

第三章

正态分布与医学参考值范围



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE

目录

- **第一节 正态分布**

- **第二节 医学参考值范围**

重点难点

掌握

正态分布的特征的两个参数，曲线下的面积分布规律及标准正态分布曲线下面积的计算；医学参考值范围的概念和两种计算方法。

熟悉

制订医学参考值范围的注意事项，尤其要注意依据专业知识进行单、双侧界值的选择。





第一节

正态分布

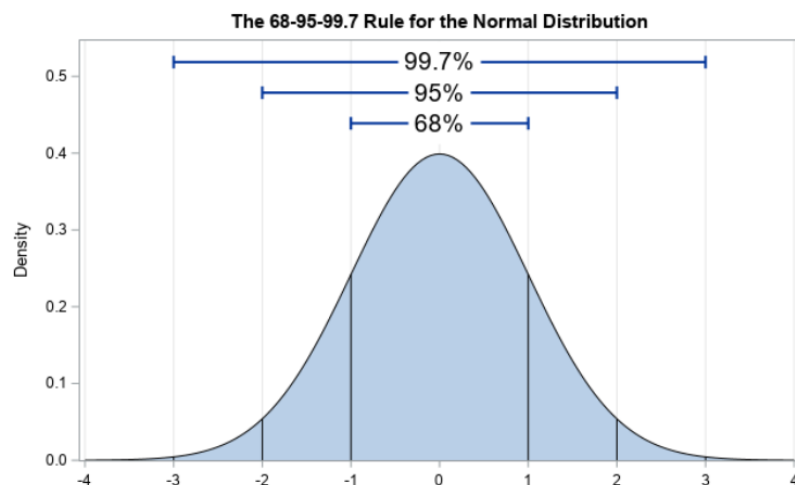


人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE

➤ **正态分布 (normal distribution) 又称为高斯分布。**首先由德国数学家和天文学家德·莫阿弗尔提出，高斯虽然发现稍晚，但他迅速将正态分布应用于天文学，并对其性质作了进一步的研究，使正态分布的应用价值广为人知。



德·莫阿弗尔

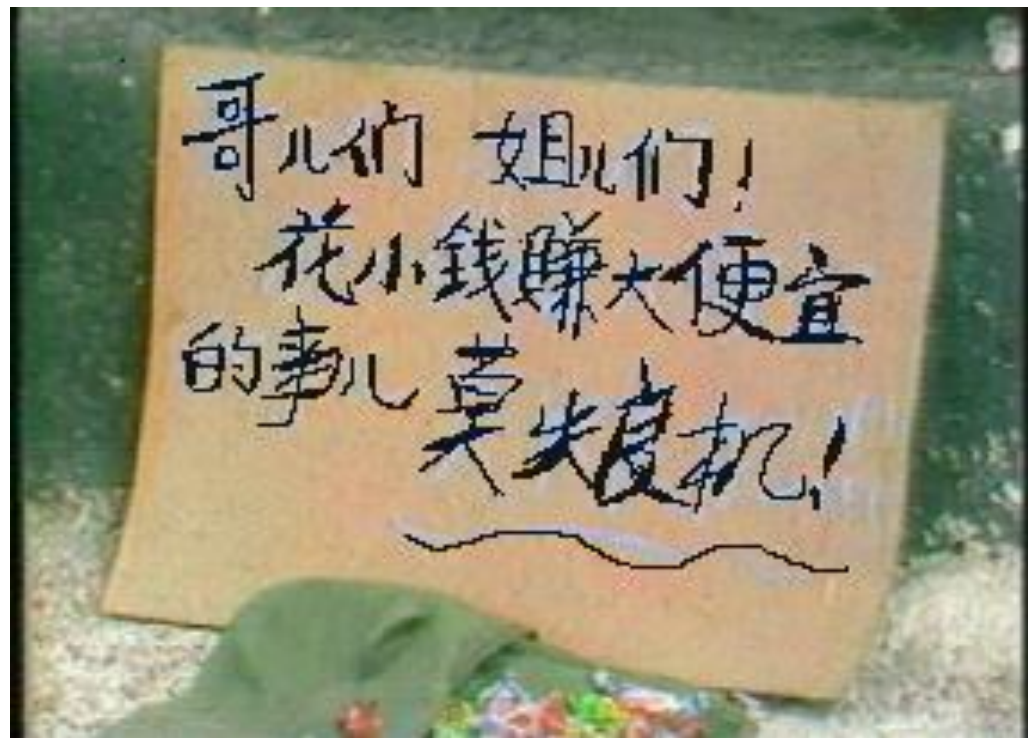


卡尔·弗里德里希·高斯

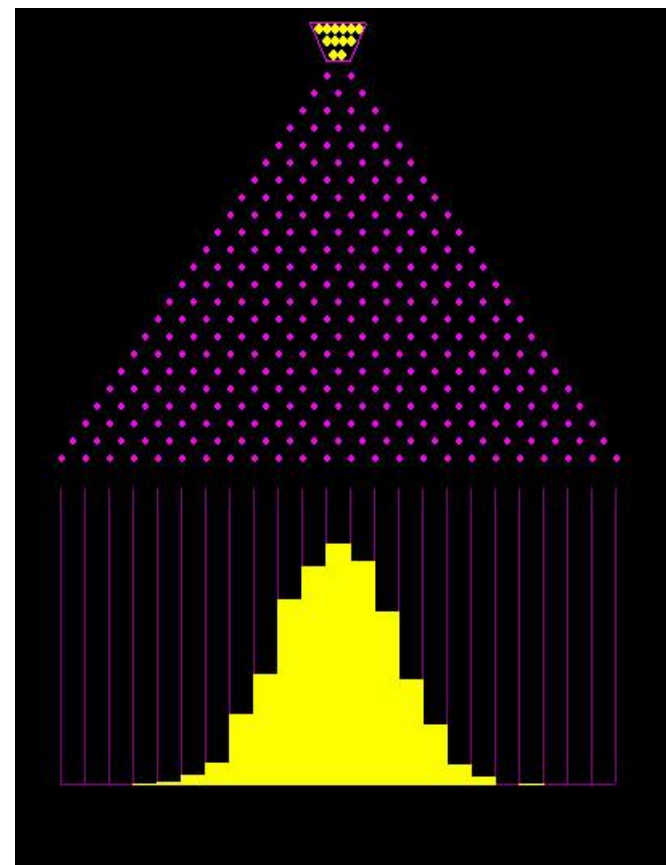
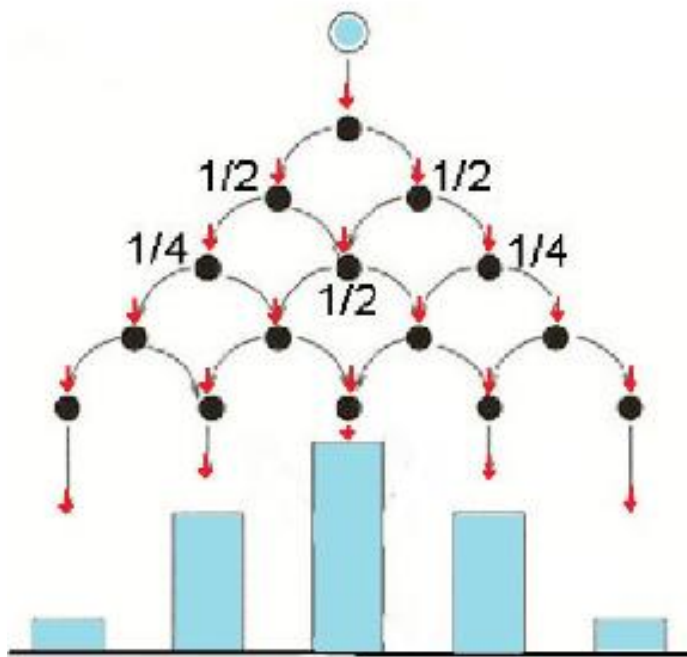
纸币上的数学家



