量进行m组不等精度测量，得到m个测量结果 $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ ，若已知单次测得值的标准差为 σ ，则各组测量结果的残差为

$$v_{\bar{x}_i} = \bar{x}_i - \bar{x}$$

乘以权重，变换为等精度测量列 $\sqrt{p_i}\bar{x}_i$ ，则有

$$\sqrt{p_i}v_{\bar{x}_i} = \sqrt{p_i}\bar{x}_i - \sqrt{p_i}\bar{x}$$

因此有

$$\sigma = \sqrt{\frac{\sum_{i=1}^m p_i v_{\bar{x}_i}^2}{(m-1)}} \xrightarrow{\text{代入}} \sigma_{\bar{x}} = \sigma_{\bar{x}_i} \sqrt{\frac{p_i}{\sum_{i=1}^m p_i}} = \frac{\sigma}{\sqrt{\sum_{i=1}^m p_i}}$$

注意这里为什么用m-1，因为求v的时候均值用的是单个小组的均值，而不是所有小组的均值，因此不是n-1；

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^m p_i v_{\bar{x}_i}^2}{(m-1) \sum_{i=1}^m p_i}}$$

用A、B两种仪器对5V稳压芯片的输出电压进行两次测量，测量结果分别是5.005V（标准差为0.006V）、5.002V（标准差为0.008V），求该输出电压的最佳估计值及其标准差。

用两种仪器进行的两次测量构成了不等精度测量列，两次测量分别测得稳压芯片的输出电压为：

$$U_A = 5.005V, \sigma_A = 0.006V; U_B = 5.002V, \sigma_B = 0.008V$$

按测量结果的标准差来确定两个测量值的权，得： $p_A : p_B = [\text{填空1}] : [\text{填空2}]$

作答



2. 随机误差的处理 计算实例2

用A、B两种仪器对5V稳压芯片的输出电压进行两次测量，测量结果分别是5.005V（标准差为0.006V）、5.002V（标准差为0.008V），求该输出电压的最佳估计值及其标准差。

用两种仪器进行的两次测量构成了不等精度测量列，两次测量分别测得稳压芯片的输出电压为：

$$U_A = 5.005V, \sigma_A = 0.006V; U_B = 5.002V, \sigma_B = 0.008V$$

按测量结果的标准差来确定两个测量值的权，得

$$p_A : p_B = \frac{1}{\sigma_A^2} : \frac{1}{\sigma_B^2} = 16 : 9$$

因此，

$$\begin{aligned}\sigma_{\bar{V}} &= \sigma_A \sqrt{\frac{p_A}{\sum_{i=1}^2 p_i}} = 0.006 \times \sqrt{\frac{16}{16+9}} V \\ &= 0.0048V \approx 0.005V\end{aligned}$$

$$\bar{U} = \frac{U_A p_A + U_B p_B}{p_A + p_B} = \frac{5.005 \times 16 + 5.002 \times 9}{16 + 9} V = 5.004V$$



2. 随机误差的处理 计算实例3

工作基准米尺在连续三天内与国家基准器比较，得到工作基准米尺的平均长度的平均长度为999.9425mm（三次测量的）、999.9416mm（两次测量的）、999.9419mm（五次测量的），每单次测量均为等精度测量，求最后测量结果。

按测量次数来确定权： $p_1 = 3, p_2 = 2, p_3 = 5$

$$\bar{x} = \frac{3 \times 999.9425 + 2 \times 999.9416 + 5 \times 999.9419}{3 + 2 + 5} = 999.9420mm$$

可得各组测量结果的残差为

$$v_1 = 0.5\mu m, v_2 = -0.4\mu m, v_3 = -0.1\mu m$$

由 $m = 3, p_1 = 3, p_2 = 2, p_3 = 5$

$$\begin{aligned}\sigma_{\bar{x}} &= \sqrt{\frac{3 \times 0.5^2 + 2 \times (-0.4)^2 + 5 \times (-0.1)^2}{(3 - 1) \times (3 + 2 + 5)}} = \frac{1.12}{20} \mu m \\ &= 0.24\mu m \approx 0.0002mm\end{aligned}$$



