



# 第八章 分析实验数据处理

## 8-1 数据的特征及分布

### 一、数理统计的某些基本概念

#### 1、总体与样本

总 体 : 被研究对象某特性值的全体。

个 体 : 组成总体的每个单元。

样本 (子样) : 自总体中随机抽取的一部分个体。

样 本 容 量 : 样本中所包含个体的数目, 用 $n$ 表示。

#### 2、随机变量

来自同一总体的无限多个测量值都是随机出现的, 叫随机变量。

多次重复测定的结果不能相同, 也不能事先知道。测量值一经取定就是一个常量, 不再有随机性。

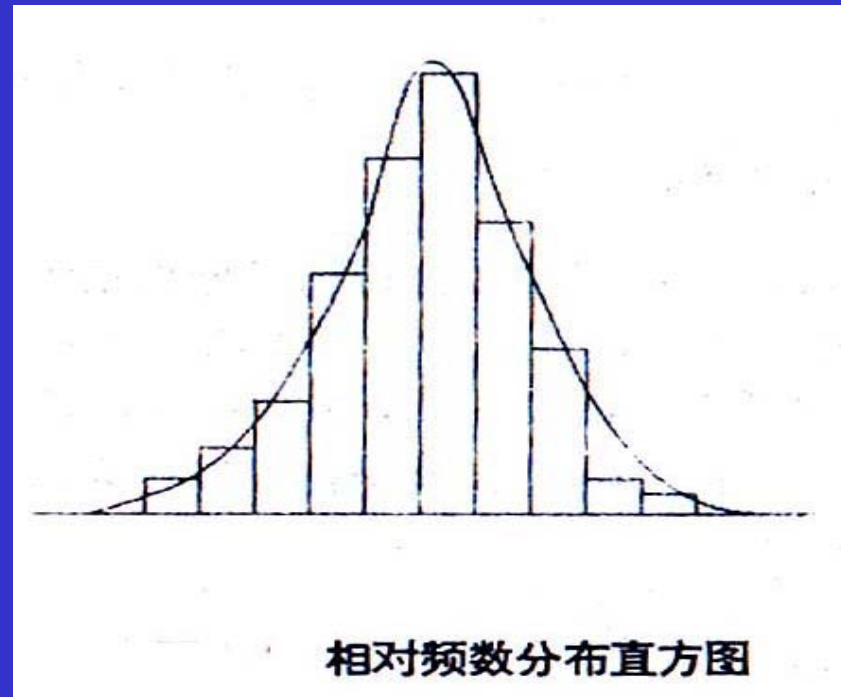
## 二、随机变量的频数分布

**例** 学生测定  $\text{BaCl}_2 \cdot 2\text{H}_2\text{O}$  试剂中 Ba 含量 (%)，共有 190 个数据，区间为 55.48%-56.46%。将这些数据按组距 0.1 来分成 10 组。

**频数**：每组中数据的个数

**相对频数**：频数在总测定次数中所占的分数

以各组区间为底相对频数为高做成一排矩形的相对频数分布直方图



## 其特点：

1. 离散特性：测定值在平均值周围波动，波动的程度用总体标准偏差表示

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

总体标准偏差

2. 集中趋势：向平均值集中

$$\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i$$

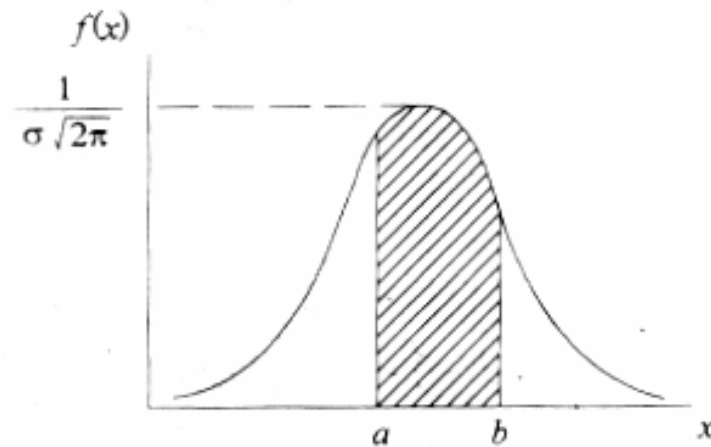
