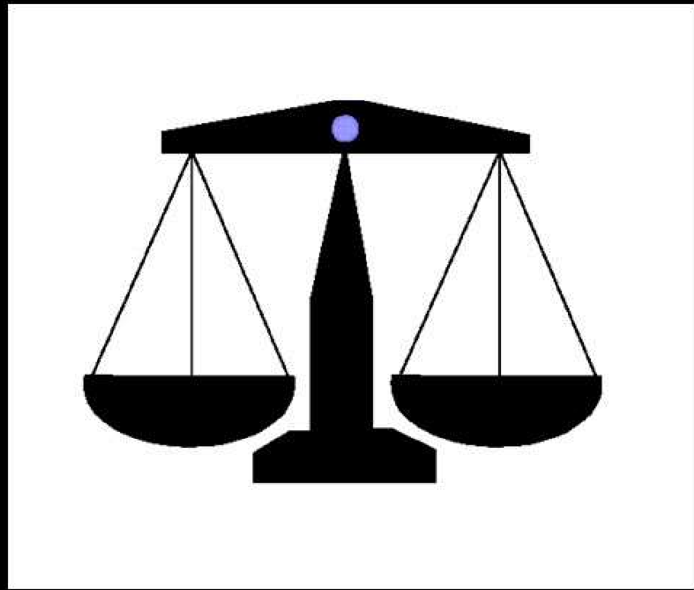
Ch. 8 误差和分析数据的处理



某金矿含Au < 1g / T,

却测成100g / T ??

某 … %,

去 …

说明误差无时不在，无处不有

公平、公正，实事求是

如何才能准呢？

例：铜矿标样12.06 %

平行测3次：

12.03，12.02，12.01(%)。 平均值12.02%

Sec.1 准确度与精密度

准确度表征测量值X与真实值T的符合程度。准确度用误差Ea表示。

$$Ea = X - T \qquad Er = \frac{Ea}{T} \times 100\%$$

精密度表征平行测量值的相互符合程度。精密度用偏差di表示。

$$\bar{d}_i = Xi - \bar{X} \qquad \overline{d_r} = \frac{\bar{d}}{\bar{X}} \times 100\%$$