2. 随机误差的处理

误差范围——置信区间

目标：真值 μ 的置信区间

已知：一系列测量值 x_i

➤ 若考虑随机误差的分布为正态分布

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

σ 未知，用 $\hat{\sigma}$ 代替
构成t分布
为什么是t分布

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}}$$

➤ 真值的置信区间

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t(n-1)$$

$$P \left\{ \left| \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} \right| \leq t_{\alpha/2}(n-1) \right\} = 1 - \alpha$$

$$P \left\{ \bar{x} - \frac{\hat{\sigma}}{\sqrt{n}} t_{\alpha/2}(n-1) \leq \mu \leq \bar{x} + \frac{\hat{\sigma}}{\sqrt{n}} t_{\alpha/2}(n-1) \right\} = 1 - \alpha$$

则 μ 的置信度为 $1-\alpha$ 的置信区间为：

$$\left[\bar{x} - \frac{\hat{\sigma}}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1), \bar{x} + \frac{\hat{\sigma}}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \right]$$

■ 卡方分布定义

假设随机变量 X_1, X_2, \dots, X_n 相互独立，且都服从 $N(0,1)$ ，则随机变量

$$Z = \sum_{i=1}^n X_i^2 \sim \chi^2(n)$$

■ t分布定义

假设 X 服从标准正态分布，即 $X \sim N(0,1)$ ， Y 服从自由度 n 的卡方分布，即 $Y \sim \chi^2(n)$ ，且 X 与 Y 相互独立，则

$$Z = \frac{X}{\sqrt{Y/n}} \sim t(n)$$

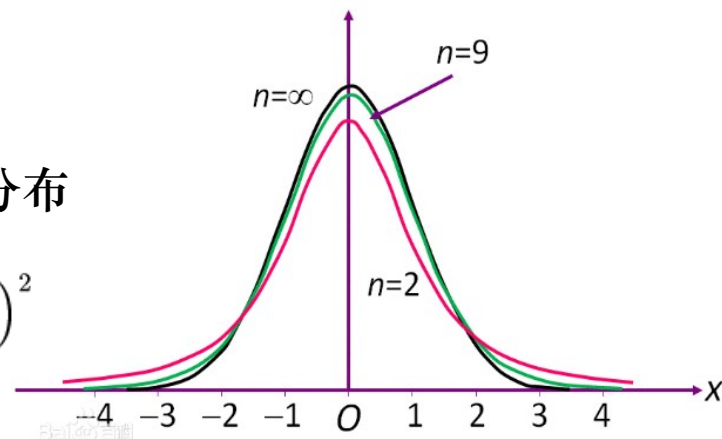
随着自由度逐渐增大，t分布逐渐接近标准正态分布

$$\bar{X}_n = (X_1 + \dots + X_n)/n$$

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

$$T = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}}$$

自由度为n-1的t分布





2. 随机误差的处理 计算实例1(续)

802.40, 802.50, 802.38, 802.48, 802.42,
802.46, 802.39, 802.47, 802.43, 802.44

求算术平均值及其测量标准差？若已知仪器误差为0.05？

算术平均值为： $\bar{X} = \frac{\sum_{i=1}^{10} x_i}{10} = 802.437 \approx 802.44$

测量结果的标准差： $\delta = \sqrt{\sum_{i=1}^{10} (x_i - \bar{x})^2 / (10-1)} = 0.04$

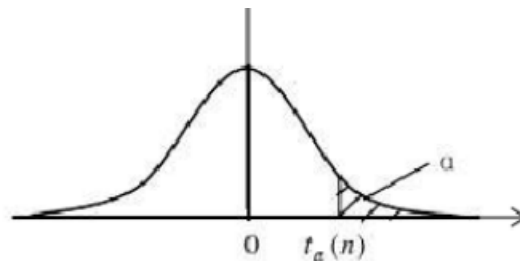
算术平均值的标准差： $\delta_x = \frac{\delta}{\sqrt{n}} = 0.013$