总体平均值

在确认消除系统误差的前提下，总体平均值就是真值。

### 三、 随机变量的正态分布

当数据无限多，组分的很细时，上述直方图则变成一条圆滑的曲线，称为**正态分布**。其数学表达式称为**正态分布密度函数**

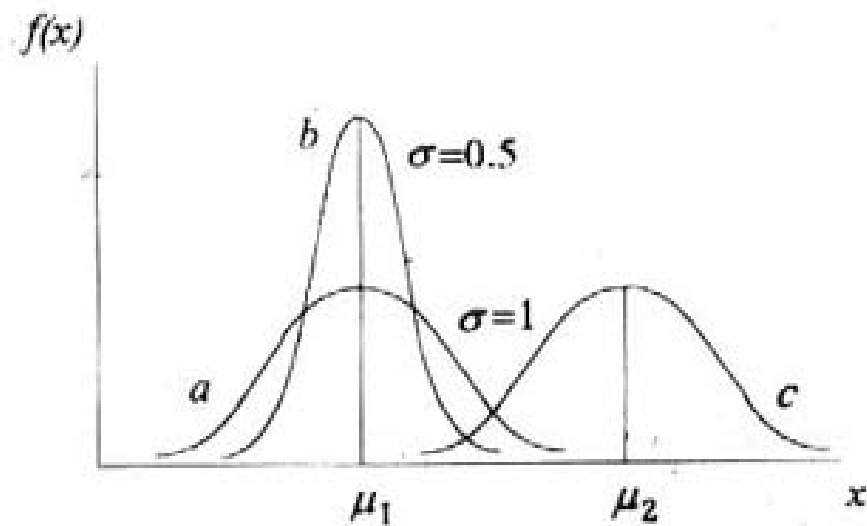
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



正态分布曲线

其两个重要参数为  $\mu$ 、 $\sigma$ ，记为  $N(\mu, \sigma^2)$ ， $\mu$  决定曲线在  $x$  轴的位置， $\sigma$  决定曲线的形状

$\sigma$  小  $\rightarrow$  曲线高、陡峭、  
精密度好；  
 $\sigma$  大  $\rightarrow$  曲线低、平坦、  
精密度差。



正态分布的两个参数

## 随机误差符合正态分布

1. 大误差出现的几率小，小误差出现的几率大。
2. 绝对值相等的正负误差出现的几率相等。
3. 误差为零的测量值出现的几率最大。

所有测量值出现的概率总和应为1，即

$$P(-\infty, \infty) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

求变量在某区间出现的概率，即对该区间求积分

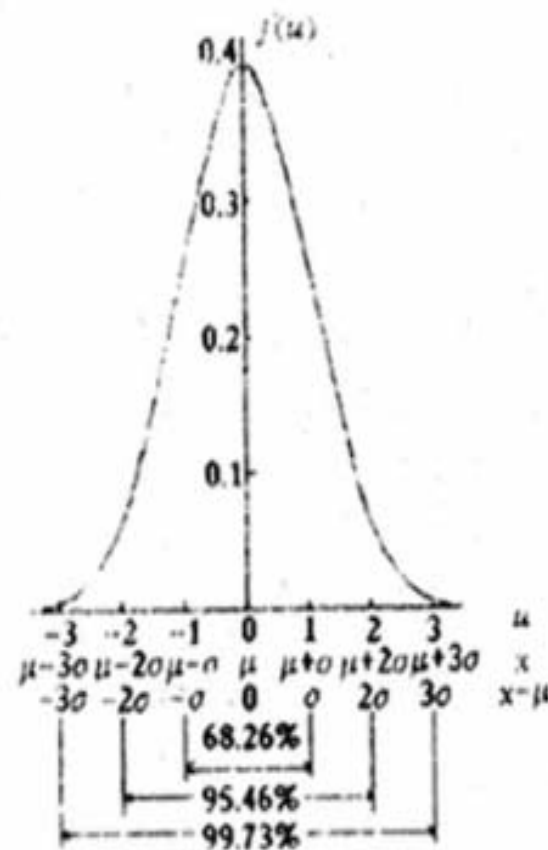
$$P(a, b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