$$\bar{X} = \frac{\sum Xi}{n} \qquad \bar{d} = \frac{\sum |di|}{n}$$

任何测量都带有误差，测量不能获得真值，可逐渐地逼近真值。

我们知道的真值有三类(相对性），相对真理

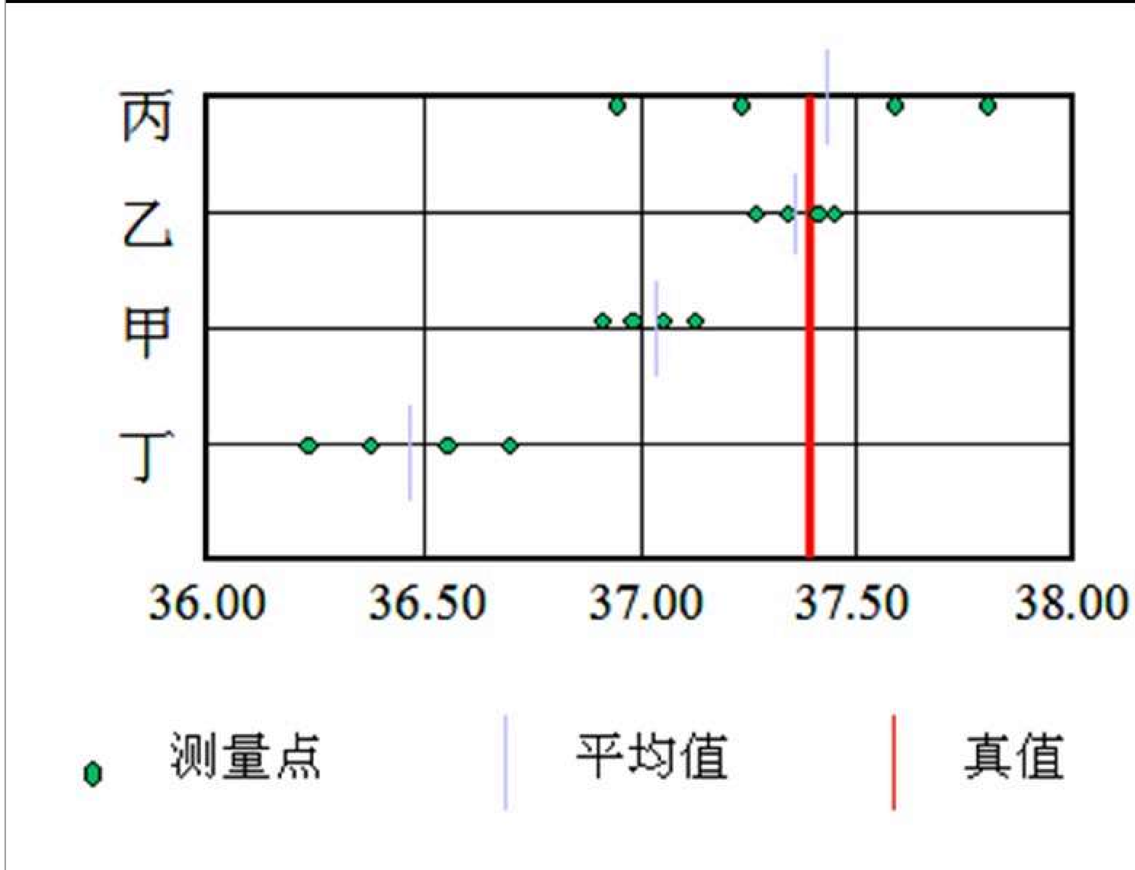
1、理论真值（如三角形三内角和等于180°、化合物的理论组成）

2、计量学约定真值（如国际计量大会确定的长度、质量、物质的量单位等等）

3、相对真值（如高一级精度的测量值相对于低一级精度的测量值；标准参考物质证书所给的数值）

准确度与精密度的关系

例1—1： 甲、乙、丙、丁四个分析工作者对同一铁标样（W_{Fe}= 37.40%)中的铁含量进行测量，得结果如图示，比较其准确度与精密度。



"准确度"高，精密度低

准确度高，精密度高

准确度低，精密度高

准确度低，精密度低

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。

精密度是保证准确度的必要条件，但不是充分条件

上张

Sec.2　系统误差与随机误差

1. 系统误差 —某种固定的因素造成的误差

2. 随机误差 —不确定的因素造成的误差

3. 过失误差—错误、责任事故

分析例1－1的原因：

分析工作者	系统误差	随机误差
甲	大	小
乙	小	小
丙	小（碰巧）	大
丁	大	大

1. 系统误差

固定原因、单向性

影响准确度

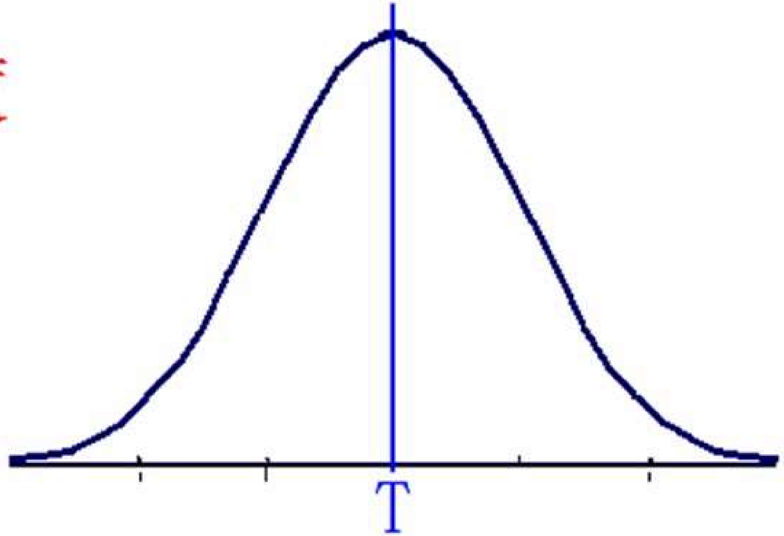
- 方法：选择方法
  - 标样
  - 标准方法
  - 回收
  - 校正结果
- 仪器、试剂：校正仪器、试剂，空白实验.....
- 操作：内检、外检、对照实验......

系统误差来自固定的原因，其大小可以估算或计算出来，从而可设法对测定值进行校正，直至消除。

上一张 下一张

2. 随机误差———影响精密度

随机、不固定原因，时大时小，时正时负，$n \to \infty$，服从正态分布。



例如：环境温度、湿度、气压的微小波动；仪器性能的微小变化；分析天平小数后四位、滴定管小数后第二位估计不准......

增加平行测定次数，在校正系统误差的前提下，$n \to \infty$，$\bar{X} \to \mu \to T$

上一张 下一张

3. 过失——这不是误差，是责任事故！应杜绝！

因缺乏经验、粗枝大叶、过渡疲劳、操作不正确引起，如跳样、加错试剂、读错刻度、记录或计算错误……

提高工作责任心！！！

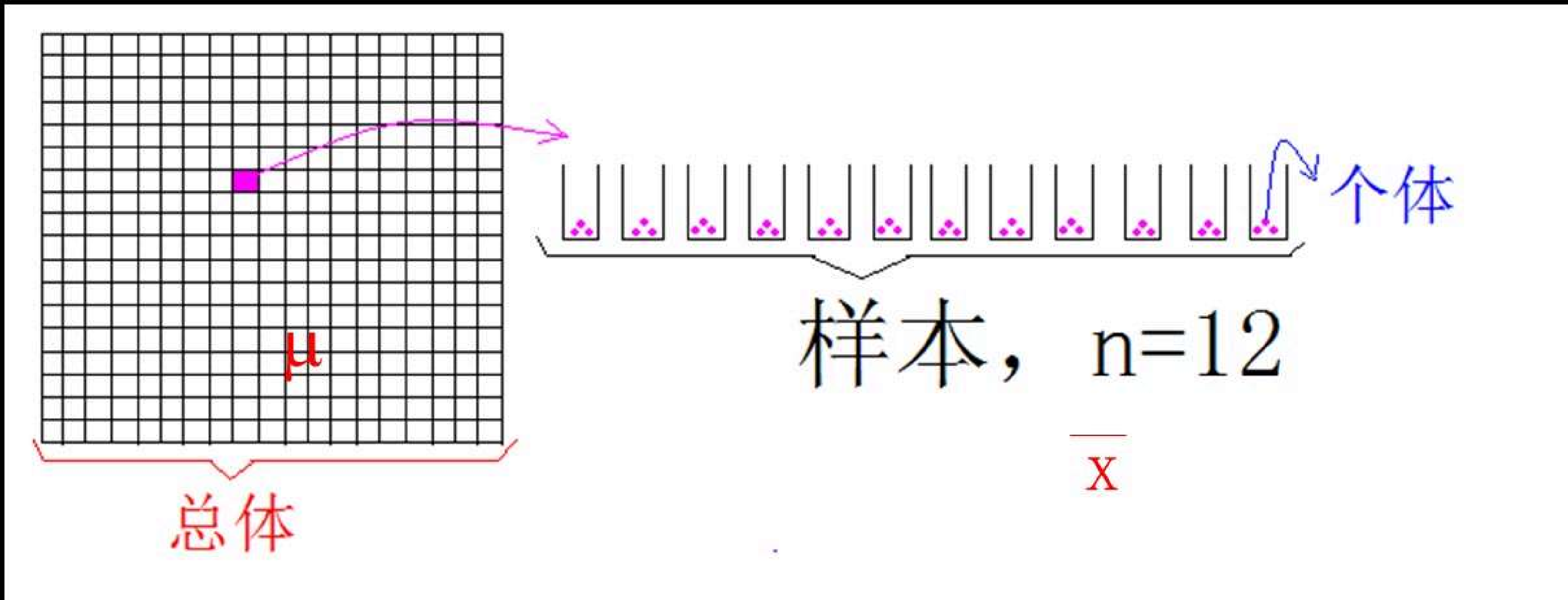
上一张

Sec.3 精密度的另种表示

1. 总体（母体）—所考察对象的全体

2. 样本（子样）—自总体中随机抽出的一组测量值

3. 样本大小（样本容量）—样本中所含测量值的数目



返回 Sec.5

样本：$\overline{X} = \dfrac{\sum Xi}{n}$　　总体：$n \to \infty, \mu = \dfrac{\sum Xi}{n}$