测量结果： 802.44 ± 0.01

真值的置信区间（置信水平为90%）： $[802.42, 802.46]$;
 $n=10$ （正态分布误差） $(0.013 * 1.833 = 0.0238)$

t分布表

$$P\{t(n) > t_{\alpha}(n)\} = \alpha$$

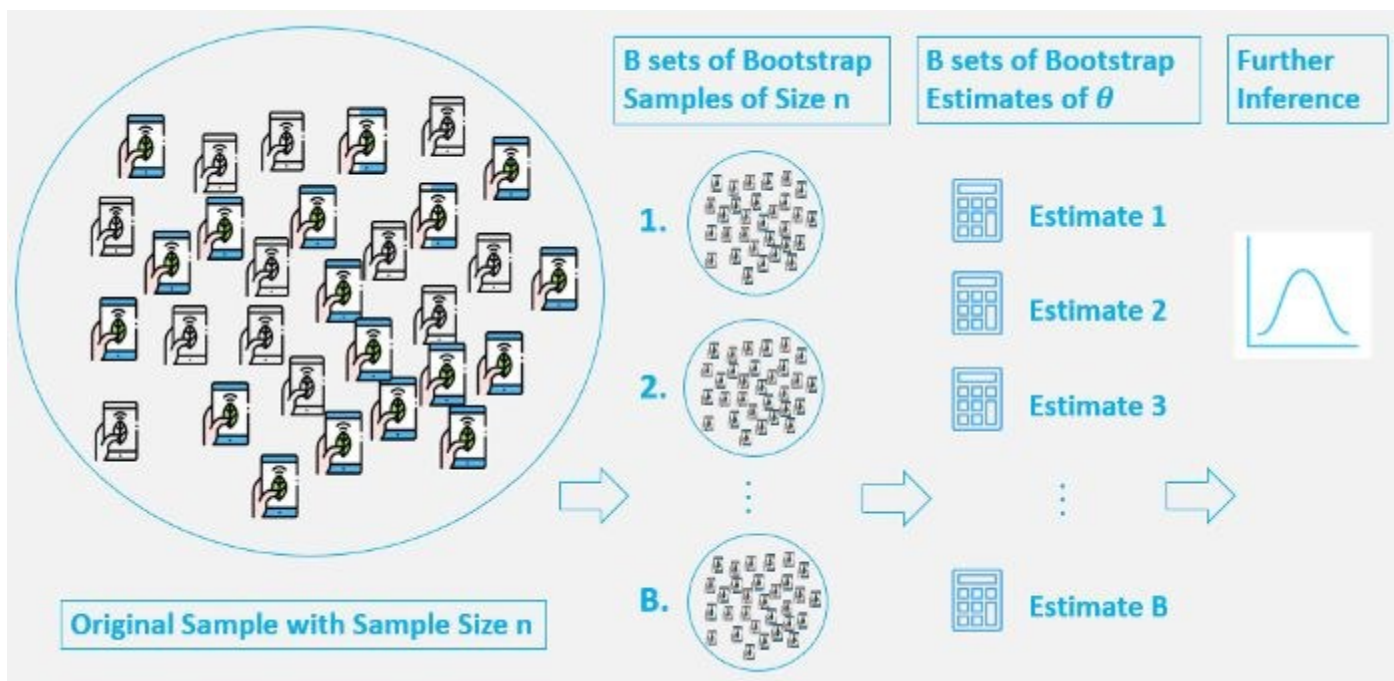


n	$\alpha=0.25$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.025$	$\alpha=0.01$	$\alpha=0.005$
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467

2. 随机误差的处理 区间估计——Bootstrap方法

(1) 使用原因：不确定数据分布类型

(2) 基本思想





2. 随机误差的处理 区间估计——Bootstrap方法

(3) 步骤

① 采样 (有放回)

如：在22个个体中作有放回的抽样，每个个体被抽中的概率为1/22，可以用计算机产生1~22的均匀分布的随机数，相应的编号即为抽中的个体，由这些个体组成的样本就是一个bootstrap样本。如，随机生成一组随机数，10, 20, 3, 19, 12, 2, 3, 2, 5, 20, 8, 13, 12, 18, 17, 22, 2, 1, 19, 7, 8, 13