面值10德国马克，1993年在德国发行。纸币正面印有高斯肖像，和他给出的正态分布函数（曲线）



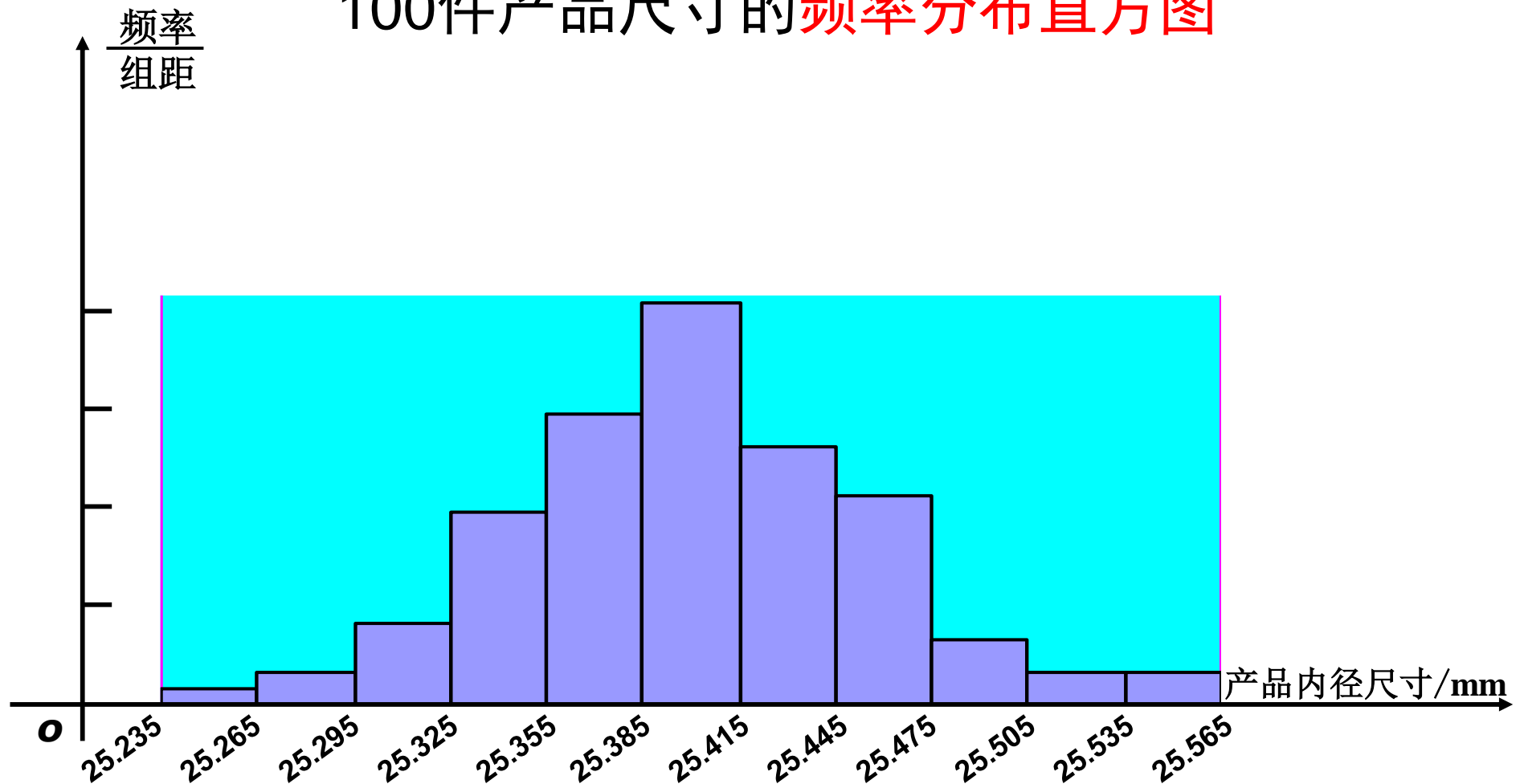
高尔顿钉板试验



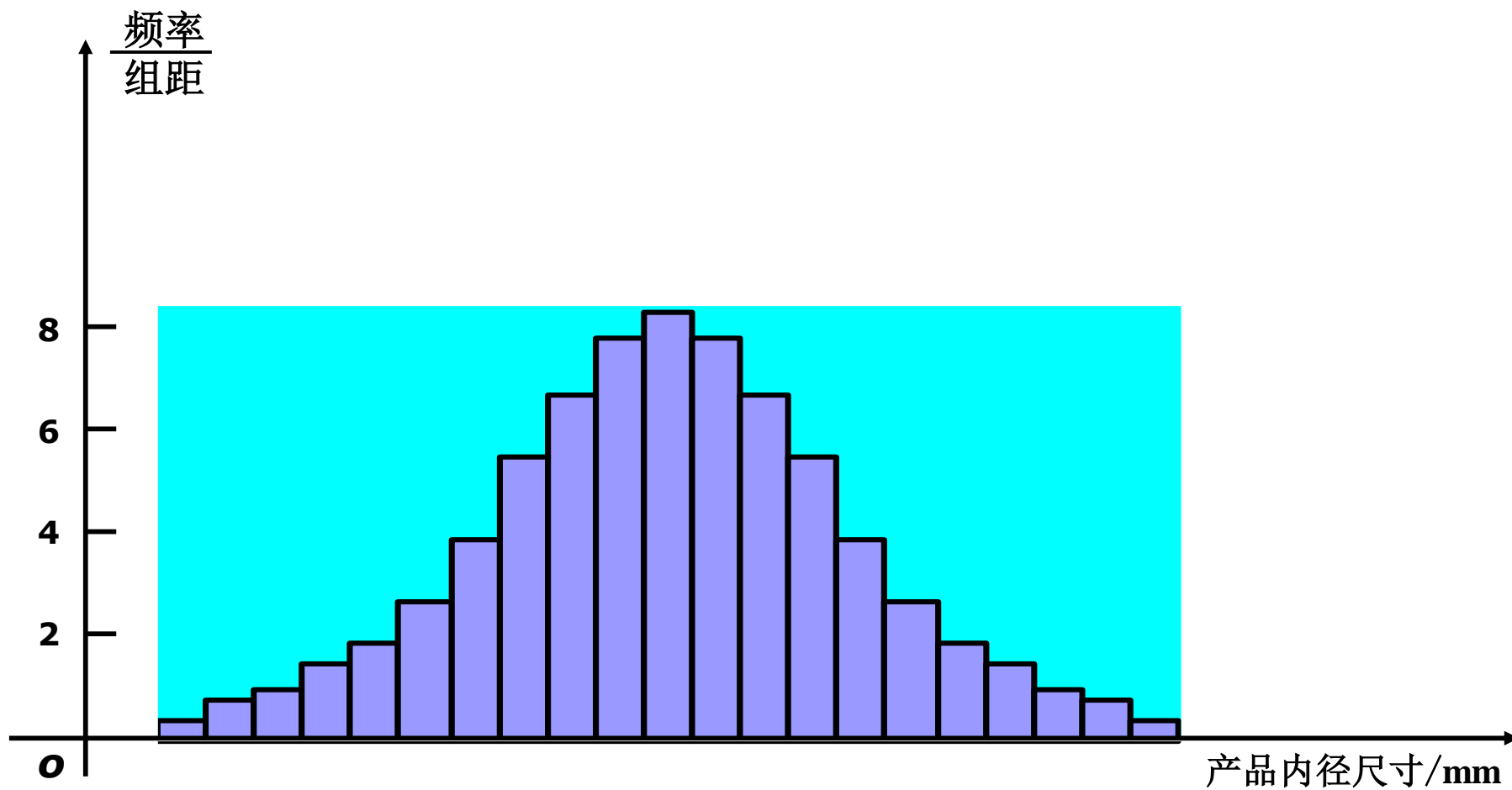
这条曲线就近似我们将要介绍的**正态分布**的密度曲线。

频率分布直方图

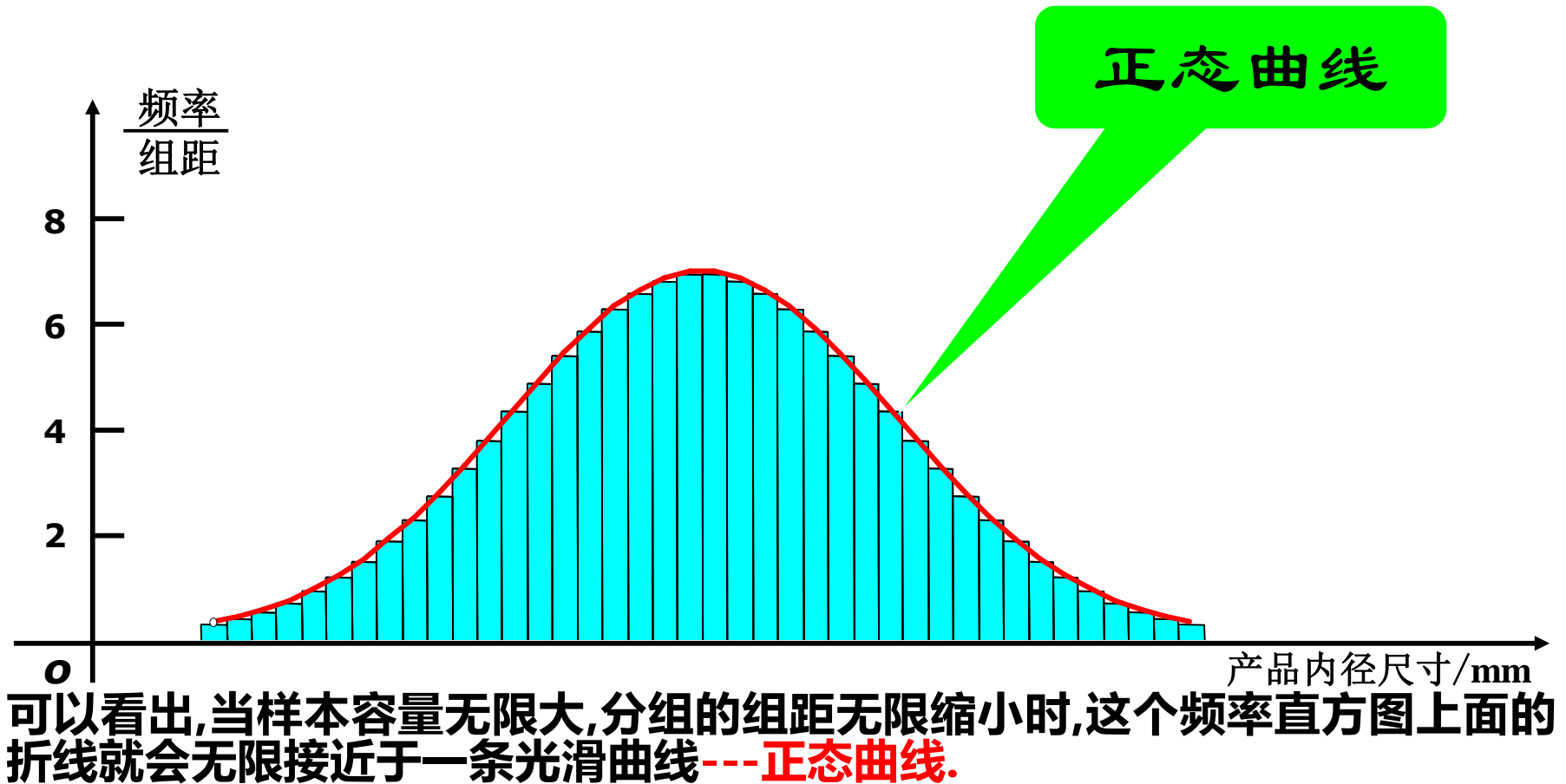
100件产品尺寸的频率分布直方图



200件产品尺寸的频率分布直方图



样本容量增大时频率分布直方图



什么样的随机变量服从正态分布呢？

经验表明，一个随机变量如果是众多的、互不相干的、不分主次的偶然因素作用结果之和，它就服从或近似服从正态分布。

在日常中，在正常生产条件下各种产品的质量指标；

在医学中，同一群体的某一特征；……；

在气象中，某地每年七月份的平均气温、平均湿度以及降雨量等，水文中的水位；

总之，正态分布广泛存在于自然界、生产及科学技术的许多领域中。

二、正态分布的特征

☑ 正态分布

- 连续型随机变量 X 服从正态分布，记为 $X \sim N(\mu, \sigma^2)$ ，正态分布有两个特征参数：
 μ 为 X 的**总体均数**、 σ^2 为 X 的**总体方差**。

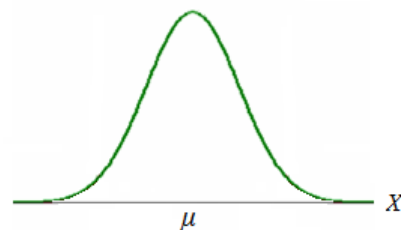
随机变量 X 的概率密度函数：

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

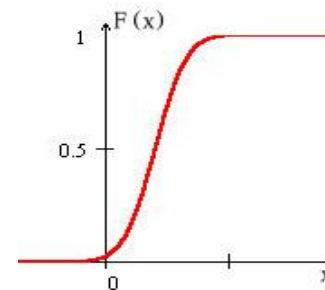
概率分布函数：

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad -\infty < x < \infty$$

正态分布密度曲线

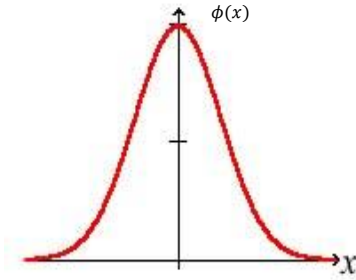


概率分布曲线



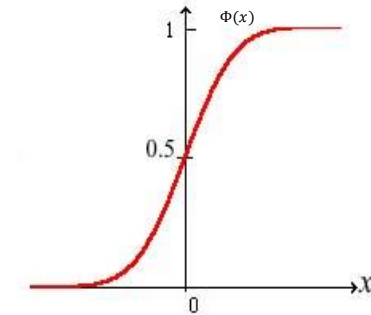