有限次数！　　无限次数！

即：$\lim\limits_{n\to\infty} \overline{X} = \mu \xrightarrow{\text{无系统误差的前提下}} T$

$$\bar{d} = \frac{\sum |Xi - \overline{X}|}{n} \qquad \bar{\delta} = \frac{\sum |Xi - \mu|}{n}$$

$$S = \sqrt{\frac{\sum (Xi - \overline{X})^2}{n-1}} = \sqrt{\frac{\sum d_i^2}{n-1}}; \qquad \sigma = \sqrt{\frac{\sum (Xi - \mu)^2}{n}} = \sqrt{\frac{\sum \delta_i^2}{n}}$$

相对标准偏差：

$$Sr = \frac{S}{\overline{X}} \times 100\%$$

$\overline{X1}$	$\overline{X2}$	$\overline{X3}$	$\overline{X4}$
$\overline{X5}$	$\overline{X6}$	$\overline{X7}$	$\overline{X8}$
$\overline{X9}$	$\overline{X10}$	$\overline{X11}$	$\cdots$

如果从同一总体中随机抽出容量相同的数个样本，由此可得到一系列样本的平均值：$\overline{X}_1, \overline{X}_2, \overline{X}_3, \cdots \overline{X}_n$

实践证明：$\overline{X}_1 \neq \overline{X}_2 \neq \overline{X}_3 \neq \cdots \neq \overline{X}_n$

这些平均值的分散程度，可以用样本平均值的标准偏差$\sigma_{\overline{X}}$表示。

统计学上业已证明：

$$S_{\overline{X}} = \frac{S}{\sqrt{n}} \qquad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

有限次

无限次

在日常分析中，一般平行测定：3—4次

较高要求：5—9次

最多：10—12次

2、适当地增加测定次数可提高结果的精密度。

粗看，杂乱无章

态分布

事实证明，在消除了系统误差的前提下，随机误差符合正态分布

一、频率分布（相

例：分析某镍试样，共

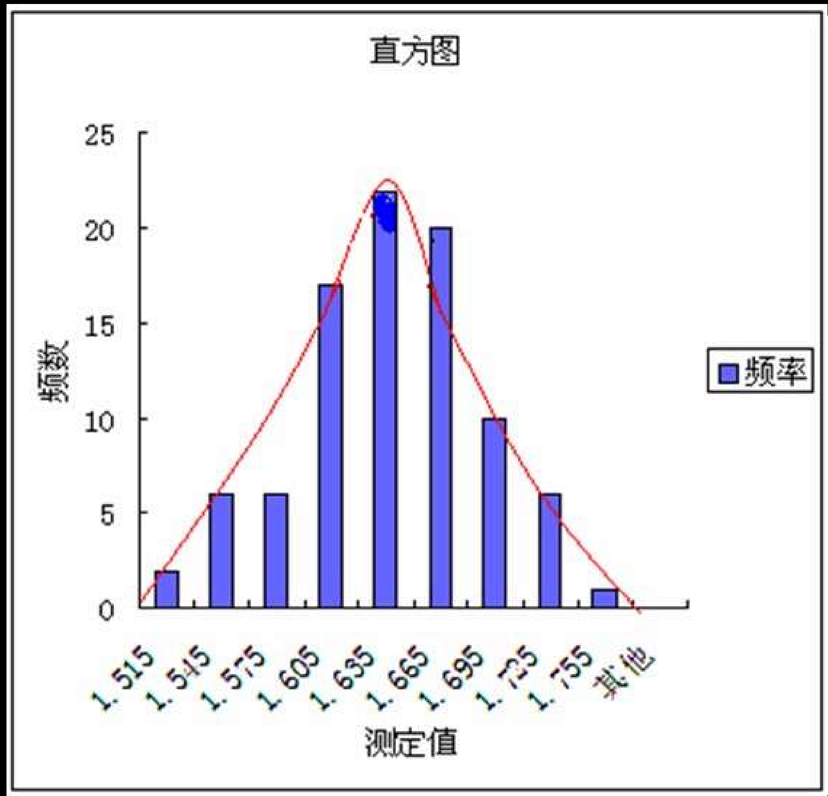
大部分介于1.57-1.67；小至1.49，大至1.74极少；基本上是围绕平均值1.62上下波动

	A	B	C	D		F	G	H	I	J
1	1.6	1.67	1.67	1.64	1.58	1.64	1.67	1.62	1.57	1.6
2	1.59	1.64	1.74	1.65	1.64	1.61	1.65	1.69	1.64	1.63
3	1.65	1.7	1.63	1.62	1.7	1.65	1.68	1.66	1.69	1.7
4	1.7	1.63	1.67	1.7	1.7	1.63	1.57	1.59	1.62	1.6
5	1.53	1.56	1.58	1.6	1.58	1.59	1.61	1.62	1.55	1.52
6	1.49	1.56	1.57	1.61	1.61	1.61	1.5	1.53	1.53	1.59
7	1.66	1.63	1.54	1.66	1.64	1.64	1.64	1.62	1.62	1.65
8	1.6	1.63	1.62	1.61	1.65	1.61	1.64	1.63	1.54	1.61
9	1.6	1.64	1.65	1.59	1.58	1.59	1.6	1.67	1.68	1.69

