

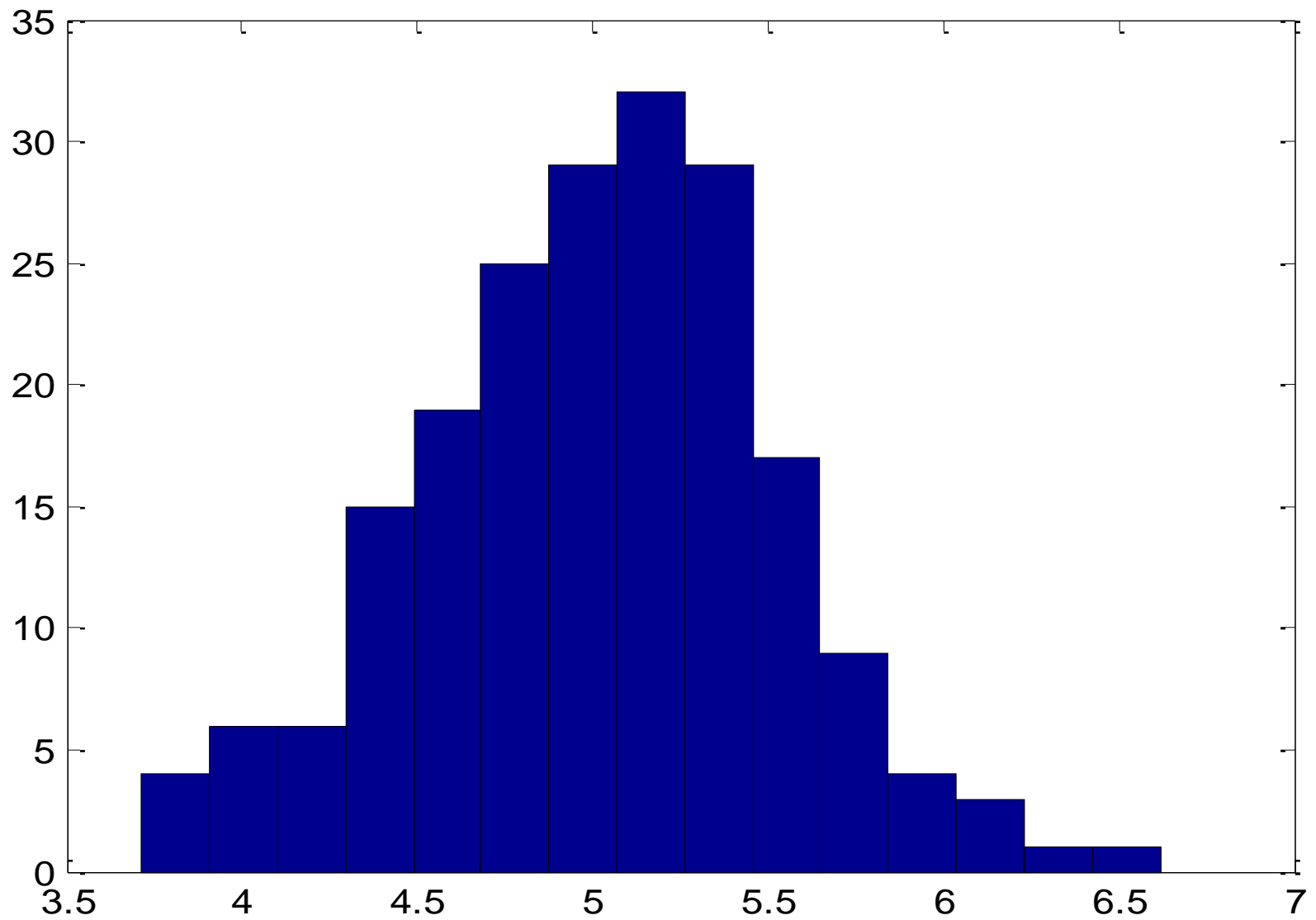


正态分布与参考值范围估计