随机变量 X 的概率密度函数:

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



概率分布函数:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad -\infty < x < \infty$$



概率分布函数 $F(X)$ 又称概率累积分布函数, 是对概率密度函数求积分

$$F(x) = \int_{-\infty}^x f(x) dx$$

概率密度函数是概率分布函数 $F(X)$ 的导函数

$$F'(x) = f(x)$$

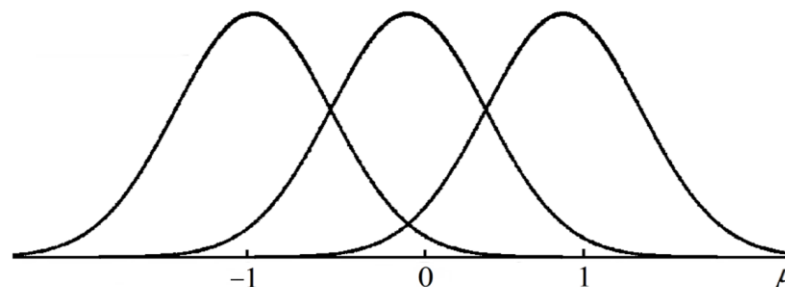
☑ 正态分布的特征

➤ 正态分布是单峰分布，以 $X=\mu$ 为中心左右完全对称，正态曲线与X轴不相交

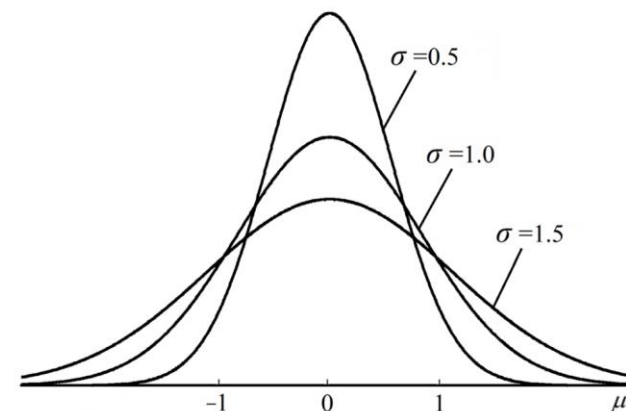
➤ 正态分布有两个参数

μ 是**位置参数**，决定着正态曲线在X轴上的位置

σ 是**形状参数**，决定着正态曲线的分布形状（胖瘦）



正态分布位置参数变化示意图 ($\sigma=1$)



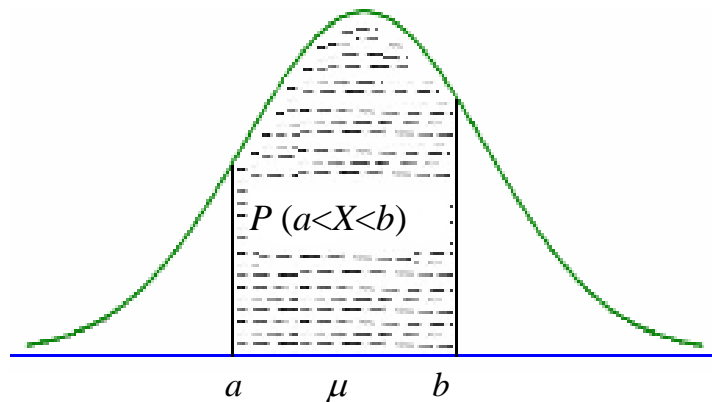
正态分布形状参数变化示意图 ($\mu=0$)

曲线下面积的分布规律

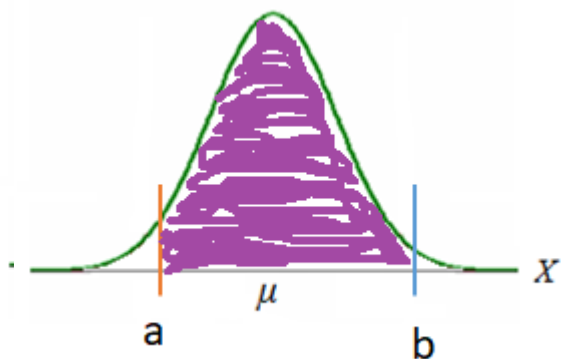
➤ 曲线下面积即为概率

如果对于任何实数 $a < b$, 随机变量 X 满足:

$$P(a < X \leq b) \approx \int_a^b f_{\mu, \sigma}(x) dx = F(b) - F(a)$$



服从正态分布的随机变量, 该随机变量在**某一区间**上的**概率**与**该区间的曲线下面积**与相等。



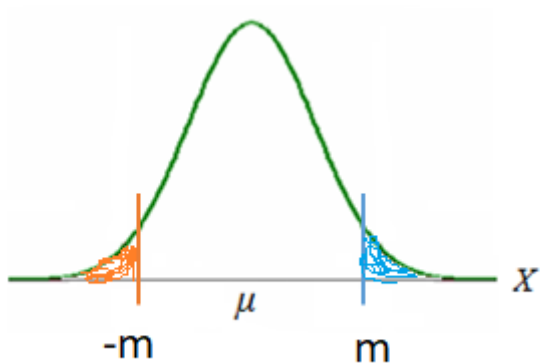
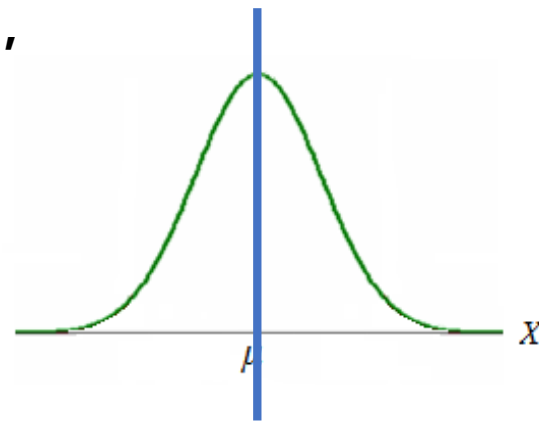
$P(X \leq a)$ = 橘色线左边的曲线下面积: $F(a)$