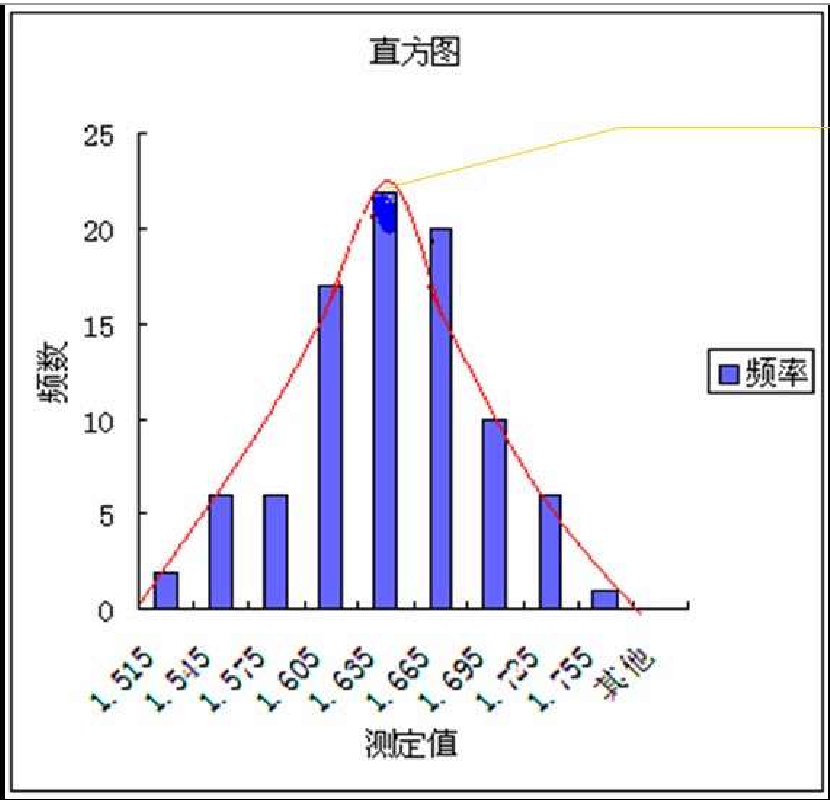
在单元格K1-K9中分别输入1.515;1.545;1.575;1.605;1.635;1.665;1.695;1.725;1.755（意思把上数据分成9组，

$$组距=\frac{极差}{组数}=\frac{1.74-1.49}{9}=0.03);$$

为避免骑墙现象，组界值比测定值多取一位。



选取【工具】、【数据分析】，在选【直方图】并输入相应的数值，可画出频率或频数直方图。



平均值1.62

1.从横轴看：对称，正、负误差出现的机会相等；

2.从纵轴看：大误差比小误差出现的机会少，极大的误差出现的机会极少。

规律：测量数据既集中又分散

二、正态分布

正态分布是法国数学家A. de Moivre 提出的，德国数学家Gauss在研究天文学中的观测误差时导出的正态分布曲线即Gauss曲线。所以正态分布又叫Gauss误差定律。

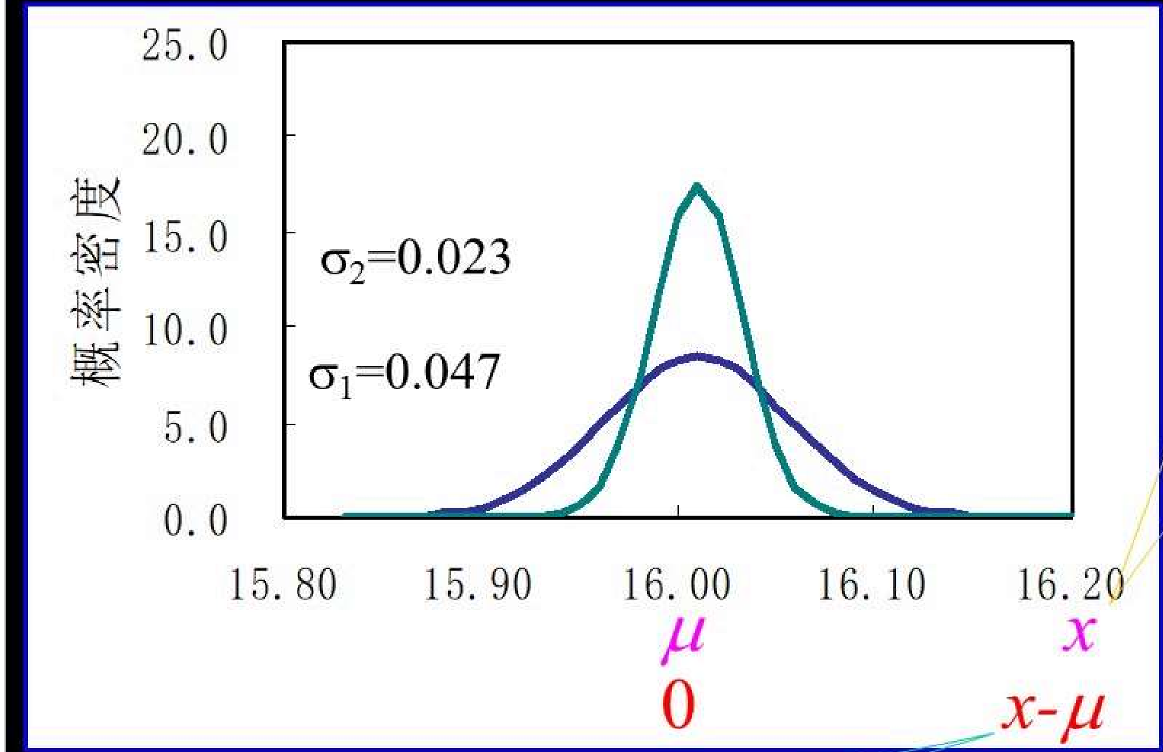
正态分布 的密度函数是:

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

y 概率密度　　x 个别测量值　　x-μ 随机误差

σ 总体标准偏差，表示无限次测量分散的程度。

μ 总体平均值，表示无限次测量值集中的趋势。



测量值的正态分布

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

正态分布曲线—

N(μ，σ²)