② 对抽取样本计算均值 $\widehat{x}_{(i)}$

③ 前两步重复B次 (如：B=1000)

$$\hat{x} = \frac{1}{B} \sum_{i=1}^B \widehat{x}_{(i)} \quad \text{Var}(\hat{x}) = \frac{1}{B-1} \sum_{i=1}^B (\widehat{x}_{(i)} - \hat{x})^2$$

④ 计算总体均值的置信区间

(a) 标准Bootstrap方法

由此得到总体均值的 $(1 - \alpha)\%$ 的置信区间为：

$$\left[\hat{x} - u_{1-\frac{\alpha}{2}} \text{Var}(\hat{x}), \hat{x} + u_{1-\frac{\alpha}{2}} \text{Var}(\hat{x}) \right]$$

$u_{1-\frac{\alpha}{2}}$ 是标准正态分布的 $1 - \frac{\alpha}{2}$ 百分位数



2. 随机误差的处理 区间估计——Bootstrap方法

(b) 百分位数Bootstrap方法

将bootstrap得到的样本均值从小到大排列，得到 \hat{x}_w^*

第 $\frac{\alpha}{2}$ 和第 $1 - \frac{\alpha}{2}$ 分位点分别是 $1 - \alpha$ 置信水平下统计量 \hat{x} 的上下限

$$[\hat{x}_{\frac{\alpha}{2}B}^*, \hat{x}_{(1-\frac{\alpha}{2})B}^*]$$

(c) t百分位数Bootstrap方法

将bootstrap得到的样本计算t统计量，并从小到大排列

$$t_{(i)} = \sqrt{\frac{\hat{x}_{(i)} - \hat{x}}{Var(\hat{x})}}$$

将结果从小到大排列，得到顺序统计量 $t_{(i)}^*$ ，当显著性水平为 α 时，第 $\frac{\alpha}{2}$ 和第 $1 - \frac{\alpha}{2}$ 分位点分别是 $\hat{t}_{\frac{\alpha}{2}B}^*, \hat{t}_{(1-\frac{\alpha}{2})B}^*$

$$[\hat{x} - \hat{t}_{\frac{\alpha}{2}B}^* \times \sqrt{Var(\hat{x})}, \hat{x} + \hat{t}_{\frac{\alpha}{2}B}^* \times \sqrt{Var(\hat{x})}]$$

3. 系统误差的处理 系统误差的分类

➤ 恒定系统误差

➤ 可变系统误差

线性系统误差

周期变化系统误差

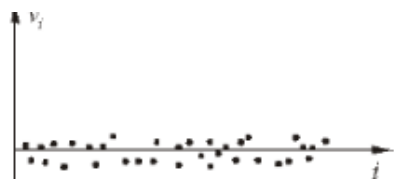
复杂规律变化系统误差

刻度值为1mm的标准刻度尺存在刻画误差 Δ mm，若用它测量某一物体，得到的值为 k ，则被测

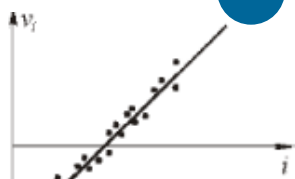
$$L = k(1 + \Delta)mm$$

仪表指针回转中心与刻度盘中心有偏离值，则指针在任一转角处的读数误差为周期变化误差

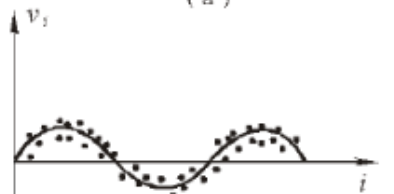
微安表指针偏转角与偏转力矩不严格保持线性关系，而表盘仍采用均匀刻度所产生的误差



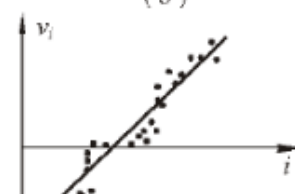
(a)