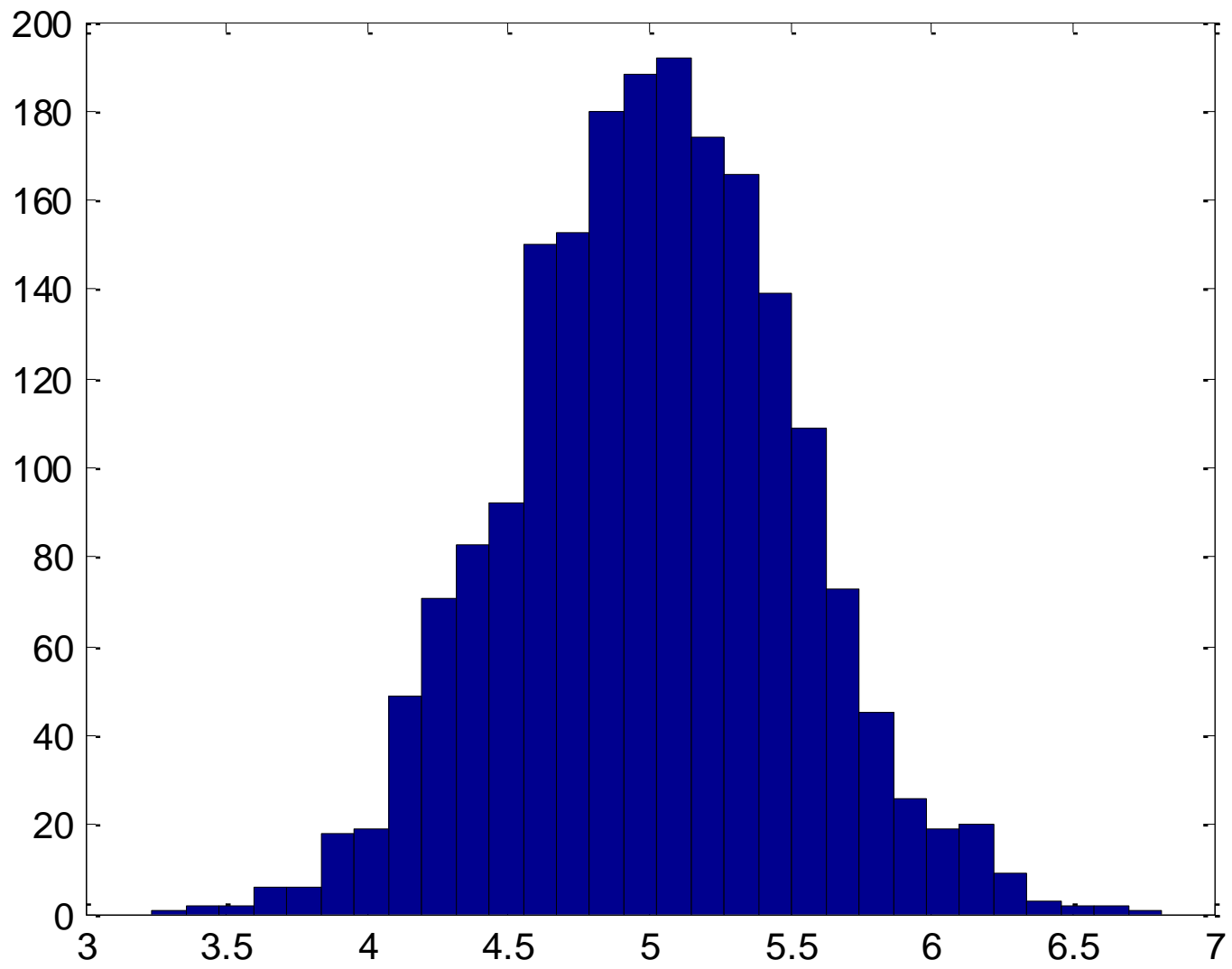
Normal distribution and estimation
of reference range