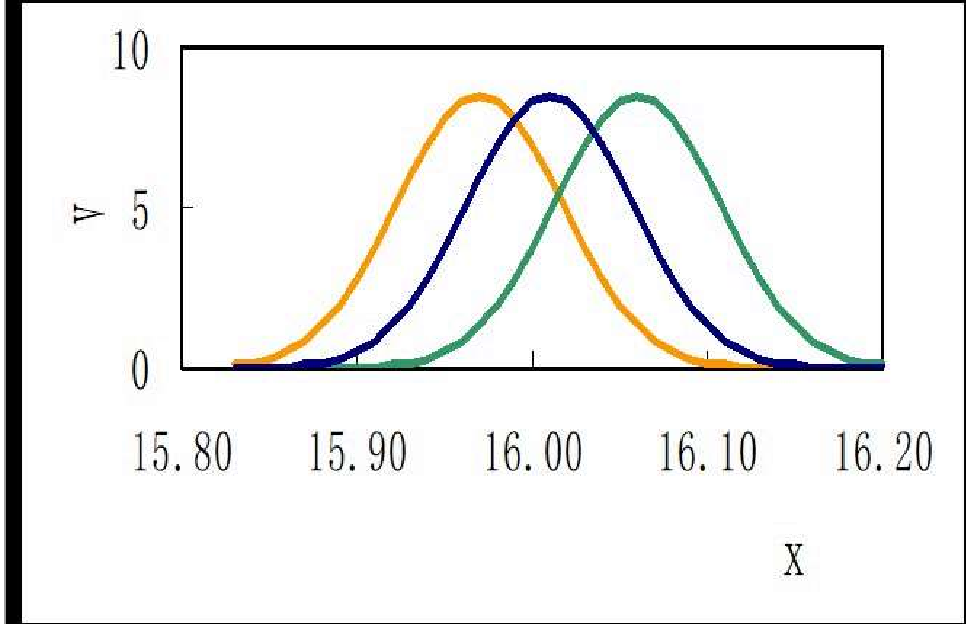
曲线的形状取决于μ，σ²，μ，σ²确定了，N(μ，σ²)也就定了。

随机误差的正态分布

标准正态分布曲线—

N(0，1)