(b)



(c)

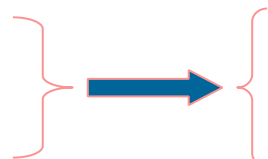


(d)



3. 系统误差的处理 系统误差的发现

实验对比法
理论分析法
数据分析法



发现恒定还是可变系统误差？
定性分析还是定量发现？

结论：

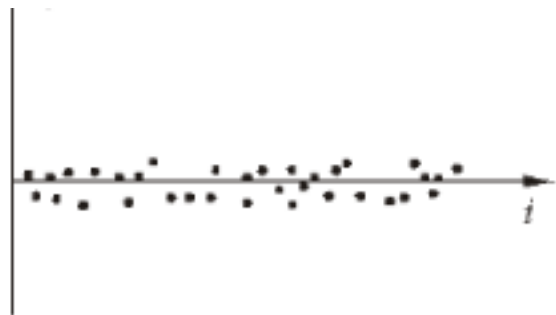
- 1) 实验对比法主要用于发现恒定系统误差；
- 2) 理论分析法主要用于定性分析；
- 3) 数据分析法主要用于定量发现。

数据分析方法有：

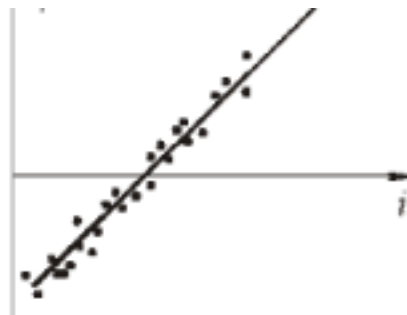
残余误差观察法、残余误差校验法、不同公式计算标准差比较法、计算数据比较法、t检验法、秩和检验法

3. 系统误差的处理 系统误差的发现

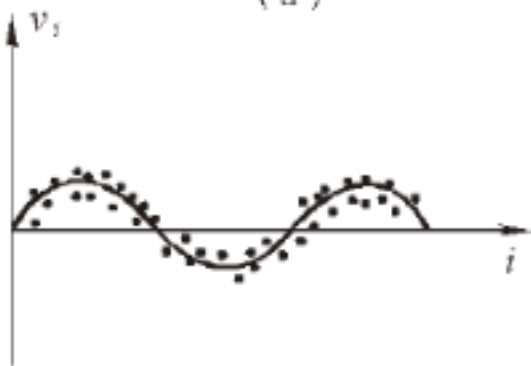
(1) 画图观察法



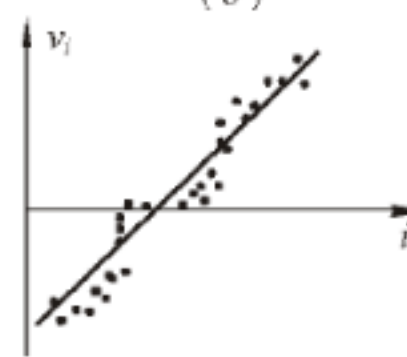
(a)



(b)



(c)



(d)



3. 系统误差的处理 系统误差的发现

(2) 不同公式计算标准差比较法 (同一组测量值)

贝塞尔公式: $\sigma_1 = \sqrt{\frac{\sum_{i=1}^n v_i^2}{n-1}}$ 别捷尔斯公式: $\sigma_2 = 1.253 \frac{\sum_{i=1}^n |v_i|}{\sqrt{n(n-1)}}$

令 $\frac{\sigma_2}{\sigma_1} = 1 + u$

若 $|u| \geq \frac{2}{\sqrt{n-1}}$

则怀疑测量列中存在系统误差



3. 系统误差的处理 系统误差的发现

(3) 计算数据比较法 (不同组测量值)

若对同一量独立测得m组结果，并知它们的算术平均值和标准差为：

$$\bar{x}_1, \sigma_1; \quad \bar{x}_2, \sigma_2; \quad \dots, \quad \bar{x}_m, \sigma_m$$