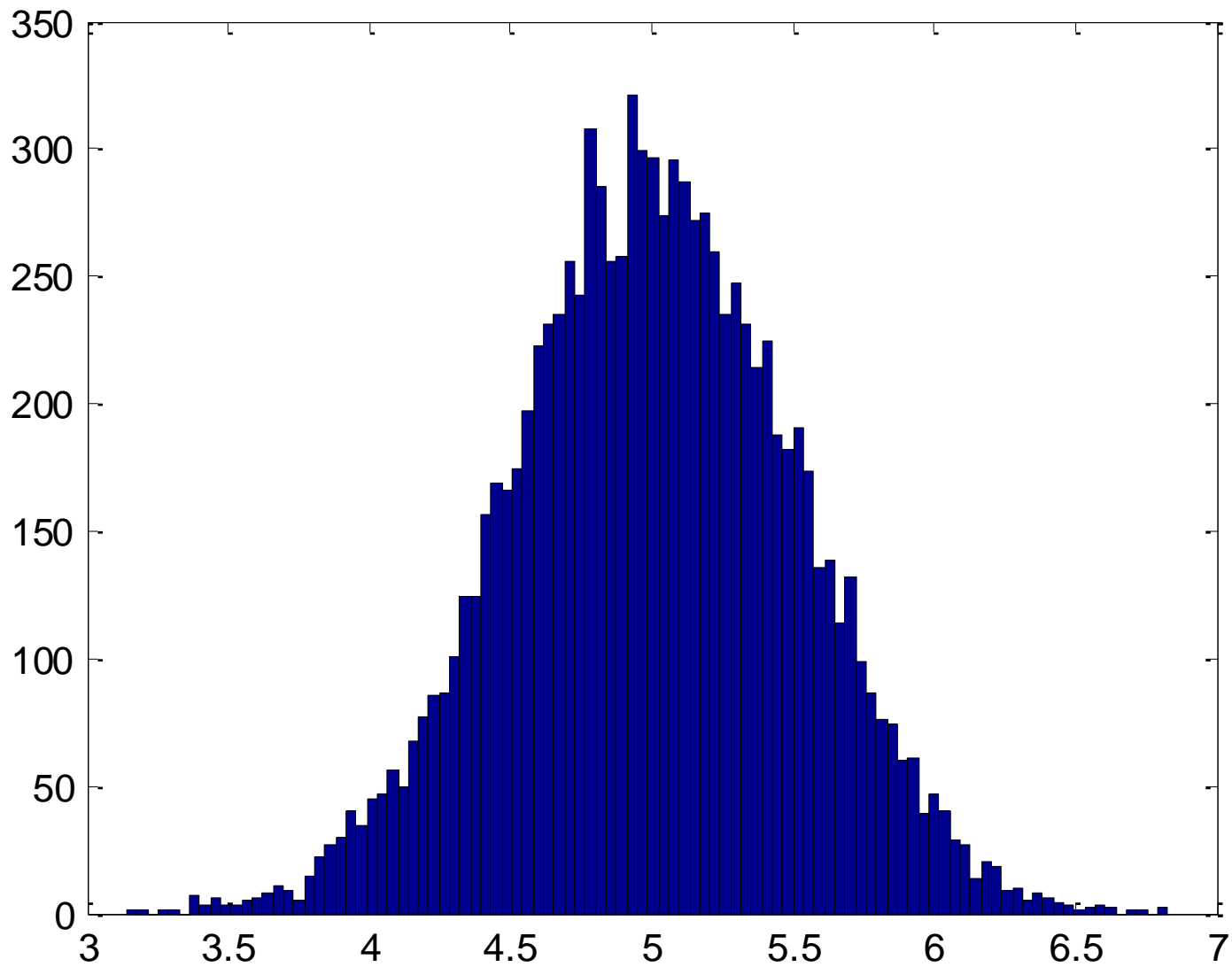


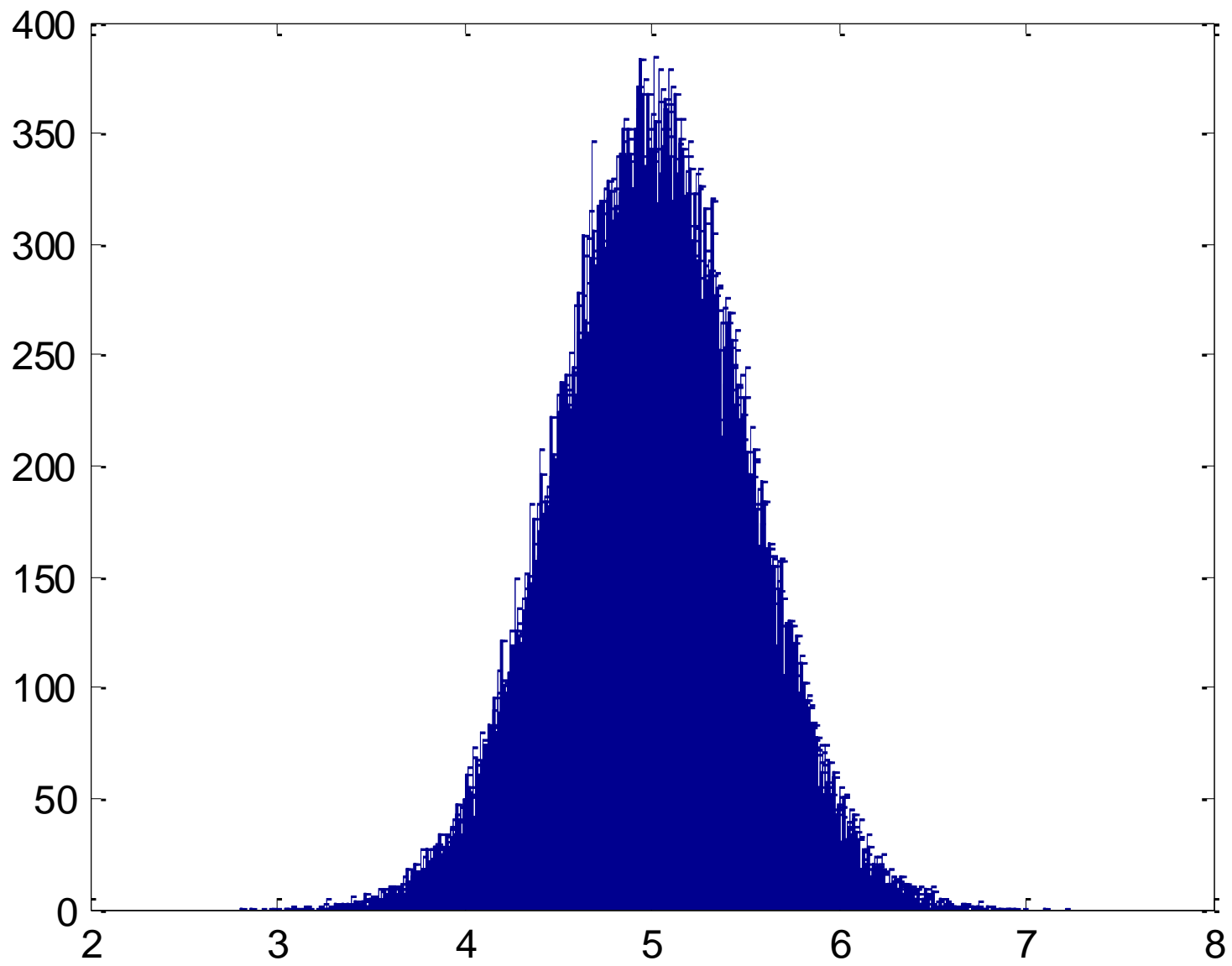
南方医科大学生物统计学系 安胜利

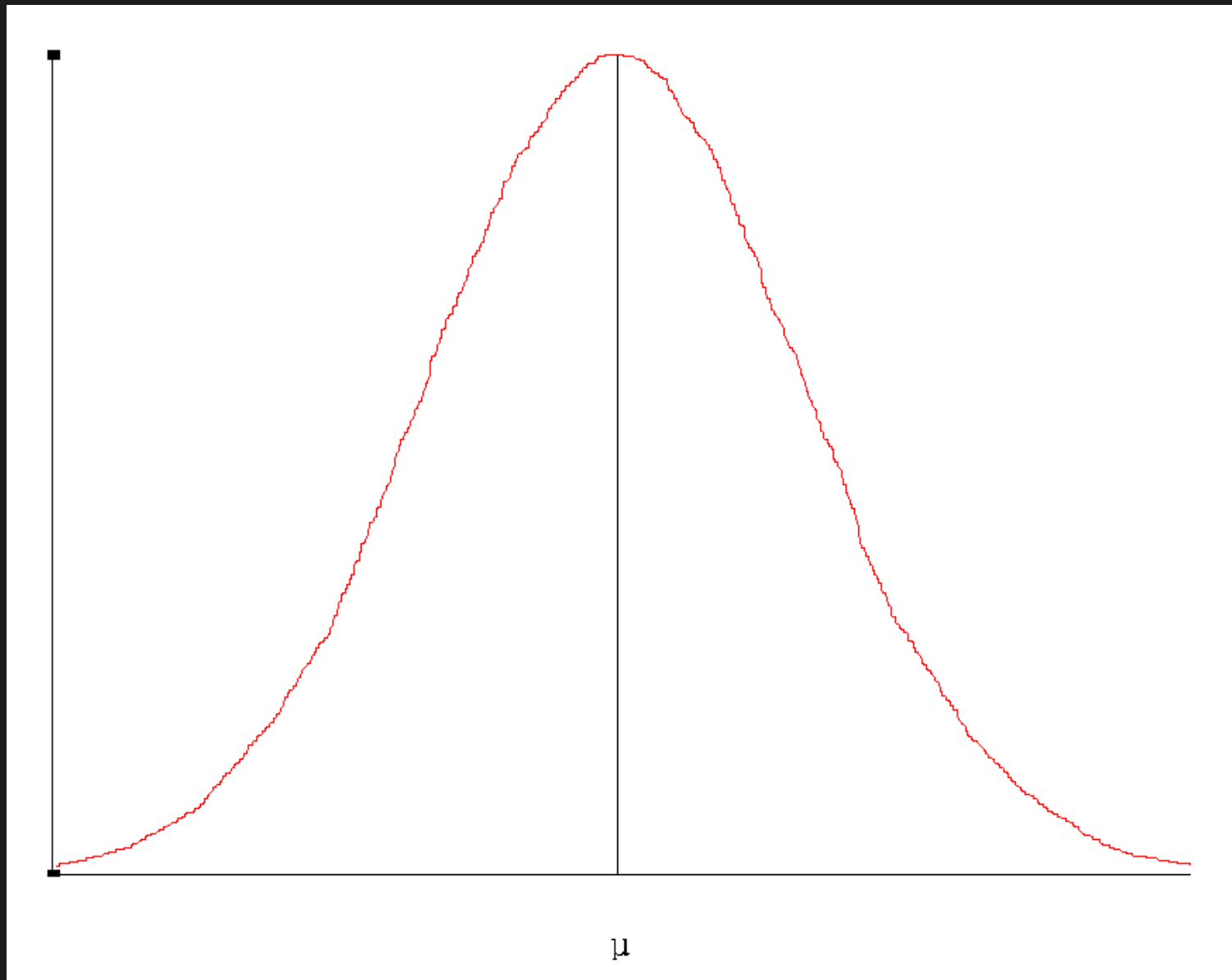


● mean=5.00, sd=0.50









- 分布：x轴是具体的测量值，y轴是相应的频数

正态分布的重要性

- 大量实际资料服从正态分布或经变量变换后服从正态分布。
- 许多分布建立在正态分布的基础之上，或在一定条件下趋于正态分布。
 - 卡方分布，t分布，F分布是在正态分布基础上推导出来。二项分布 $n > 30$, poisson分布在 $\lambda \geq 20$ 时，趋近正态分布
- 多数参数统计方法的前提条件是资料服从正态分布（来自正态分布总体）。

德国数学家A de Moivre 1733年提出

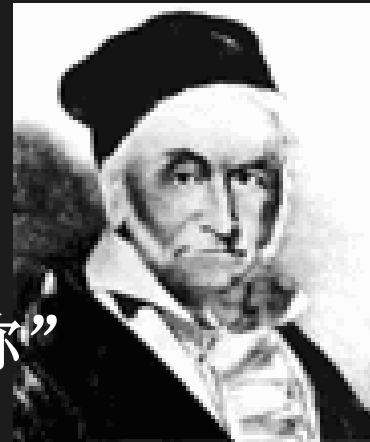
(统计学家 K .Pearson 1924年发现)

德国数学家CF.Gauss 1809年

normal 不是指“正常”、“对称”



Karl Pearson



正态分布的概率密度函数 (probability-density function, pdf)

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

$$y = \varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad -\infty < u < +\infty$$