后面详细介绍。

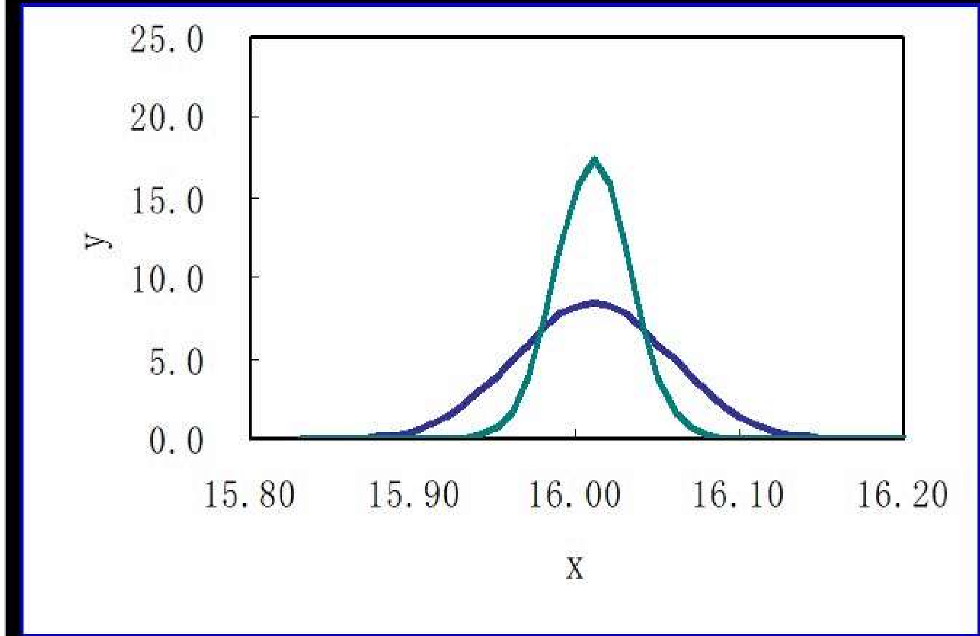


总体标准偏差σ 相同，
总体平均值μ不同

原因：

1、总体不同

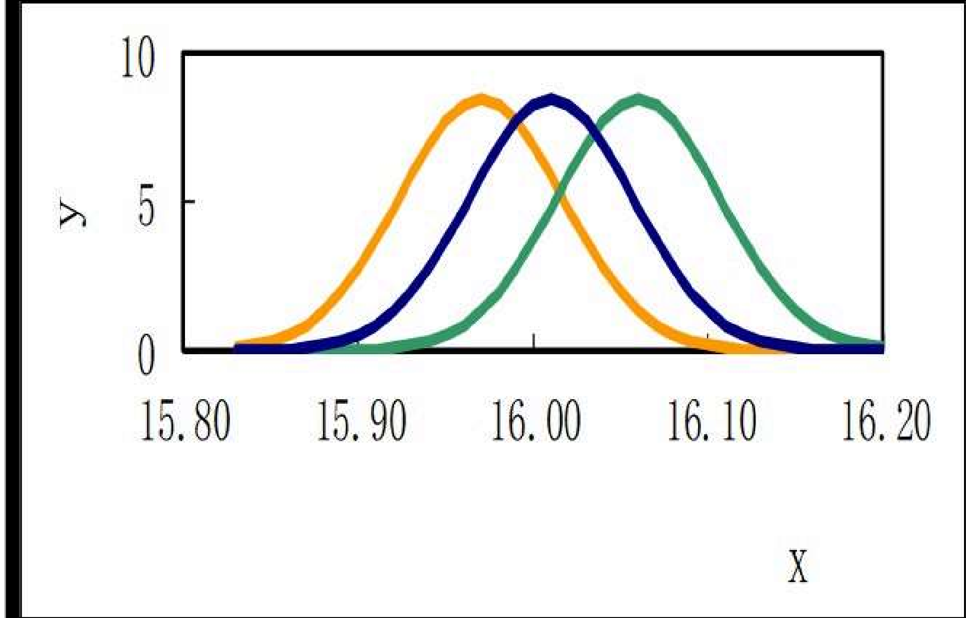
2、同一总体，存在系统误差

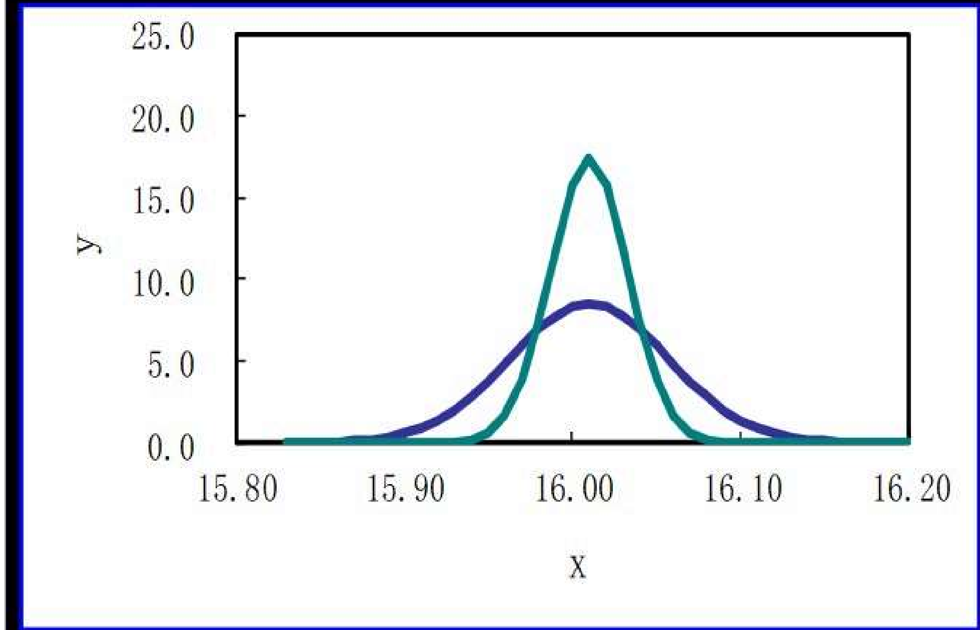


总体平均值μ相同，总体标准偏差σ不同

原因：

同一总体，精密度不同





不论怎样，μ与σ不同，图形就不同。应用起来不方便。解决方法：坐标变换

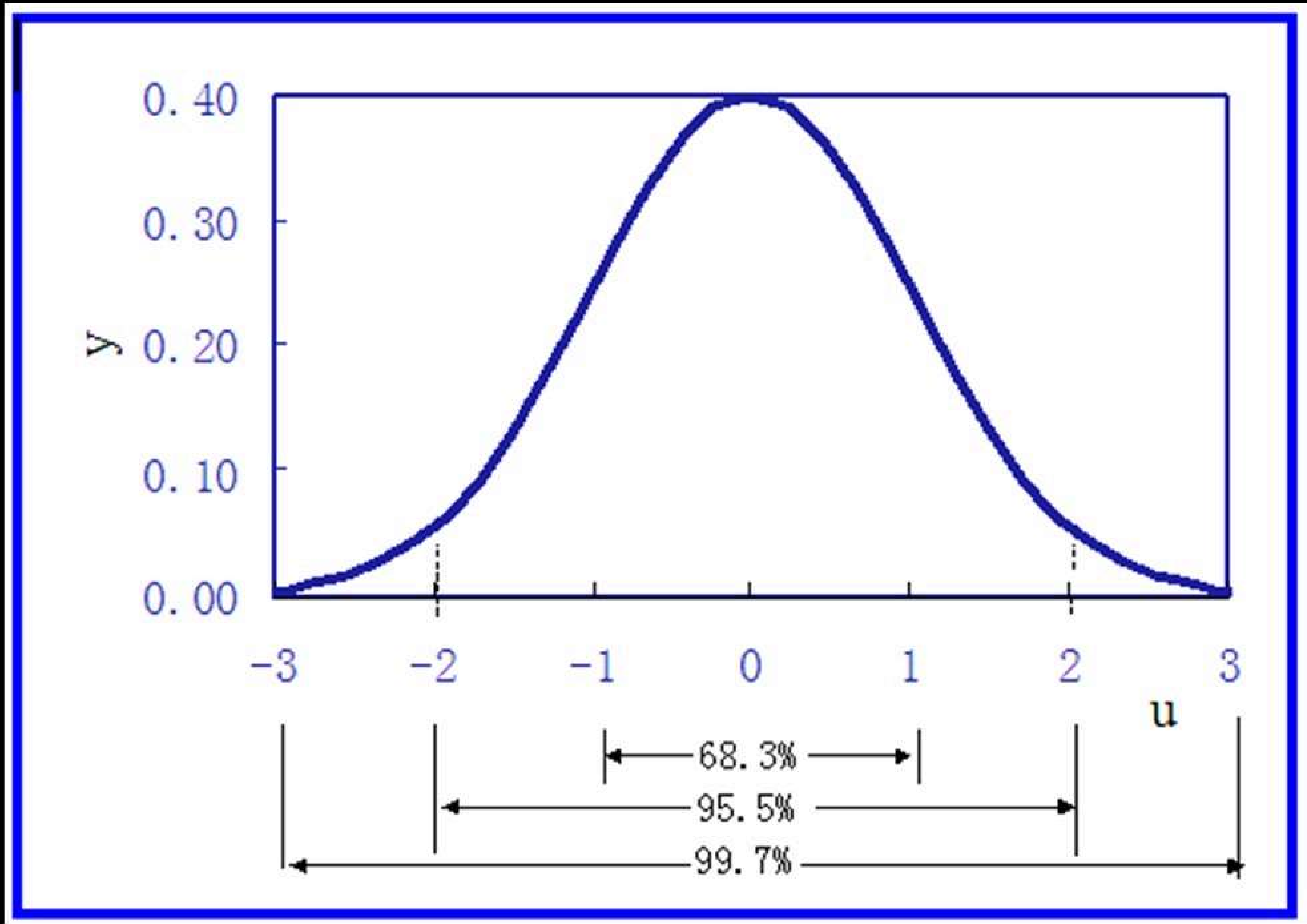
令：$u=\dfrac{X-\mu}{\sigma}$, 则$du=\dfrac{dx}{\sigma}$, $dx=\sigma du$

(1) 式 ：$y=f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ ……(1) 可变为：

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

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

$\therefore y=\Phi(u)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$ ……………(2), 以N(0,1)表示。