而任意两组结果之差为：

$$\Delta = \bar{x}_i - \bar{x}_j$$

其标准差为：

$$\sigma = \sqrt{\sigma_i^2 + \sigma_j^2}$$

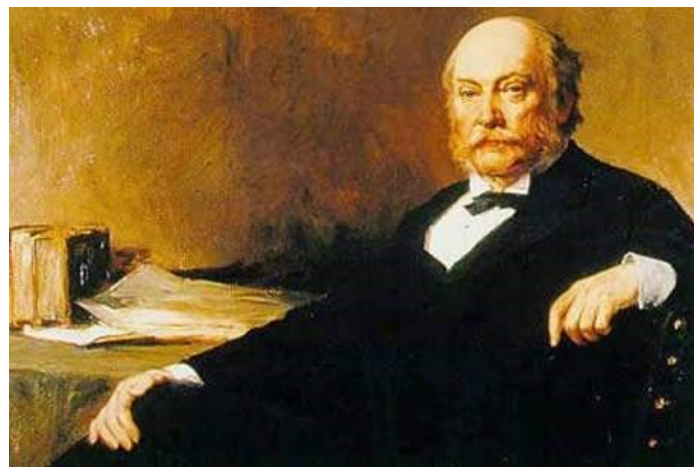
任意两组结果间不存在系统误差的标志是 $|\bar{x}_i - \bar{x}_j| < 2\sqrt{\sigma_i^2 + \sigma_j^2}$



3. 系统误差的处理 计算实例5

物理学家雷莱(Rayleigh)发现惰性气体

1893年, 雷莱用精密天平测量常见的各种气体的密度时发现, 凡从空气中得到的氮气每升重1.2575克, 而从氮的化合物中提取的氮气在相同条件下每升重为1.2505克。



例: 由化学法制取氮: $\bar{x}_1 = 1.2505$, $\sigma_1 = 0.00041$
由大气中提取氮: $\bar{x}_2 = 1.2572$, $\sigma_2 = 0.00019$
两者差值: $\Delta = \bar{x}_2 - \bar{x}_1 = 0.0067$

而标准差: $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{0.00041^2 + 0.00019^2} = 0.00045$

$$\Delta \gg 2\sqrt{\sigma_1^2 + \sigma_2^2} = 2 \times 0.00045 \\ = 0.0009$$

并未企图设法改进实验方法缩小误差, 而是强调两种方法的实质差别
最终发现惰性气体的存在



4. 粗大误差的处理 粗大误差的判别

➤ 基本思想

给定一置信概率，确定相应的置信区间，凡超过置信区间的误差就认为是粗大误差。

➤ 问题

是否已知误差分布规律？
若不知，测量点数是否足够多？
测量点数较少时如何处理？

➤ 两个判断准则

- 1) 3σ 准则；
- 2) 格鲁布斯准则



4. 粗大误差的处理 3 σ 准则

➤ 计算过程

- 1) 假设数据只含随机误差，计算标准偏差；
- 2) 确定置信区间；
- 3) 进行粗大误差剔除； $|v_i| > 3\sigma$
- 4) 重复上述步骤，直至粗大误差全部剔除。

关键问题

- 1) 应用前提条件！
- 2) 真值的计算；
- 3) 标准偏差的计算；
- 4) 逐步剔除。

几点说明：

小于10个点： $|v_i| > 2.5\sigma$

10-50个点： $|v_i| > 3\sigma$

50个点以上： $|v_i| > 3.5\sigma$



4. 粗大误差的处理 计算实例5

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
结果	20.42	20.43	20.40	20.43	20.42	20.43	20.39	20.30	20.40	20.43	20.42	20.41	20.39	20.39	20.40