对于不同的 $\mu$ 、 $\sigma$ 有不同的曲线，积分比较麻烦，为简化做数学上的变量转换，令  $u = \frac{x - \mu}{\sigma}$

将上式两端微分得  $\sigma du = dx$

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

此时正态分布转换为标准正态分布，记为 $N(0,1^2)$ 。

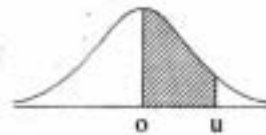


标准正态分布



积分  $P = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{u^2}{2}} du$  已被做成概率积分表

正态分布概率积分表



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

u	面积	u	面积	u	面积
0.0	0.0000	1.0	0.3413	2.0	0.4773
0.1	0.0398	1.1	0.3643	2.1	0.4821
0.2	0.0793	1.2	0.3849	2.2	0.4861
0.3	0.1179	1.3	0.4032	2.3	0.4893
0.4	0.1554	1.4	0.4192	2.4	0.4918
0.5	0.1915	1.5	0.4332	2.5	0.4938
0.6	0.2258	1.6	0.4452	2.6	0.4953
0.7	0.2580	1.7	0.4554	2.7	0.4965
0.8	0.2881	1.8	0.4641	2.8	0.4974
0.9	0.3159	1.9	0.4713	2.9	0.4987

$u$	$x - \mu$	$x$	$P\%$
$u = \pm 1$	$\pm \sigma$	$\mu - \sigma$ $\mu + \sigma$	$P = 2 \times 0.3413 = 0.6826 = \mathbf{68.26\%}$
$u = \pm 2$	$\pm 2\sigma$	$\mu - 2\sigma$ $\mu + 2\sigma$	$P = 2 \times 0.4773 = 0.9546 = \mathbf{95.46\%}$
$u = \pm 3$	$\pm 3\sigma$	$\mu - 3\sigma$ $\mu + 3\sigma$	$P = 2 \times 0.4987 = 0.9974 = \mathbf{99.74\%}$

<例> 按照正态分布 $x$ 在区间 $(\mu - 0.5\sigma, \mu + 1.5\sigma)$ 出现的概率

解: 根据  $u = \frac{x - \mu}{\sigma}$

可将  $\mu - 0.5\sigma$      $x$      $\mu + 1.5\sigma$  变换为  $-0.5$      $u$      $1.5$

查表  $u = 0.5$  时 面积为0.1915

$u = 1.5$  时 面积为0.4332

则  $-0.5$      $u$      $1.5$  的总面积即为 $x$ 在区间 $(\mu - 0.5\sigma, \mu + 1.5\sigma)$ 出现的概率  $P = 0.1915 + 0.4332 = 0.6247$

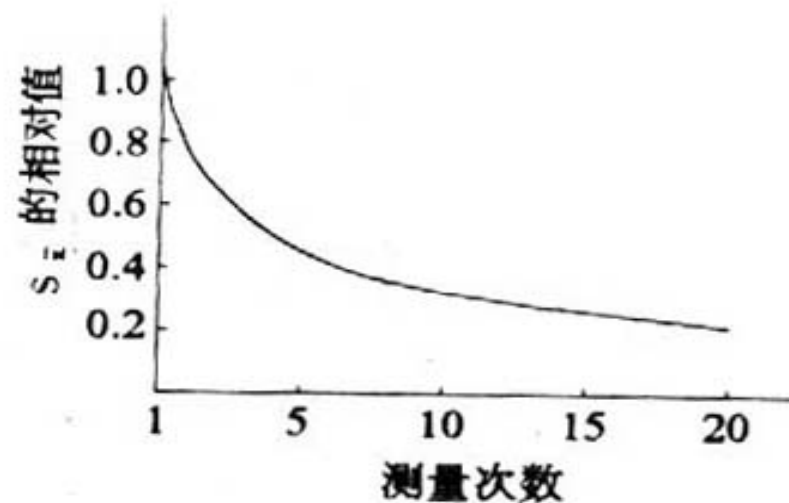
## § 8-2 总体平均值的估计

### 一、平均值的标准偏差

多个样本有多个平均值  $\bar{x}_1$ 、 $\bar{x}_2$  ……  $\bar{x}_n$  也遵循正态分布

根据数学推导  $S_{\bar{x}} = \frac{S}{\sqrt{n}}$        $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