$P(X \leq b)$ = 蓝色线左边的曲线下面积: $F(b)$

$P(a \leq X \leq b)$ = 橘色线与蓝色线之间的曲线下面积: $F(b) - F(a)$

曲线下面积的分布规律

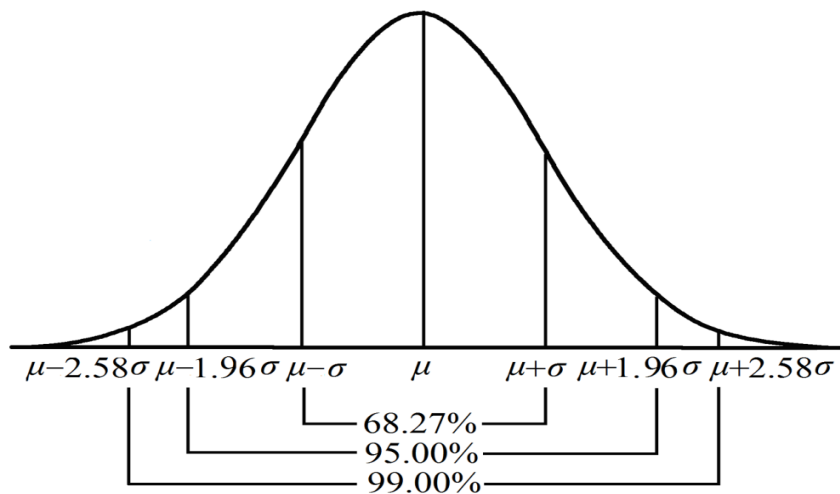
- 曲线下总面积为1或100%，以 μ 为中心左右两侧面积各占50%，
- 正态分布是单峰分布，以 $X=\mu$ 为中心左右完全对称



$$P(X \leq -m) = P(X \geq m)$$

曲线下面积的分布规律

- 所有正态曲线:
- ◆ 区间 $\mu \pm \sigma$ 范围内的面积约为**68.27%**,
- ◆ 区间 $\mu \pm 1.96\sigma$ 范围内的面积约为**95%**,
- ◆ 区间 $\mu \pm 2.58\sigma$ 范围内的面积约为**99%**。



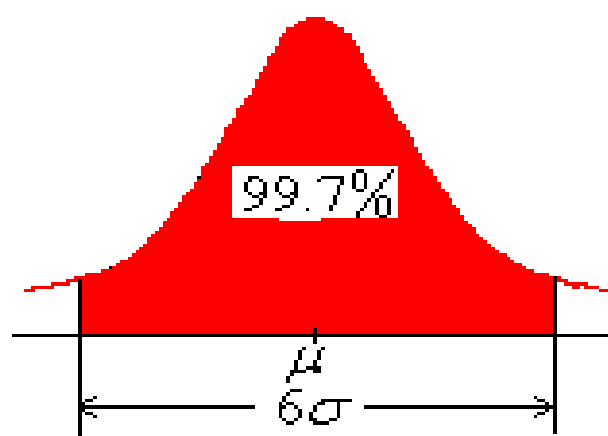
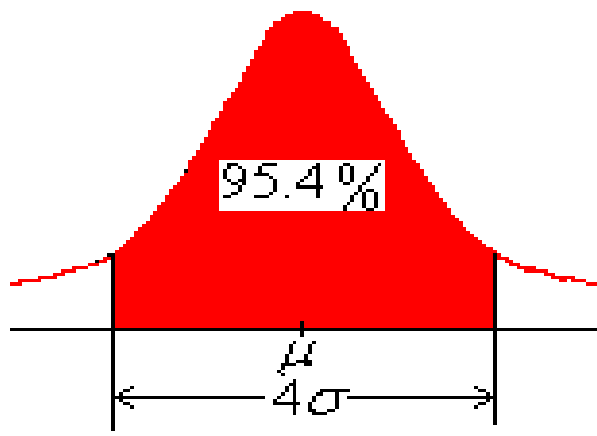
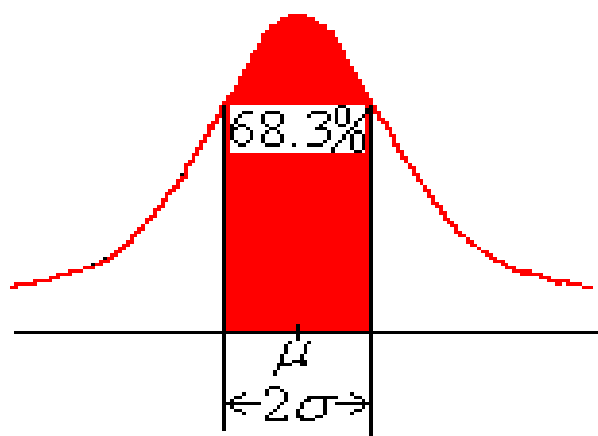
正态曲线下面积的分布规律示意图

几个特殊区间的概率:

区 间	取值概率
$(\mu - \sigma, \mu + \sigma)$	68.3%
$(\mu - 2\sigma, \mu + 2\sigma)$	95.4%
$(\mu - 3\sigma, \mu + 3\sigma)$	99.7%

正态总体几乎总取值于 $(\mu - 3\sigma, \mu + 3\sigma)$ 区间内,而在此区间以外取值的概率只有0.26%,通常认为这种情况在一次试验中几乎不可能发生.

在实际应用中,通常认为服从于正态分布 $N(\mu, \sigma^2)$ 的随机变量只取 $(\mu - 3\sigma, \mu + 3\sigma)$ 之间的值,并称为**3 σ 原则**.



三、标准正态分布

➤ 标准正态分布概念:

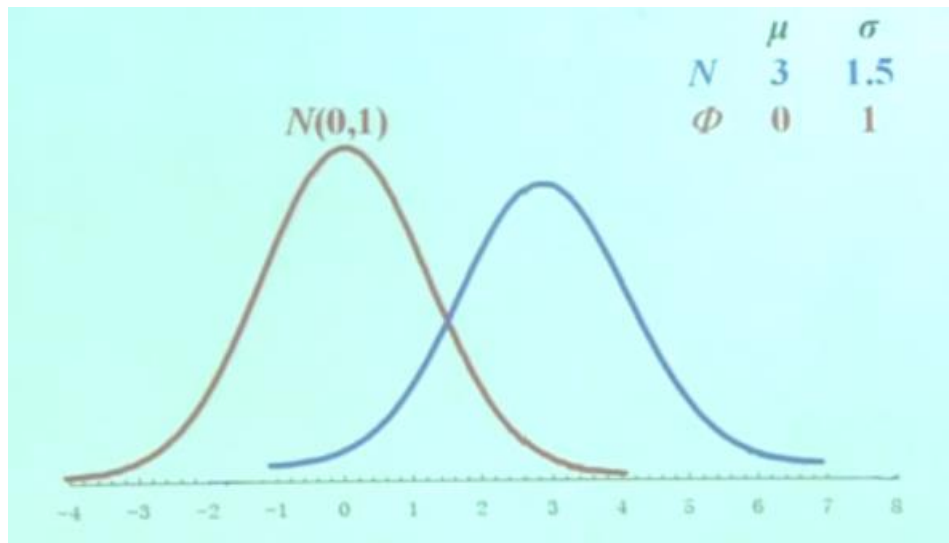
$\mu=0$ 、 $\sigma=1$ 的**正态分布**即为标准正态分布,记为 $N(0,1)$, 即 z 分布。

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

z 变换

- 对任意一个服从正态分布 $X \sim N(\mu, \sigma^2)$ 的随机变量 X ，经过如下变换，都可以转换为标准正态分布：



$$z = \frac{X - \mu}{\sigma}$$

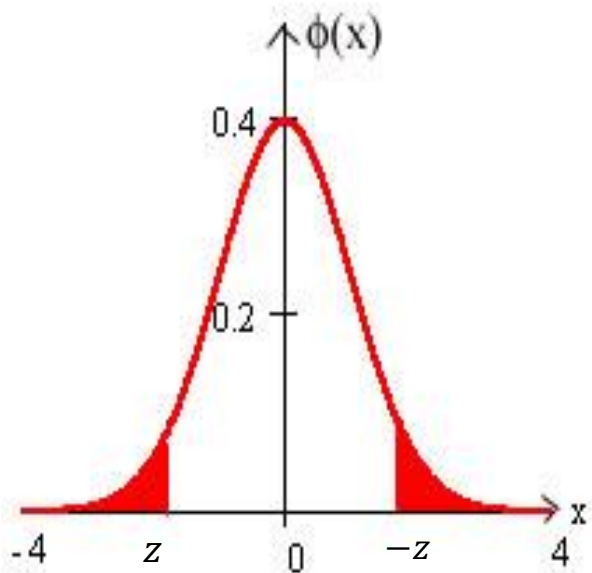
- z 变换在 μ 和 σ 未知时，可以利用样本均数 \bar{X} 和标准差 S 计算

$$z = \frac{X - \bar{X}}{S}$$

经 z 变换后，可把求解任意一个正态分布曲线下面积的问题，转化成标准正态分布曲线下面积的问题。具体值可通过查附表1得到。

正态分布表

书末附表1是标准正态分布函数数值表，有了它，可以解决一般正态分布的概率计算查表。



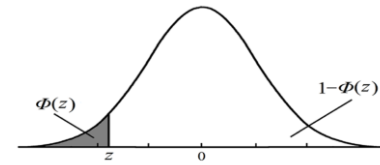
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

附表1中给的是 $z < 0$ 时, $\Phi(x)$ 的值.