Standard (or unit) normal distribution

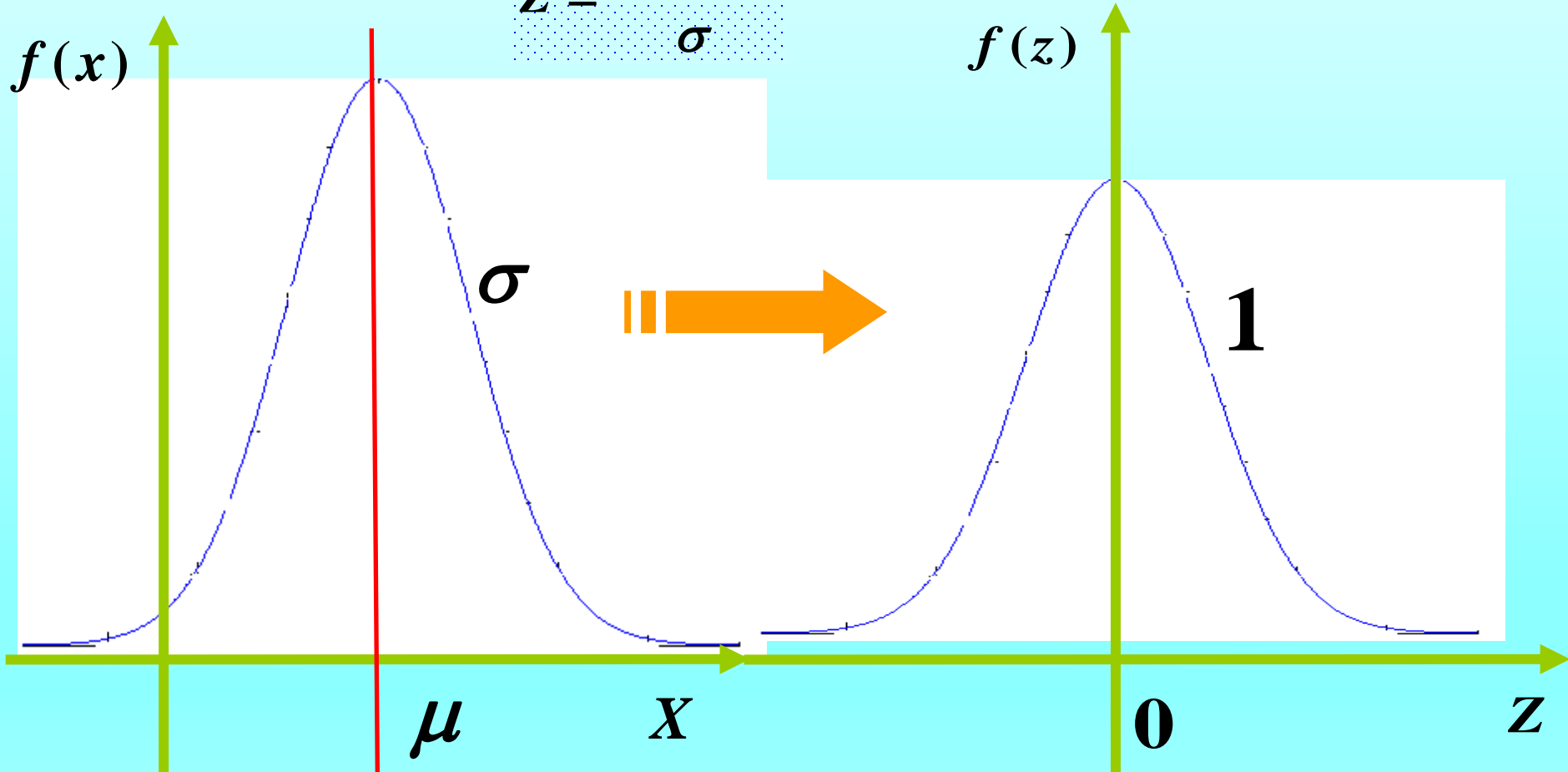
- 任何正态分布变量 X 若做**标准化变换**，即可得标准正态分布变量 u (或 z)

$$u_i = \frac{X_i - \mu}{\sigma}$$

- u 称为
 - 标准正态变量、
 - 标准正态离差、
 - 标准分 (z-score)

正态分布与标准正态分布的转换

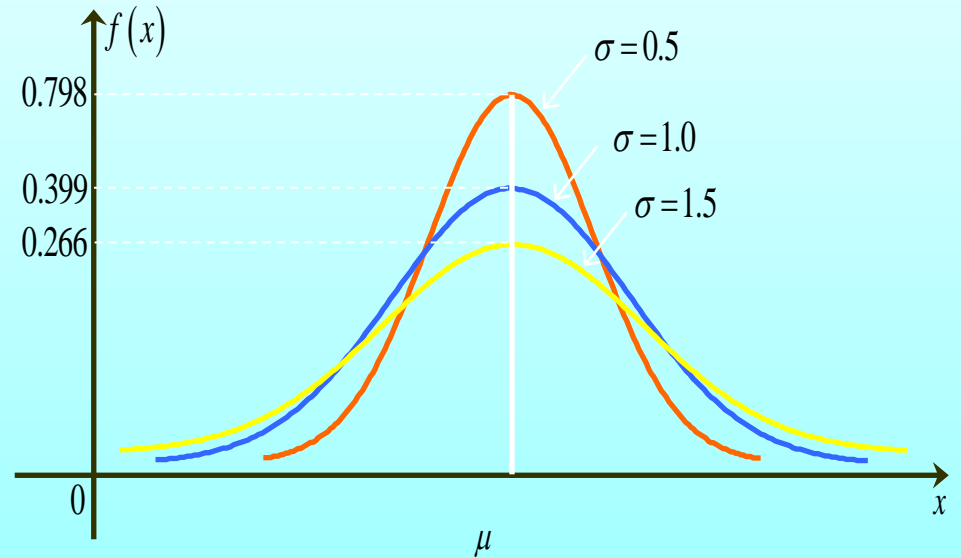
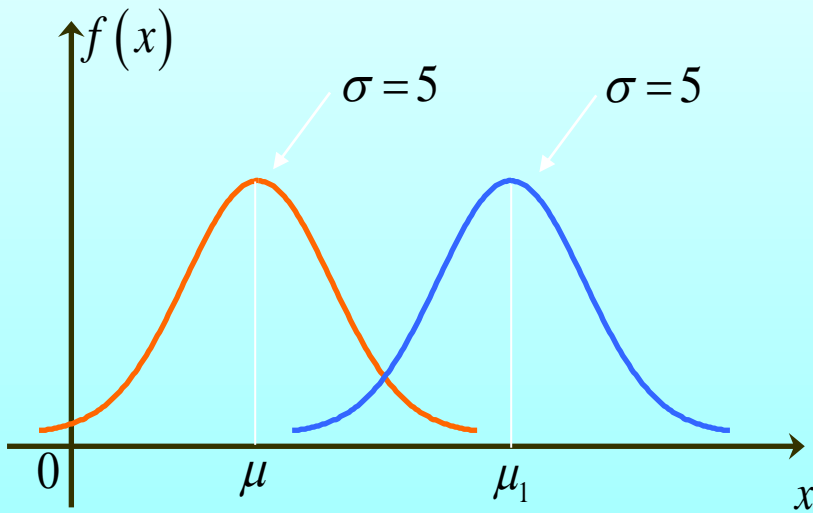
$$Z = \frac{(X - \mu)}{\sigma}$$