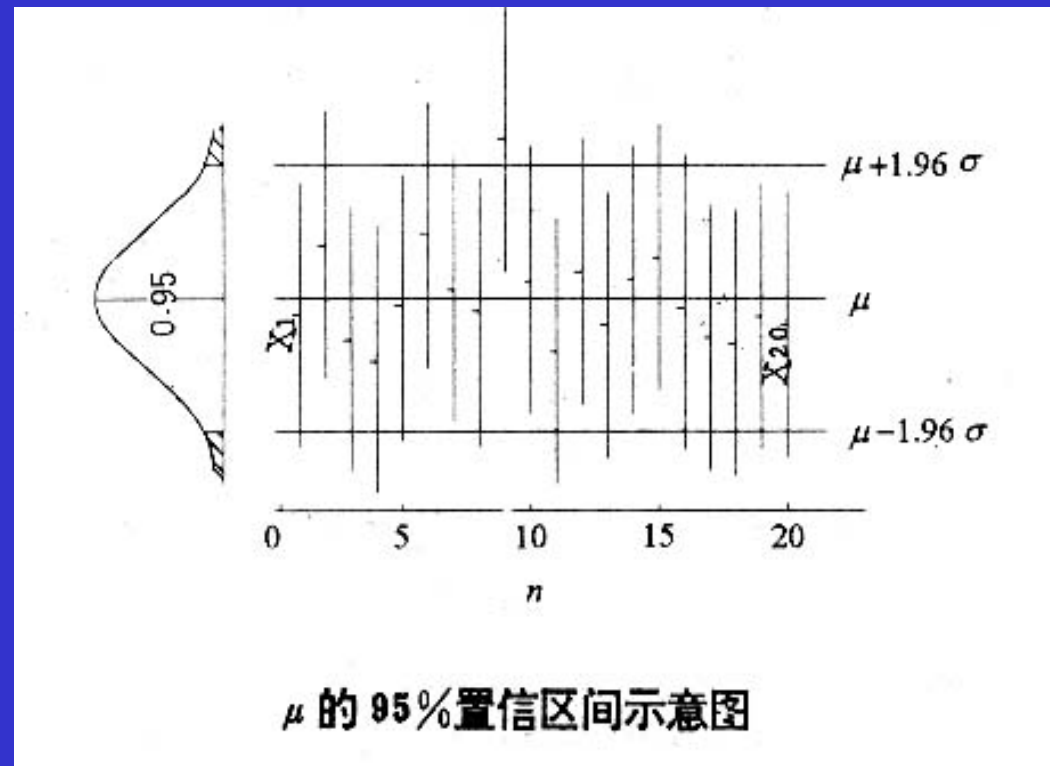
由此可见增加测定次数  
可使平均值的标准偏差  
减小，一般分析工作平  
行测定4-6次即可。



## 二、置信区间与置信度

**置信度**：做某种判断的把握性，用 $P$ 表示。

若 $x$ 落在以 $\mu$ 为中心 $1.96\sigma$ 为半长的区间里，那么以这个 $x$ 值为中心， $1.96\sigma$ 为半长的区间必然将 $\mu$ 包含在内。一切可能的区间里有95%的区间包含有 $\mu$ 。