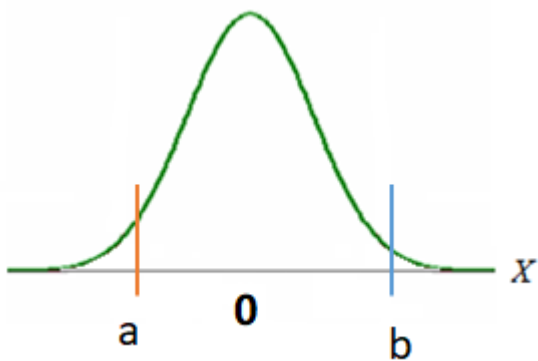
当 $-z > 0$ 时

$$\Phi(-z) = 1 - \Phi(z)$$

附表1 标准正态分布曲线下左侧尾部面积



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455



假设左侧的曲线是服从标准正态分布的 $X \sim N(0,1)$

$P(X \leq a)$, 由图可知 $a < 0$, 可通过查附表1求得左侧尾部面积, 即 $X \leq a$ 的概率值。

$P(X \leq b)$, 由图可知 $b > 0$, 可通过转换为 $1 - P(-b)$ 求得, 其中 $p(-b)$ 可通过查附表1, 求得左侧尾部面积, 即 $X \leq b$ 的概率值。

$P(a \leq X \leq b)$, 转化为 $p(b) - p(a) = 1 - p(-b) - p(a)$

- $\mu \pm 1.96\sigma$ 范围内曲线下的面积占总面积的95%
- $\mu \pm 2.58\sigma$ 范围内曲线下的面积占总面积的99%

例题

例3-1 若 $X \sim N(\mu, \sigma^2)$, 试计算X取值在区间 $\mu \pm 1.96\sigma$ 上的概率

$$P(X_1 < X < X_2) = P(\mu - 1.96\sigma < X < \mu + 1.96\sigma)$$

$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{(\mu - 1.96\sigma) - \mu}{\sigma} = -1.96$$

$$z_2 = \frac{X_2 - \mu}{\sigma} = \frac{(\mu + 1.96\sigma) - \mu}{\sigma} = 1.96$$

$$\begin{aligned} P(-1.96 < z < 1.96) &= \Phi(1.96) - \Phi(-1.96) = (1 - \Phi(-1.96)) - \Phi(-1.96) \\ &= 1 - 2\Phi(-1.96) = 1 - 2 \times 0.025 = 0.95 \end{aligned}$$

问题

➤ 试给出 $\mu \pm 2.58\sigma$ 范围内曲线下的面积值?

$$P(X_1 < X < X_2) = P(\mu - 2.58\sigma < X < \mu + 2.58\sigma)$$

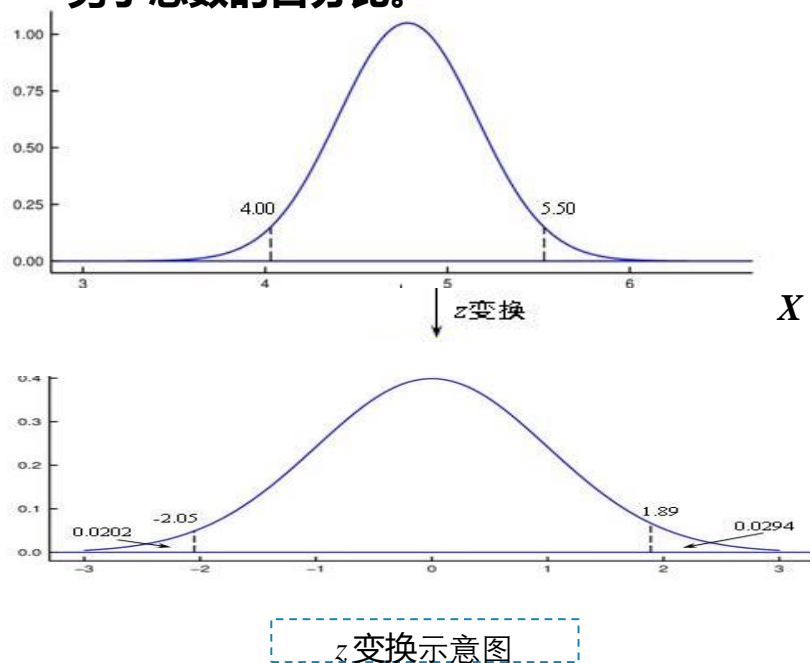
$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{(\mu - 2.58\sigma) - \mu}{\sigma} = -2.58$$

$$z_2 = \frac{X_2 - \mu}{\sigma} = \frac{(\mu + 2.58\sigma) - \mu}{\sigma} = 2.58$$

$$\begin{aligned} P(-2.58 < z < 2.58) &= \phi(2.58) - \phi(-2.58) = 1 - \phi(-2.58) - \phi(-2.58) \\ &= 1 - 2\phi(-2.58) = 1 - 2 \times 0.0049 = 0.99 \end{aligned}$$

例题

例3-2 已知某地140名正常成年男子红细胞计数近似服从正态分布, $\bar{X} = 4.78 \times 10^{12}/L$, $S = 0.38 \times 10^{12}/L$, 试估计: 1.该地正常成年男子红细胞计数在 $4.0 \times 10^{12}/L$ 以下者占该地正常成年男子总数的百分比。2.该地正常成年男子红细胞计数在 $4.0 \times 10^{12} \sim 5.5 \times 10^{12}/L$ 者占该地正常成年男子总数的百分比。



题1 将 $X=4$ 代入式 (3-8), 得

$$z = \frac{X - \bar{X}}{S} = \frac{4.0 - 4.78}{0.38} = -2.05 \quad \text{查表求得}$$

$\phi(-2.05) = 0.0202$, 该地正常成年男子红细胞计数在 $4.0 \times 10^{12}/L$ 以下者占该地正常成年男子总数的所占比例为2.02%。

题2

$$P(4.0 < X < 5.5) = P\left(\frac{4.0 - 4.78}{0.38} < \frac{X - \bar{X}}{S} < \frac{5.5 - 4.78}{0.38}\right)$$

z

$$\downarrow \\ P(-2.05 < z < 1.89)$$

$$(1 - \Phi(-1.89)) - \Phi(-2.05) = (1 - 0.0294) - 0.0202 = 0.9504$$

该地正常成年男子红细胞计数在 $4.0 \times 10^{12} \sim 5.5 \times 10^{12}/L$ 者占该地正常成年男子总数的95%。

问题

正态分布由两个参数决定，一个为**总体均数**，一个**总体方差**。

位置参数为： μ ，决定曲线在X轴的位置

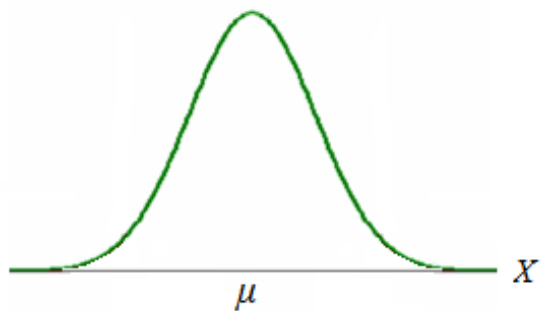
形状参数为： σ ，决定曲线的胖瘦

重要内容回顾

☑ 正态分布

- 连续型随机变量 X 服从正态分布，记为 $X \sim N(\mu, \sigma^2)$ ，正态分布有两个参数：

μ 为X的**总体均数**、 σ^2 为X的**总体方差**。



正态分布示意图

☑ 标准正态分布

- 对任意一个服从正态分布 $X \sim N(\mu, \sigma^2)$ 的随机变量 X , 经过如下变换, 都可以转换为标准正态分布:

$$z = \frac{X - \mu}{\sigma}$$

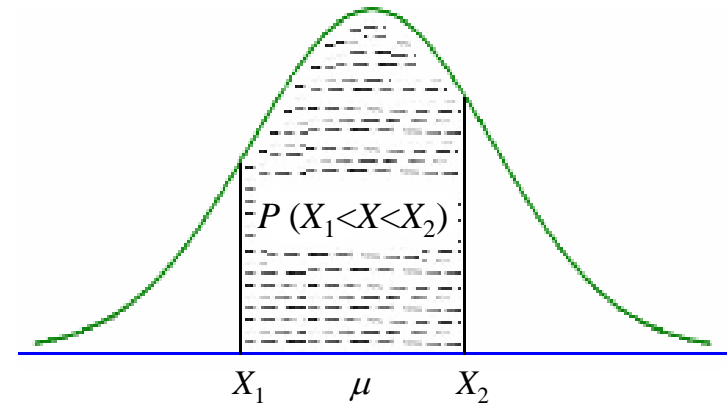
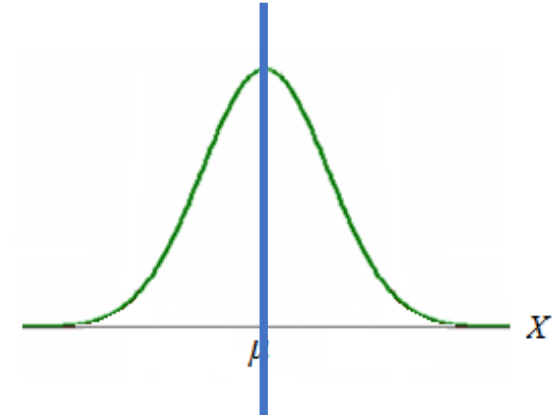
- z 变换在 μ 和 σ 未知时, 可以利用样本均数 \bar{X} 和标准差 S 计算

$$z = \frac{X - \bar{X}}{S}$$

☑ 曲线下面积的分布规律

- 曲线下面积即为概率
- 曲线下总面积为1或100%，以 μ 为中心左右两侧面积各占50%

服从正态分布的随机变量在**某一区间**上的曲线下面积与该随机变量在**同一区间**上的**概率**相等。





第二节

医学参考值范围



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE

一、医学参考值范围的概念

- 医学参考值范围 (medical reference range) : 指“正常人”的解剖、生理、生化指标等数据**大多数个体值的波动范围**。
- 由于存在生物个体变异, 虽然都是正常人, 但指标测量值也会有所不同, 即使是同一个体的数据, 也会随环境、时间等变化而变化, 因此不能以某一个测量数据作为标准, 必须确定一个波动范围。