第1种方法

算术平均值是: $\bar{X} = 20.404$

标准差是: $\delta = 0.033$

第8个测量误差的绝对值是: $0.104 > 2.5\delta = 0.0825$

结论：第8个点为粗大误差，剔除后再进行计算均在误差要求范围之内。



4. 粗大误差的处理 格鲁布斯准则

原则：**逐个剔除**，按误差分布符合t分布进行计算

计算过程：

1) 将数据从小到大排列 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

2) 计算统计量

$$g_{(1)} = \frac{\bar{x} - x_{(1)}}{\hat{\sigma}}$$

$$g_{(n)} = \frac{x_{(n)} - \bar{x}}{\hat{\sigma}}$$

3) 获取临界值 $g_0(n, \alpha)$

$$Pr\left(\frac{\bar{x} - x_{(1)}}{\hat{\sigma}} \geq g_0(n, \alpha)\right) = \alpha$$

$$Pr\left(\frac{x_{(n)} - \bar{x}}{\hat{\sigma}} \geq g_0(n, \alpha)\right) = \alpha$$

4) 将统计量与临界值比较

$$\text{if } g_{(1)} \geq g_0(n, \alpha)$$

$$\text{if } g_{(n)} \geq g_0(n, \alpha)$$

注意，1次剔除1个



4. 粗大误差的处理 格鲁布斯准则

原则：**逐个剔除**，按误差分布符合t分布进行计算
计算过程：

1) 测量数据，如测量10次得到：

8.2, 5.4, 14.0, 7.3, 4.7, 9.0, 6.5, 10.1, 7.7, 6.0

2) 计算平均值 \bar{x} 和标准差 σ

$\bar{x} = 7.89$; $\sigma = 2.704$

3) 计算偏离值：7.89-4.7 = 3.19; 14.0-7.89 = 6.11

4) 确定一个可疑值：14.0

5) 计算 $G_i = \frac{X_i - \bar{x}}{\sigma}$ ： $\frac{14.0 - 7.89}{2.704} = 2.260$ 。

6) 查表得相应临界值 $GP(n, \alpha)$

$n = 10, \alpha = 0.05$, 临界值 $GP(10, 0.05) = 2.176$

7) 剔除粗大误差：若 $G_i > GP(n, \alpha)$

8) 余下数据继续按照上述步骤考虑，直至没有异常值

和t分布的关系？



格鲁布斯准则 (Grubbs) 临界值检验表

n	90.00%	95.00%	97.50%	99.00%	99.50%
3	1.148	1.453	1.155	1.155	1.155
4	1.425	1.463	1.481	1.492	1.496
5	1.602	1.672	1.715	1.749	1.764
6	1.729	1.822	1.887	1.944	1.973
7	1.828	1.938	2.020	2.097	2.139
8	1.909	2.032	2.126	2.221	2.274
9	1.977	2.110	2.215	2.323	2.387
10	2.036	2.176	2.290	2.410	2.482
11	2.088	2.234	2.355	2.485	2.564
12	2.134	2.285	2.412	2.550	2.636
13	2.175	2.331	2.462	2.607	2.699
14	2.213	2.371	2.507	2.659	2.755
15	2.247	2.409	2.549	2.705	2.806
16	2.279	2.443	2.585	2.747	2.852
17	2.309	2.475	2.620	2.785	2.894
18	2.335	2.501	2.651	2.821	2.932
19	2.361	2.532	2.681	2.954	2.968
20	2.385	2.557	2.709	2.884	3.001



4. 粗大误差的处理 计算实例5 (续)

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
结果	20.42	20.43	20.40	20.43	20.42	20.43	20.39	20.30	20.40	20.43	20.42	20.41	20.39	20.39	20.40

第1种方法

算术平均值是: $\bar{X} = 20.404$

标准差是: $\delta = 0.033$

第8个测量误差的绝对值是: $0.104 > 3\delta = 0.099$

结论：第8个点为粗大误差，剔除后再进行计算均在误差要求范围之内。

第2种方法： 算术平均值是:20.404 标准差是: $\delta = 0.033$

确定一个可疑值：20.30 计算 $G_i = \frac{20.404 - 20.30}{0.033} = 3.152$

查表得相应临界值 $GP(15, 0.05) = 2.409 < 3.152$

结论：第八个值确为粗大误差，应予剔除。再往下继续计算的结果表明剩余14个测量结果不含有粗大误差。



谢谢
Q & A