**置信区间**：在一定的概率下，以测量值为中心包含总体平均值在内的区间。

置信区间表示如下：

$$\mu = x \pm u\sigma \quad (u \text{ 是由所定概率决定的})$$

$P$

$$90\% \quad u=1.64 \quad \mu = x \pm 1.64$$

$$95\% \quad u=1.96 \quad \mu = x \pm 1.96$$

$$99\% \quad u=2.58 \quad \mu = x \pm 2.58$$

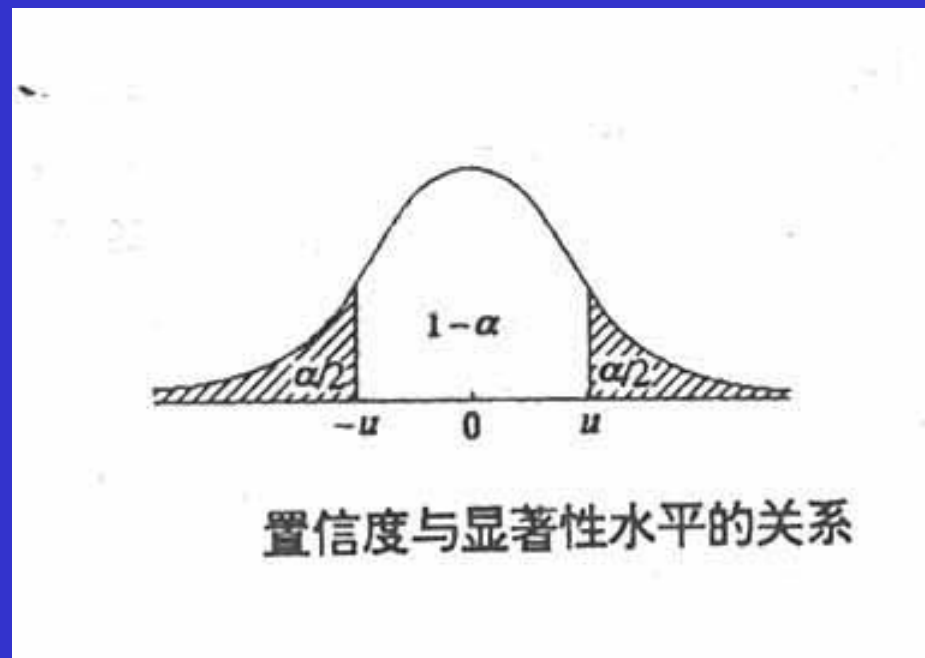
置信度太高无意义，太低不可信，一般定为90% 或95%。

$n$ 次测定总体平均值的置信区间为

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

### 三、显著性水平

表示测定值落在置信区间以外的概率，用  $\alpha$  表示， $\alpha=1-P$



<例> 已知测定NaCl试剂中Cl含量方法的标准偏差  $\sigma=0.05\%$ ，若分析结果为60.60%，计算95%置信度时总体平均值的置信区间，若（a）此结果为单次测定（b）此结果为四次测定的平均值。

解：单次测定  $\mu = x \pm u\sigma = 60.60 \pm 1.96 \times 0.05 = 60.60 \pm 0.10$

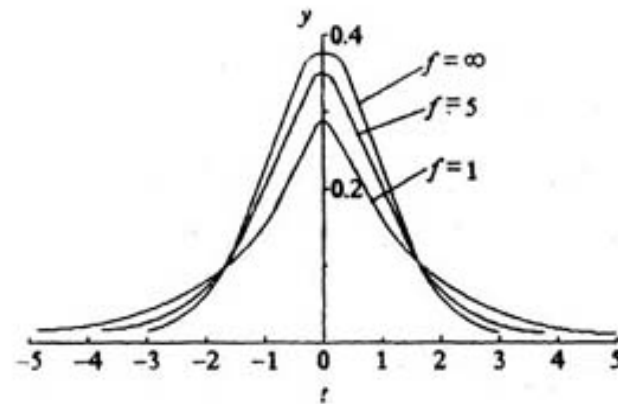
四次测定  $\mu = \bar{x} \pm u \frac{\sigma}{\sqrt{n}} = 60.60 \pm 1.96 \times \frac{0.05}{\sqrt{4}} = 60.60 \pm 0.05$