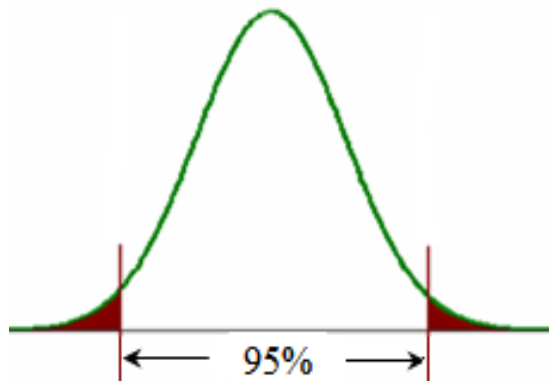
检验报告单 请妥善保管下次就诊携带 第二军医大学附属长海医院 检验编号:S00245

姓名:	编号	项目	结果	参考值	编号	项目	结果	参考值
性别: 女 年龄: 30	1	WBC	11.82 H	4.0-10.6x10 ⁹ /L	15	MCV	86.20	80-100 fL
病人ID:	2	LYM#	3.78	0.8-4.0x10 ⁹ /L	16	MCH	29.4	26-34 pg
科室: 急救科	3	MO#	0.83 H	0.12-0.6x10 ⁹ /L	17	MCHC	340	320-360 g/l
病区: 床号:	4	GRAN#	7.12 H	2-7x10 ⁹ /L	18	RDW-CV	12.1	11.9-14.5%
标本种类血	5	EO#	0.08	0.05-0.6x10 ⁹ /L	19	PLT	224	100-300 x10 ⁹ /L
标本说明:	6	BASO#	0.01	<0.1x10 ⁹ /L	20	MPV	8.70L	9.4-12.6 fL
送检日期:	7	LYM%	32.0	20-40 %	21	PCT	0.19	0.16-0.35%
类别:	8	MO%	7.0	3-8 %	22	PDW	8.90L	9.8-16.1 fL
临床诊断:	9	EO%	0.7	0.5-5 %	23	P-LCR	15.8L	19.2-47 %
	10	BASO%	0.1	<1 %				
	11	GRAN%	60.2	50-70 %				
	12	RBC	4.94	4-5.5x10 ¹² /L				
	13	HGB	145.0	120-160 g/L				
	14	HCT	0.426	0.36-0.53				

此结果仅对本
检验标本负责
注: H-偏高, L-偏低 送检 收到 报告

确定医学参考值范围时，选择数据样本需注意：

- 1、选择足够数量的正常人作为调查对象。
 - 2、样本含量足够大
- 其确切含义是，将取得样本的所有个体观察值，用统计学方法建立百分位数界限，由此得到个体观察值的波动区间。通常情况使用的是95%参考值范围。



95%参考值范围示意图

➤ 确定医学参考值范围的意义：

- ◆ 一是基于临床实践，从个体角度，作为临床上判定正常与异常的参考标准，即用于划分界限或分类；如成年人的白细胞计数。
- ◆ 二是基于预防医学实践，从人群角度，可用来评价儿童的发育水平，如制订不同年龄、性别儿童某项发育指标的等级标准。

二、制订医学参考值范围的注意事项

1. 确定同质的参照总体

- 一般选择“正常人”，主要是排除了对研究指标有影响的疾病或有关因素的同质人群。如确定血红蛋白的参考值范围，“正常人”就不能包括贫血的人，其他如身体残疾的，而不存在贫血，是可以纳入“正常人”范围的。

2. 选择足够例数的参照样本

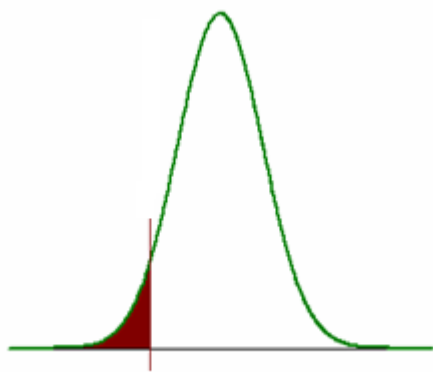
- 通常情况下，确定参考值范围需要大样本，如果例数过少，确定的参考值范围往往不够准确。

3. 控制检测误差

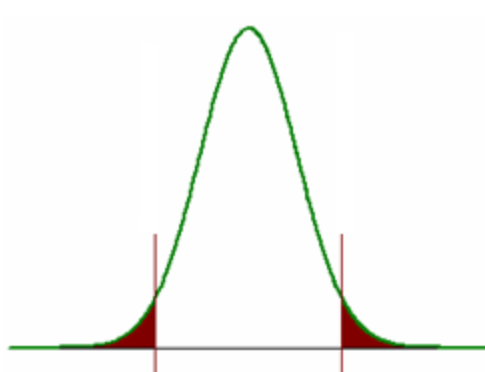
- 为保证原始数据可靠，检测过程中要严格控制随机误差，避免系统误差。

4. 选择单、双侧界值

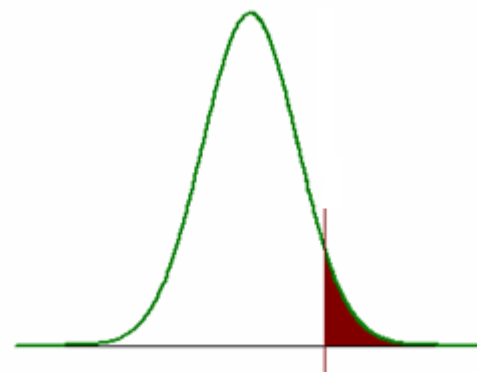
- 依据专业知识确定，研究指标无论过高或过低均属异常，采用双侧界值；有些指标仅过大或者过小为异常，采用单侧界值。



肺活量参考值范围



白细胞数参考值范围



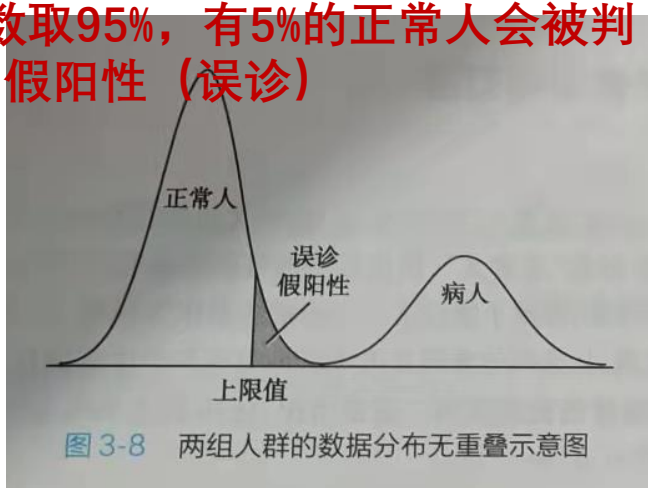
血铅参考值范围

5. 选择适当的百分数范围

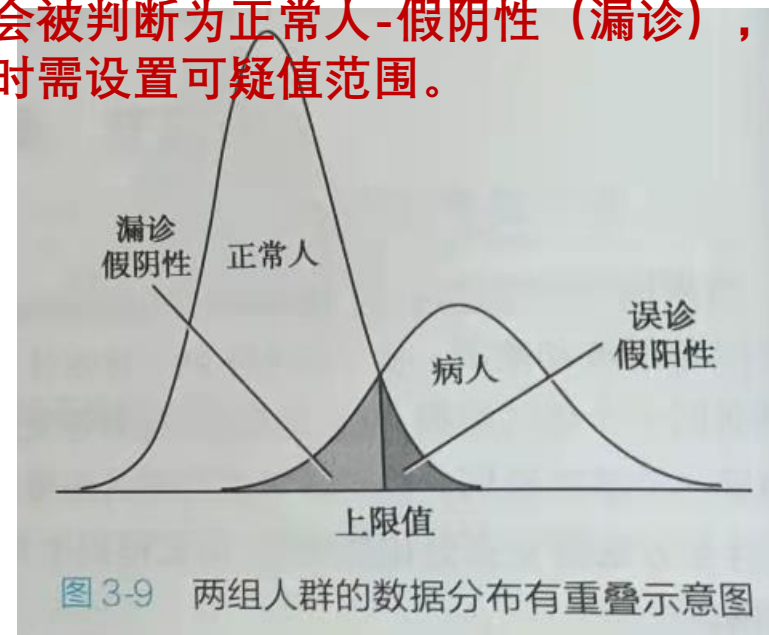
- ▶ 参考值范围是指参照总体中绝大多数“正常人”测量值的所在范围。
“绝大多数”是多少，应结合专业知识，数据分布特征、研究目的确定百分数范围。