正态分布特征

- 1.正态曲线在横轴上方均数处最高,即频数最大
- 2.正态分布以均数为中心,左右对称,无限接近于x轴
- 3.曲线与横轴所围面积为1。正态曲线下面积分布有一定规律
- 4.正态分布有两个参数:均数,标准差
一般用 $N(\mu, \sigma^2)$ 示正态分布

正态分布的参数



μ : 位置参数(决定对称轴位置)

σ : 形状参数(决定曲线形状)
反映 x 取值分散度的指标

正态曲线下面积的分布规律

(cumulative-distribution function, cdf)

正态曲线下一定区间的面积

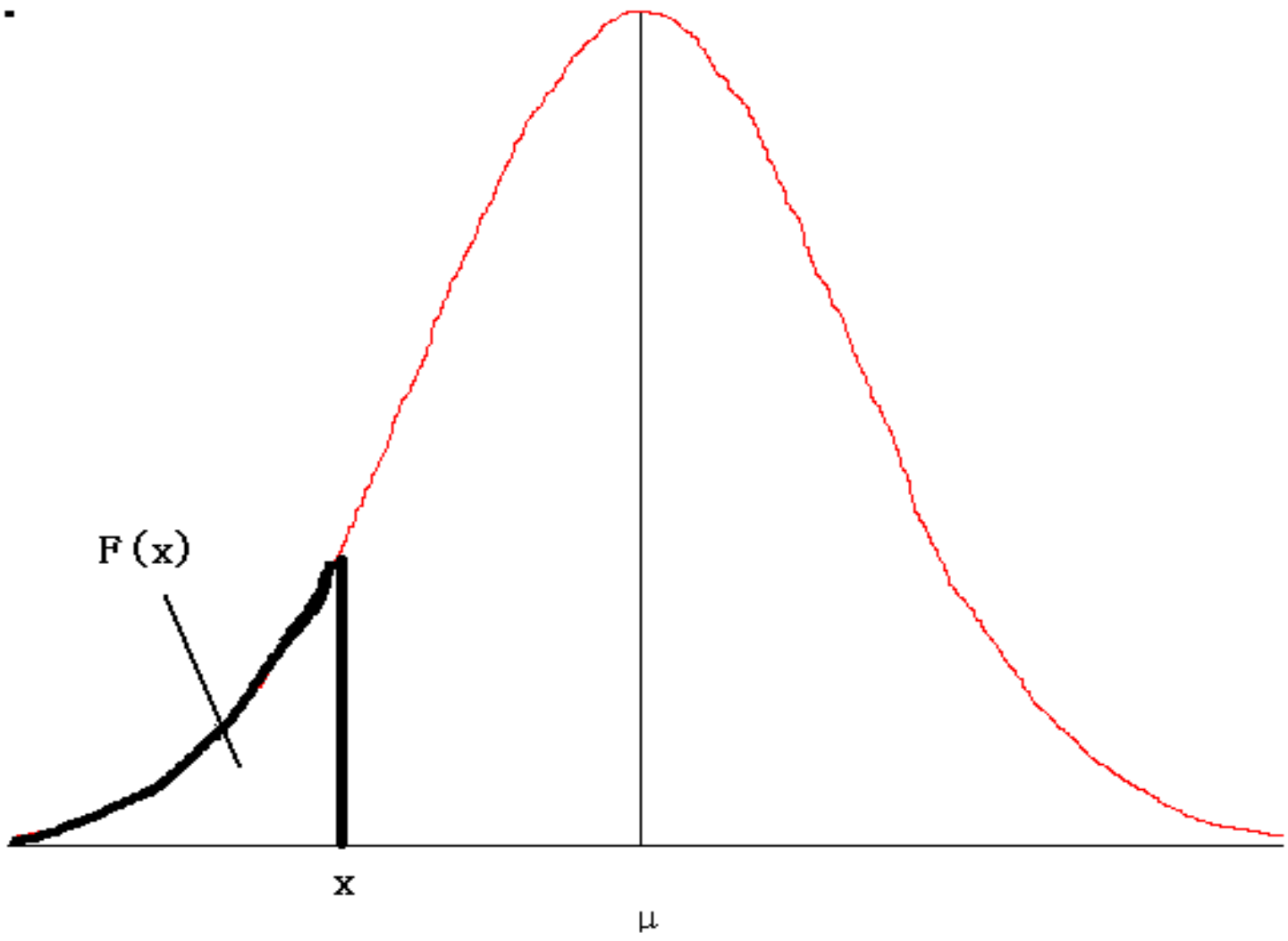
$$P(X \leq x) = F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(x-\mu)^2/(2\sigma^2)} dx$$