## 四、少量实验数据的统计处理

如果只做 $n$ 次测量， $\sigma$ 不知道，用 $S$ 代替 $\sigma$ ，要想得到同样的置信度，必须用一个比 $u$ 大的因子 $t$ 代替 $u$ ，随着测定次数 $n$ 的增加， $t$ 趋于 $u$  (用 $t$ 代替 $u$ 得到 $t$ 分布)

$$t\text{因子} \quad t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

$t$ 分布曲线与自由度  
( $f=n-1$ )有关



总体平均值的置信区间  $\mu = \bar{x} \pm t \frac{S}{\sqrt{n}}$

其中  $t_{\alpha, f}$  与置信度（用  $\alpha$  表示）和自由度有关，可查书上的  $t$  表（双侧表）。

<例> 测某Cu矿中Cu含量的四次结果分别为40.53, 40.48, 40.57, 40.42 求90%, 95% 置信水平时总体平均值的置信区间。

解：  $\bar{x} = 40.50$   $S=0.06$

查表  $P=90\%$  时  $t_{0.10, 3}=2.35$

$$\mu = \bar{x} \pm t \frac{S}{\sqrt{n}} = 40.50 \pm 2.35 \frac{0.06}{\sqrt{4}} = 40.50 \pm 0.07$$



查表  $P=95\%$  时  $t_{0.05,3}=3.18$

$$\mu = \bar{x} \pm t_{0.05,3} \frac{S}{\sqrt{n}} = 40.50 \pm 3.18 \frac{0.06}{\sqrt{4}} = 40.50 \pm 0.10$$

由此可见置信度越高，置信区间越大。

若  $x$  为6次测定的平均值，

则  $t_{0.05,5}=2.57$

$$\mu = \bar{x} \pm t_{0.05,5} \frac{S}{\sqrt{n}} = 40.50 \pm 2.57 \frac{0.06}{\sqrt{6}} = 40.50 \pm 0.06$$

## § 8-3 显著性检验

### 一、显著性检验的步骤

➤ 提出原假设 $H_0$ ：二者无显著性差异，二者相等，差异是由随机误差造成的，来自同一总体。

备择假设 $H_1$ ：二者不等（或大，或小）。

➤ 确定是单侧检验还是双侧检验，决定用什么样的表，或换算。

➤ 选定检验统计量，决定检验方法。

➤ 选定显著性水平，查出表值。

➤ 由样本值计算统计量，与表值比较，小于表值接受原假设，大于表值拒绝原假设而接受备择假设。

## 二、 $u$ 检验法

已知 $\sigma$ ，比较 $\bar{x}$ 与 $\mu$       统计量  $u = \frac{|\bar{x} - \mu|}{\frac{\sigma}{\sqrt{n}}}$

<例> 某钢铁厂在生产正常的情况下产品含碳量服从 $N(4.55, 0.11^2)$ ，某天某炉铁水含碳量(%)的分析结果为4.48, 4.40, 4.45, 4.46, 4.50, 问这炉铁水是否正常( $P=95\%$ )。

解：  $H_0: \bar{x}=\mu$  ,  $H_1: \bar{x} \neq \mu$  , 此题为双侧检验，使用 $u$ 表。

查表 $\alpha=0.05$ 时，  $u_{0.05}=1.96$ 。      求得 $\bar{x}=4.46$

$$u = \frac{|\bar{x} - \mu|}{\frac{\sigma}{\sqrt{n}}} = \frac{|4.46 - 4.55|}{\frac{0.11}{\sqrt{5}}} = 1.83 < u_{\text{表}}$$

接受原假设, 有95%的把握认为此炉铁水含碳正常。

### 三、t 检验法

不知道 $\sigma$ , 检验  $\bar{x}$  与  $\mu$  或  $\bar{x}_1$  与  $\bar{x}_2$

(1) 比较  $\bar{x}$  与  $\mu$

统计量

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

(2) 比较  $\bar{x}_1$  与  $\bar{x}_2$

统计量

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

$$\bar{S}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

<例> 某药厂产品中杂质铁含量为0.14%，改革生产工艺后取样分析5次结果得到  $\bar{x}=0.13$   
 $S=0.01$ ，问杂质含量是否明显降低  
( $P=95\%$ )。

解：  $H_0: \bar{x} = \mu$      $H_1: \bar{x} < \mu$

单侧检验，  $\alpha = 0.05$

$t$ 表为双侧表，查 $\alpha=0.10$ 表，得 $t_{0.10,4}=2.13$

$$t = \frac{|\bar{x} - \mu|}{\frac{S}{\sqrt{n}}} = \frac{|0.13 - 0.14|}{\frac{0.01}{\sqrt{5}}} = 2.24 > t_{\text{表}}$$