根据数据特征确定百分数范围

由于两组人群数据基本无交叠，若正常人百分位数取95%，有5%的正常人会被判断为病人-假阳性（误诊）



由于两组人群数据交叠较多，存在正常人会被判断为病人-假阳性（误诊），病人会被判断为正常人-假阴性（漏诊），此时需设置可疑值范围。



- ✓ 如研究目的是确诊病人→确实有病的才包含进来，减少正常人被包含进来的风险，需要减少假阳性率（误诊率），参考值范围的百分数范围要取 **大一些**。
- ✓ 如研究的目的是为初筛病人→尽量将病人都包含进来，则需要减少假阴性率（漏诊率），参考值百分数范围可取 **小一些**。
- 百分数范围可以取80%、90%、95%或99%等，其中以95%最为常见。

6. 选择计算参考值范围的方法

- 根据资料的分布类型，样本含量的多少和研究目的等，选用适当的方法确定参考值范围：
 - ◆ 近似服从正态分布或能转换为正态分布的资料，可选用正态近似法；
 - ◆ 不服从正态分布的资料，可选用百分位数法等进行计算。

三、医学参考值范围的计算方法

☑ 最基本的两种方法

- **百分位数法**适合于**任何分布类型的资料**，在实际中最为常用。由于参考值范围所涉及的常常是波动较大的两端数据，使用百分位数法必须要有**较大的样本含量**，否则结果不稳定。
- **正态分布法**要求**资料服从或近似服从正态分布**，优点是结果比较稳定，在样本含量不是很大的情况下仍然能够进行处理；若偏态分布经变量变换能转换为正态分布或近似正态分布，仍可用正态分布法。

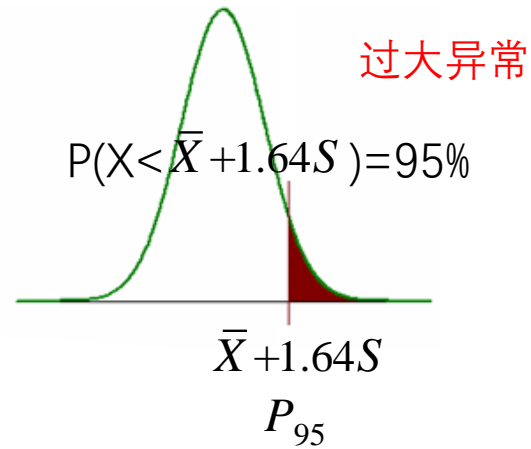
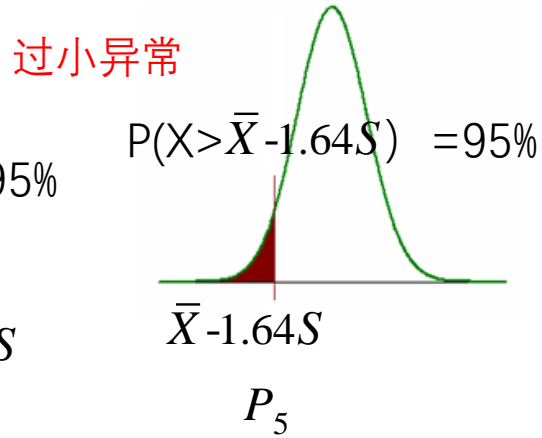
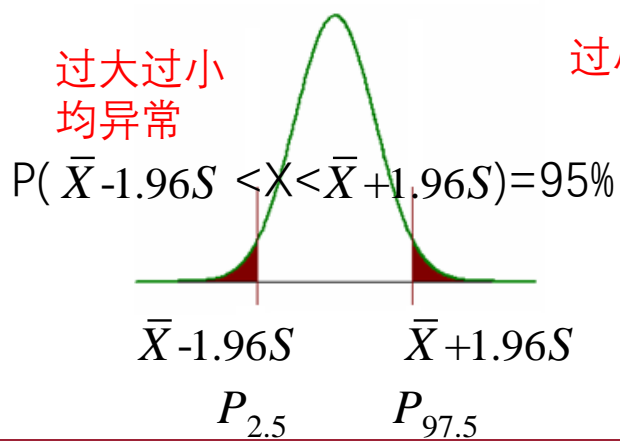
医学参考值范围的正态分布法和百分位数法的计算

概率 (%)	正态分布法			百分位数法		
	双侧	单侧		双侧	单侧	
		下限	上限		下限	上限
90	$\bar{X} \pm 1.64S$	$\bar{X} - 1.28S$	$\bar{X} + 1.28S$	$P_5 \sim P_{95}$	P_{10}	P_{90}
95	$\bar{X} \pm 1.96S$	$\bar{X} - 1.64S$	$\bar{X} + 1.64S$	$P_{2.5} \sim P_{97.5}$	P_5	P_{95}
99	$\bar{X} \pm 2.58S$	$\bar{X} - 2.33S$	$\bar{X} + 2.33S$	$P_{0.5} \sim P_{99.5}$	P_1	P_{99}

双侧参考值范围：95%

单侧参考值范围（下限）：95%

单侧参考值范围（上限）：95%



☑ 计算医学参考值范围的正态分布法

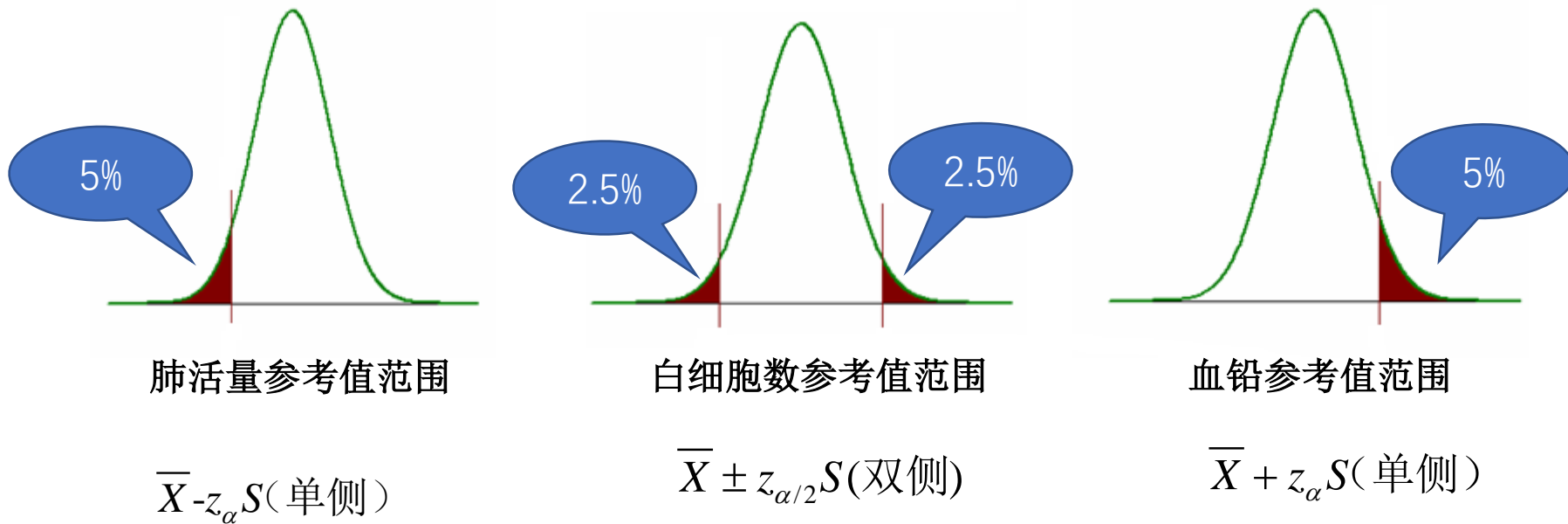
计算步骤:

- 对资料进行正态性检验: P200页
- 若资料服从正态分布, 计算资料的均数 \bar{X} 和标准差 S
- 计算单侧或双侧参考值范围 $\bar{X} \pm z_{\alpha/2} S$ (双侧)
 $\bar{X} + z_{\alpha} S$ 或 $\bar{X} - z_{\alpha} S$ (单侧)

其中, α 表示犯错误的概率。

☑ 关于 α 的确定

确定医学参考值范围为95%， $\alpha=0.05$ ， $\alpha/2=0.025$



例题

- **例3-3** 已知某地140名正常成年男子红细胞计数近似服从正态分布, $\bar{X} = 4.78 \times 10^{12}/L$, $S = 0.38 \times 10^{12}/L$, 试估计该地正常成年男子红细胞计数95%参考值范围。

近似正态分布资料可按正态分布法处理, 因红细胞计数值过大或过小均为异常, 故应估计**双侧**95%参考值范围:

$$\bar{X} \pm z_{0.05/2} S = 4.78 \pm 1.96 \times 0.38 = (4.04, 5.52)$$

即该地正常成年男子红细胞计数的95%参考值范围为 $4.04 \times 10^{12} \sim 5.52 \times 10^{12}/L$ 。

例题

- **例3-4** 某年某地测得 100 名正常成年人的血铅含量值 ($\mu\text{g}/\text{dl}$)，试确定该地正常成年人血铅含量的95%参考值范围。

4	4	5	5	6	6	7	7	7	7	7	8	8	8	8	8	8	8	9	9
10	10	10	10	10	10	10	10	11	11	11	12	13	13	13	13	13	13	13	13
13	13	14	14	14	15	15	16	16	16	16	16	16	16	16	17	17	17	17	17
18	18	18	18	19	20	20	20	20	21	21	22	22	22	23	24	24	25	25	26
26	26	27	27	28	28	29	30	30	31	31	32	32	32	33	35	41	44	50	51