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

根据上式，编制附表

$$x \longrightarrow u \longrightarrow P$$

-



• $x \longrightarrow u \longrightarrow P$

查表注意:

1. 利用均数和标准差先对 x 做标准化变换, 求得 u 值

$$u = (x - \bar{X}) / s$$

$$x = \bar{X} + us$$

2. 曲线下对称于0的区间, 面积相等

3. 曲线下横轴上的总面积为100%或1

区间 $(2.58, \infty)$ 的面积 $\varphi(-2.58) = 0.005$

区间 $(-\infty, 2.58)$ 的面积

$$\varphi(2.58) = 1 - \varphi(-2.58) = 0.995$$

$$P(|u| > 1.96) = 2\varphi(-1.96) = 0.05$$

$$P(|u| < 2.58) = \varphi(2.58) - \varphi(-2.58) = 0.99$$

- 求两个特定结果之间的面积，即概率
(由 X 求 u 及 P)

- **例** 随机抽取某市120名成年男子测红细胞计数, 得 $\bar{X}=4.7168$, $S=0.5665$ 。求红细胞计数在 $4-5 \times 10^{12}/L$ 之间的人数及所占比例。

- 解: 已知 $\bar{X}=4.7168$, $S=0.5665$

$$X_1=4 \text{ 时, } u_1=(4-4.7168)/0.5665=-1.265$$

$$X_2=5 \text{ 时, } u_2=(5-4.7168)/0.5665=0.500$$

$$P = \Phi(0.5) - \Phi(-1.265) = (1 - 0.3085) - 0.1029 \\ = 0.5886$$

- 人数: $120 \times 0.5886 \approx 71$ 人 (实际人数为70人)

- **例**某市1982年110名7岁男童身高，均数为119.95cm，标准差为4.72cm，1.试估计该地7岁男童身高在110cm以下者占该地男童总数的百分数。2.分别求下述范围内人数占总数的百分数。

$$\bar{X} \pm 1s$$

$$\bar{X} \pm 1.96s$$

$$\bar{X} \pm 2.58s$$

- **解：**

$$u = \frac{x - \bar{X}}{s} = \frac{110 - 119.65}{4.72} = -2.11$$

$$\Phi(u) = \Phi(-2.11) = 0.0174$$

得该地男童身高在110cm以下者，估计约占1.74%

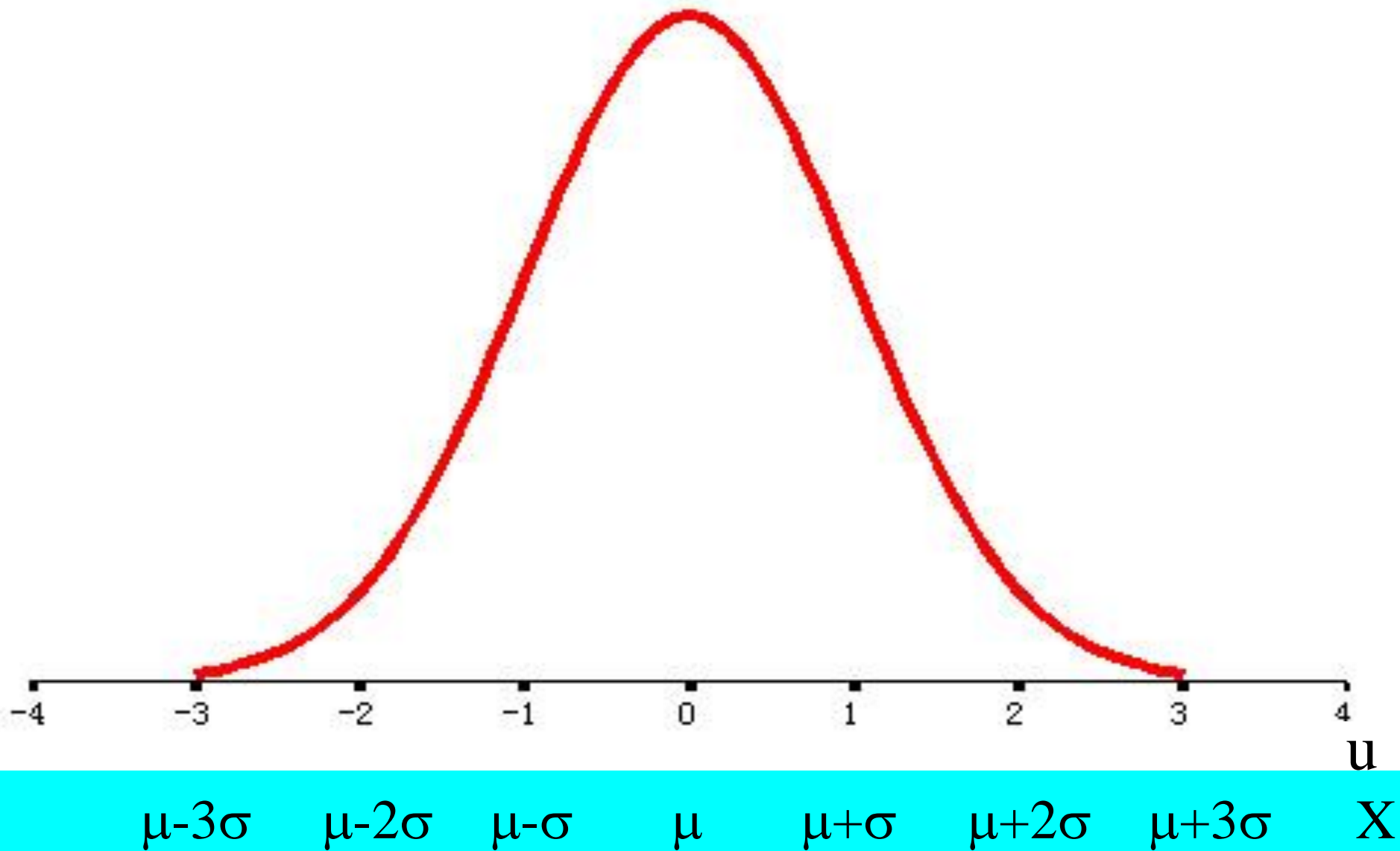
$$\Phi\left(\frac{\bar{X} + 1.96s - \bar{X}}{s}\right) - \Phi\left(\frac{\bar{X} - 1.96s - \bar{X}}{s}\right)$$