有显著性差异

有95%把握认为改革后杂质Fe含量降低。

## 四、F 检验法

比较  $S_1$  与  $S_2$  是否显著，F 表为单侧表

( $\alpha=0.05$ )。

统计量  $F = \frac{S_1^2}{S_2^2}$  (其中  $S_1 > S_2$ )

<例> 某试样用两种方法测定，结果如下：

方法 1  $\bar{x}_1 = 32.34$

$S_1 = 0.10$

$n_1 = 5$

方法 2  $\bar{x}_2 = 32.54$

$S_2 = 0.16$

$n_2 = 6$

试比较  $\bar{x}_1$  与  $\bar{x}_2$  是否有显著性差异 (P=90%)

解：先比较  $S_1$  与  $S_2$ ， $H_0: S_1 = S_2$   $H_1: S_1 \neq S_2$

双侧检验，选定  $\alpha = 0.10$  查表  $F_{0.05,5,4} = 6.26$

$$F = \frac{S_2^2}{S_1^2} = \frac{0.16^2}{0.10^2} = 2.56 < 6.26 \quad \text{不显著, 方差可以合并}$$

$$\bar{S} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{4 \times 0.1^2 + 5 \times 0.16^2}{5 + 6 - 2}} = 0.14$$

再比较  $\bar{x}_1$  与  $\bar{x}_2$ ,  $H_0: \bar{x}_1 = \bar{x}_2$ ,  $H_1: \bar{x}_1 \neq \bar{x}_2$ , 双侧检验

查表  $t_{0.10, 9} = 1.83$

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\bar{S}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{|32.34 - 32.54|}{0.14} \sqrt{\frac{5 \times 6}{5 + 6}} = 2.36 > t_{\text{表}}$$

有90%的把握认为两种方法间有系统误差

## 五、可疑值取舍

### 1、格鲁布斯 (Grubbs) 法

数据由小到大排列 求出  $\bar{x}$  与  $S$

统计量  $T = \frac{x - x_1}{S}$  ( $x_1$  为可疑值)

或  $T = \frac{x_n - x}{S}$  ( $x_n$  为可疑值)

将统计量值与表值  $T_{\alpha, n}$  比较,  $T_{\text{计}} > T_{\text{表}}$  时舍弃

<例> 一产品中杂质含量为 1.25%, 1.27%, 1.31%, 1.40%, 问 1.40% 是否保留 ( $P=95\%$ )。

解:  $\bar{x} = \frac{1.25 + 1.27 + 1.31 + 1.40}{4} = 1.31$        $S = 0.067$

$$T = \frac{1.40 - 1.31}{0.067} = 1.34$$

查表  $T_{0.05, 4} = 1.46$

$T < T_{\text{表}}$       应保留



## 2、Q 检验法

数据由小到大排列

$$\text{统计量 } Q = \frac{x_n - x_{n-1}}{x_n - x_1} \quad (x_n \text{ 为可疑值})$$

$$\text{或 } Q = \frac{x_2 - x_1}{x_n - x_1} \quad (x_1 \text{ 为可疑值})$$

将统计量值与表值  $Q_{p,n}$  比较,  $Q_{\text{计}} > Q_{\text{表}}$  舍弃

<例> 上例中的1.40是否保留 (置信度90%)。

$$Q_{\text{计}} = \frac{x_n - x_{n-1}}{x_n - x_1} = \frac{1.40 - 1.31}{1.40 - 1.25} = 0.60$$

查表  $Q_{0.90,4} = 0.76$   $Q_{\text{计}} < Q_{\text{表}}$  应保留



# 总结

离解常数、形成常数、逐级累积形成常数、总反应平衡常数、分布系数、副反应系数、条件常数

可行性判据、化学计量点、滴定突跃、滴定误差、指示剂作用原理、常用指示剂、标准溶液、应用实例、含量计算

沉淀的形成、沉淀的纯度、沉淀的条件

分离方法的原理（萃取）、应用

比耳定律、光度法的准确度、应用（定量、常数测定）

误差、偏差 ( $\bar{d}, S$ )、提高准确度的方法、有效数字

正态分布、置信区间、显著性检验