血铅含量的原始数值表现为正偏态分布，通过对数变换且经正态性检验可知对数值服从正态分布，故根据对数值的均数和标准差采用正态分布法求95%参考值范围。

表 3-2 某年某地 100 名正常成年人血铅含量 ($\mu\text{g}/\text{dl}$) 原始数值与对数值频数表

原始数值组段 (1)	频数 (2)	累计频数 (3)	对数值组段 (4)	频数 (5)	累计频数 (6)
4 ~	11	11	0.6 ~	4	4
8 ~	20	31	0.7 ~	2	6
12 ~	16	47	0.8 ~	5	11
16 ~	18	65	0.9 ~	9	20
20 ~	10	75	1.0 ~	12	32
24 ~	9	84	1.1 ~	15	47
28 ~	7	91	1.2 ~	18	65
32 ~	5	96	1.3 ~	14	79
36 ~	0	96	1.4 ~	12	91
40 ~	1	97	1.5 ~	5	96
44 ~	1	98	1.6 ~	3	99
48 ~ 52	2	100	1.7 ~ 1.8	1	100
合计	100	—	合计	100	—

- 根据以上频数表，可以看到，原始数据呈正偏态分布
- 经过对数变换后，数据呈近似的正态分布。

➤ 依据表3-2，设 X 为对数组段的组中值， $n=100$ ， $\sum fX = 120$ ， $\sum fX^2 = 149.73$ ，则对数值的均数和标准差为：

$$\bar{X} = \frac{\sum fX}{n} = \frac{120}{100} = 1.2 \text{ (}\mu\text{g/dl)}$$

$$S = \sqrt{\frac{\sum fX^2 - (\sum fX)^2 / n}{n-1}} = \sqrt{\frac{149.73 - 120^2 / 100}{100-1}}$$

$$S = 0.2406 \text{ (}\mu\text{g/dl)}$$

因为血铅含量仅过大异常，参考值范围为单侧，单侧95%的上限值

$$\lg^{-1}(\bar{X} + 1.64S) = \lg^{-1}(1.2 + 1.64 \times 0.2406) = 39.3173$$

即该地正常成年人血铅含量95%参考值范围为小于39.3173 $\mu\text{g}/\text{dl}$ 。

例题

- **例3-5** 依据表2-4某地630名50~60岁正常女性血清甘油三酯含量 (mmol/L) 的资料, 估计其血清甘油三酯含量的**单侧95%**参考值范围, 为该地50~60岁女性高血脂诊断与治疗提供参考依据。

某地630名正常女性血清甘油三酯含量(mmol/L)

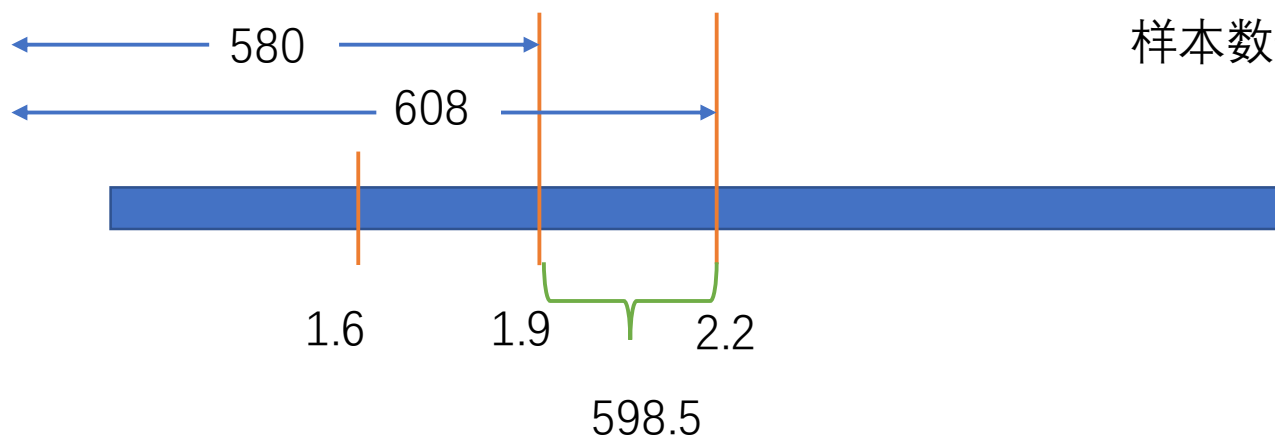
甘油三酯(mmol/L) (1)	频数 (2)	累积频数 (3)	累积频率(%) (4)
0.10 ~	27	27	4.29
0.40 ~	169	196	31.11
0.70 ~	167	363	57.62
1.00 ~	94	457	72.54
1.30 ~	81	538	85.40
1.60 ~	42	580	92.06
1.90 ~	28	608	96.51
2.20 ~	14	622	98.73
2.50 ~	4	626	99.37
2.80 ~	3	629	99.84
3.10 ~	1	630	100.00
合计	630	-	-

资料显现出血清甘油三酯含量数值偏小的人数较多，呈正偏态分布，故选用百分位数法计算参考值范围，依据专业知识，应计算单侧95%界值 P_{95} 。

$$P_{95}: 630 \times 0.95 = 598.5$$

$$P_{95} = 1.90 + (630 \times 95\% - 580) \times \frac{0.3}{(608 - 580)} = 2.098 \text{ mmol/L}$$

↓
样本数*密度



即该地50~60岁正常女性血清甘油三酯含量的单侧95%参考值范围为小于2.098 mmol/L。

正态性判定

- 判定样本资料是否服从正态分布

方法:

- 频数表、直方图
- P200, 如何进行正态性检验

本章小结

1. 正态分布 $X \sim N(\mu, \sigma^2)$ 有两个参数，总体均数 μ 是位置参数，总体标准差 σ 是形状参数；正态曲线下的面积即为概率，利用其面积分布规律可估计频数分布和确定医学参考值范围。
2. $\mu=0$ 、 $\sigma=1$ 的正态分布称作标准正态分布，即 z 分布；随机变量 X 的任意正态分布都可经 z 变换转化成标准正态分布。



本章小结

3. 医学参考值范围指同质总体中某医学指标大多数个体值的波动范围，若资料服从正态分布或转换值服从正态分布，可用正态近似法计算；若资料不服从正态分布或未知分布类型，可用百分位数法计算。



请同学讲解第三章习题二-1、二-2

课后习题二-1

1、50kg以上占的百分比。

$$z = \frac{X - \bar{X}}{\sigma} = \frac{50 - 36.3}{6.19} = 2.213$$

$$\phi(2.213) = 1 - \phi(-2.213) = 0.0136 = 1.36\%$$

附表 1 得 $\Phi(-2.21) = 0.0136$ ，即理论上该地 12 岁健康男童体重在 50kg 以上者占该地 12 岁健康男童总数的 1.36%。

2、30kg-40kg占的百分比。

$$z = \frac{X - \bar{X}}{\sigma} = \frac{40 - 36.3}{6.19} = 0.597$$

$$z = \frac{X - \bar{X}}{\sigma} = \frac{30 - 36.3}{6.19} = -1.01$$

$$P(-1.02 < z < 0.6) = 1 - \phi(-0.6) - \phi(-1.02) = 1 - 0.2743 - 0.1539 = 0.5718 = 57.18\%$$

即理论上该地 12 岁健康男童体重在 30~40kg 者占该地 12 岁健康男童总数的 57.18%。

3、80%的集中范围。

$$X = \bar{X} \pm 1.28S = 36.3 \pm 1.28 \times 6.19 = 28.38 \sim 44.22 \text{ 公斤}$$

该地 80% 的 12 岁健康男童体重值集中在区间 $\bar{X} \pm 1.28S$ 内, 即 28.38~44.22kg。

4、95%的参考范围。

$$X = \bar{X} \pm 1.96S = 36.3 \pm 1.96 \times 6.19 = 24.17 \sim 48.43 \text{ 公斤}$$

即该地 12 岁健康男童体重的 95% 参考值范围为 24.17~48.43kg。

课后习题二-2

题中所给资料属于正偏态分布资料，所以宜用百分位数法计算其参考值范围。

又因血铅含量仅过大为异常，故应计算只有上限的单侧范围，即 P_{95} 。

血铅含量	0~<0.2 4	0.24~ <0.48					1.44~<1.68	1.68~<1.9 2	1.92~	2.16~<2. 4	2.4~<2.6 4	2.64~
频数	7	49	45	32	28	13	14	4	4	1	2	1
累计频数	7	56	101	133	161	174	188	192	196	197	199	200

$$1.68 + \frac{1.92 - 1.68}{4} \times (200 \times 0.95 - 188) = 1.8$$

该地正常成年人血铅含量的 95% 参考值范围为小于 $1.80 \mu\text{mol/L}$ 。

第一次作业

例1、在某次数学考试中，考生的成绩 X 服从一个正态分布，即 $X \sim N(90,100)$ 。

(1) 试求考试成绩 X 位于区间 $(70,110)$ 上的概率是多少？

(2) 若这次考试共有2000名考生，试估计考试成绩在 $(80,100)$ 间的考生大约多少人？

作业提交

https://zhangy-lab.cn/teaching.html

Zhang Lab

[About](#) [Member](#) [Publication](#) [Teaching](#) [Contact](#)

医学统计学

第一章 绪论

第二章 定量数据的统计描述

第三章 正态分布与医学参考值范围

作业提交

请注意作业文件的提交格式为“姓名_学号”!

姓名

学号

作业 未选择任何文件

第几次作业

提交

重置

Word文档



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE

谢谢观看