标准正态分布曲线N(0，1)就是以μ为原点，σ为单位的曲线，它对于不同的μ和σ的任何测量值都是通用的，如上图所示。

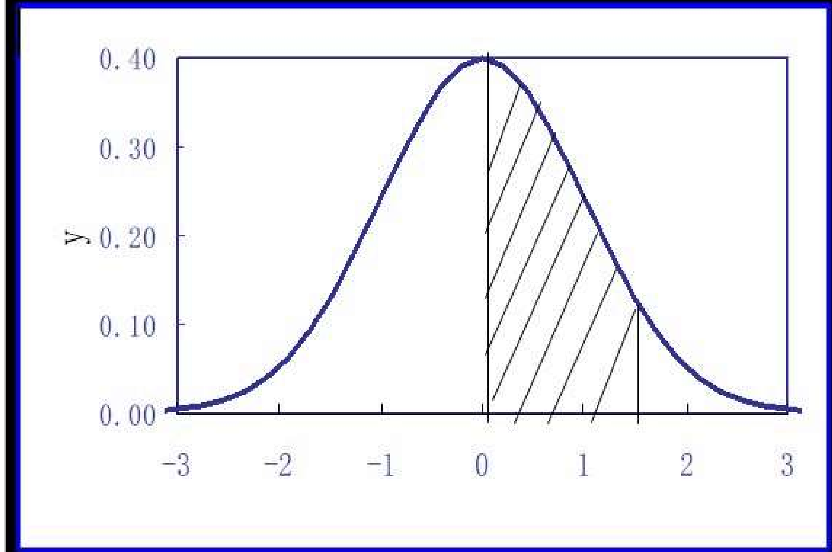
正态分布曲线与横坐标-∝到+∝之间所夹的总面积，代表所有数据出现的概率总和。其值应为1。即：

$$P=\int_{-\infty}^{+\infty}f(x)dx=\int_{-\infty}^{+\infty}\Phi(u)du=1$$

对于N(0, 1)，测量值的随机误差在某一区间内出现的概率（不同u值所占的面积）已用积分法求得，列于书P54页表3-1。表中所列之值为单边值。

随机误差的区间概率

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



$$概率 = \phi(u)du$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{u} e^{-u^2/2} du = 面积$$

正态分布概率积分表（部分数值）

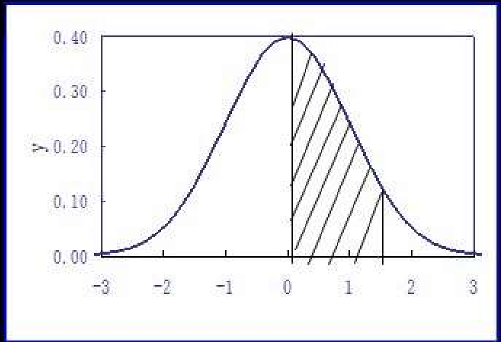
|u|	面积	|u|	面积	|u|	面积	|u|	面积
0.674	0.2500	1.000	0.3413	1.645	0.4500	1.960	0.4750
2.000	0.4773	2.576	0.4950	3.000	0.4987	∞	0.5000

返回例4—3　　返回例题

正态分布概率积分表（部分数值）返回例题4-1　返回例题

|u|	面积	|u|	面积	|u|	面积	|u|	面积
0.674	0.2500	1.000	0.3413	1.645	0.4500	1.960	0.4750
2.000	0.4773	2.576	0.4950	3.000	0.4987	∞	0.5000
0.500	0.1915	1.500	0.4332	2.500	0.4938		

测量值与随机误差的区间概率



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

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

随机误差出现的区间u（以σ为单位）	测量值x出现的区间	概率%
(-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

例4-1： 已知某试样中Co的标准值为1.75%，测得σ＝0.10, 又知测量时无系统误差，求结果落在（1）1.75 ±0.15% 概率； （2）测量值大于2 %的概率。

解（1）找u值：

$$u = \pm\frac{x-\mu}{\sigma} = \pm\frac{0.15}{0.10} = \pm 1.5$$

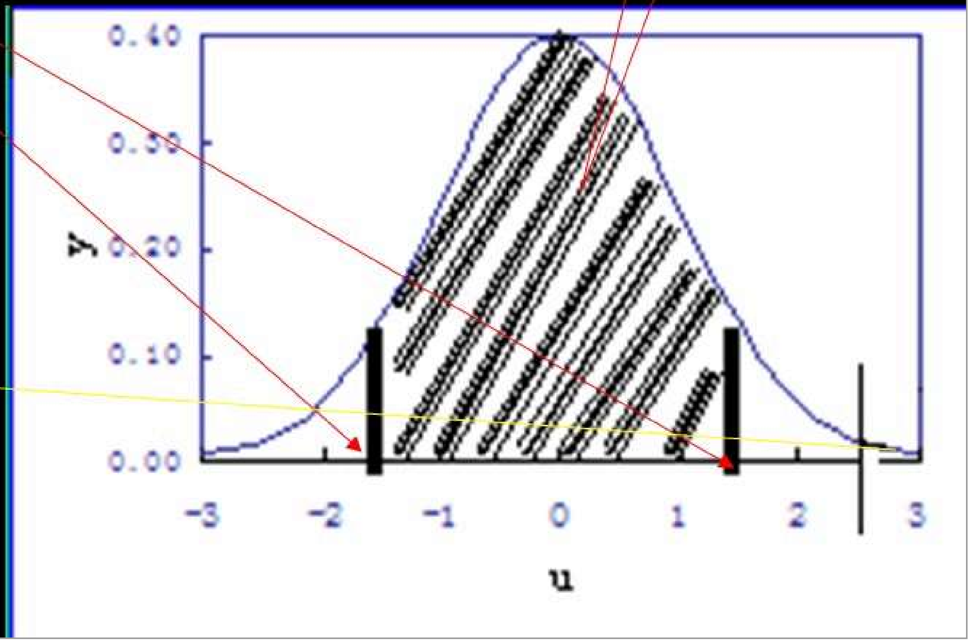
查表：u= ±1.5 时， 概率为：$2 \times 0.4332 = 0.866 = 86.6\,\%$

（2） $$u = \frac{2-1.75}{0.10} = 2.5$$

查表：u >2.5 时， 概率为：

$$0.5 - 0.4938 = 0.0062$$

$$= \boxed{0.62\%}$$

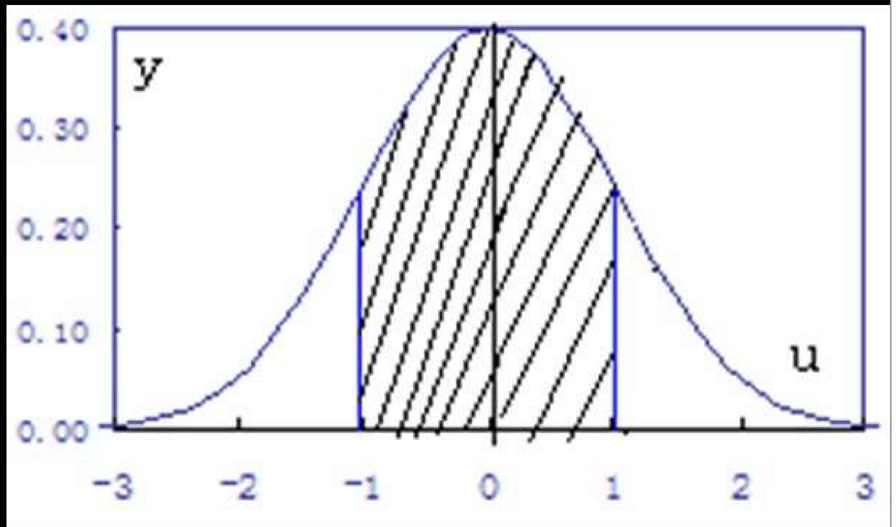


例4—3：根据正态分布概率积分表，计算单次测量值的偏差绝对值分别小于1 σ和大于1 σ的概率.

解：(1)单次测量值的偏差绝对值小于1 σ的概率，即：

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

现 $|x-\mu| < 1\sigma$，脱绝对值符号：



$$(x-\mu) < 1\sigma \qquad -(x-\mu) < 1\sigma$$

$$x-\mu < 1\sigma \qquad x-\mu > -1\sigma$$

$$\frac{x-\mu}{\sigma} < 1 \qquad \frac{x-\mu}{\sigma} > -1$$

即：$u > -1$；$u < 1$，属于双边内侧检验（图中蓝色阴影部分）

查表3－1

u=±1，面积0.3413，故P=0.3413×2=68.26%

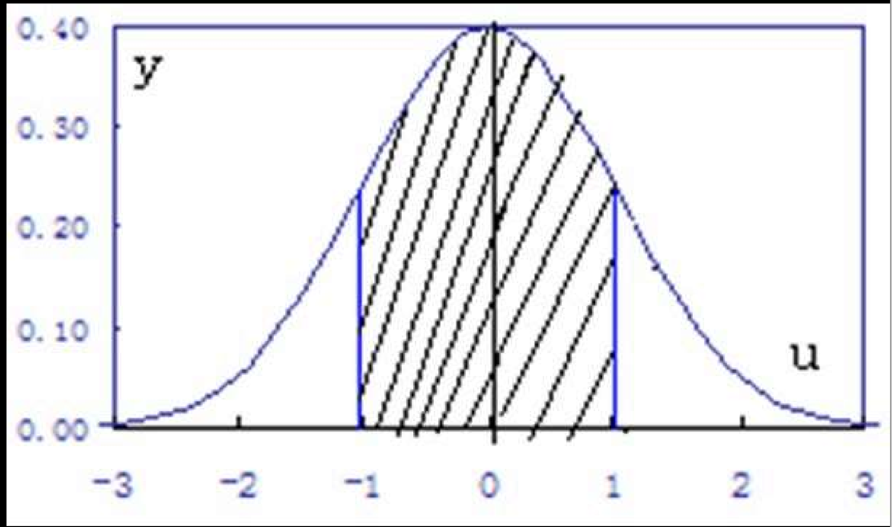
（2）单次测量值的偏差绝对值大于1σ的概率，即：

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

现 $|x-\mu| > 1\sigma$，脱绝对值符号：

$(x-\mu) > 1\sigma \qquad -(x-\mu) > 1\sigma$

$$\frac{x-\mu}{\sigma} > 1 \qquad \frac{x-\mu}{\sigma} < -1$$

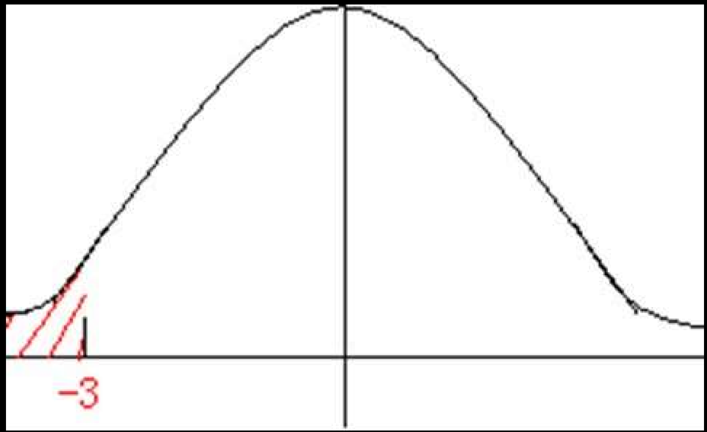


即：$u > 1$；$u < -1$，属于双边外侧检验（图中无阴影部分）

故：$P = 1 - 0.3413 \times 2 = 31.7\%$

例4－4：已知某金矿试样中含金量的标准值为12.2 g/T, σ = 0.2 g/T, 求分析结果小于11.6 g/T的概率.

解: 既然不是绝对值小于，而仅仅是小于，属单边检验



$$u = \frac{x-\mu}{\sigma} = \frac{11.6-12.2}{0.2} = -3$$

求x<11.6的概率，μ, σ为常数；也就是求u < -3的概率。

查表，u=3，面积＝0.4987

故P=0.5-0.4987=0.13%