$$= \Phi(1.96) - \Phi(-1.96)$$

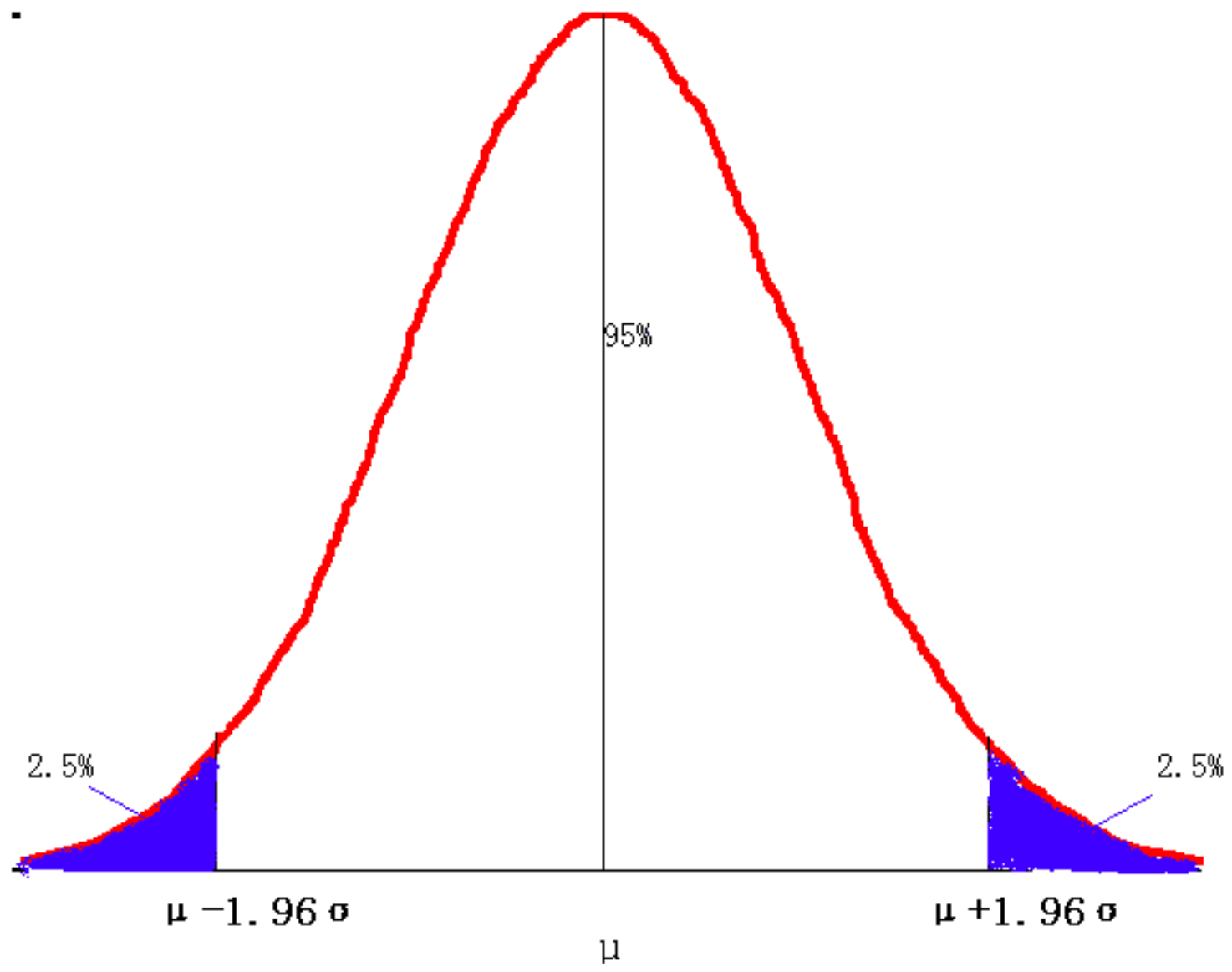
$$= 1 - 2\Phi(-1.96)$$

$$= 1 - 2 \times 0.025 = 95\%$$

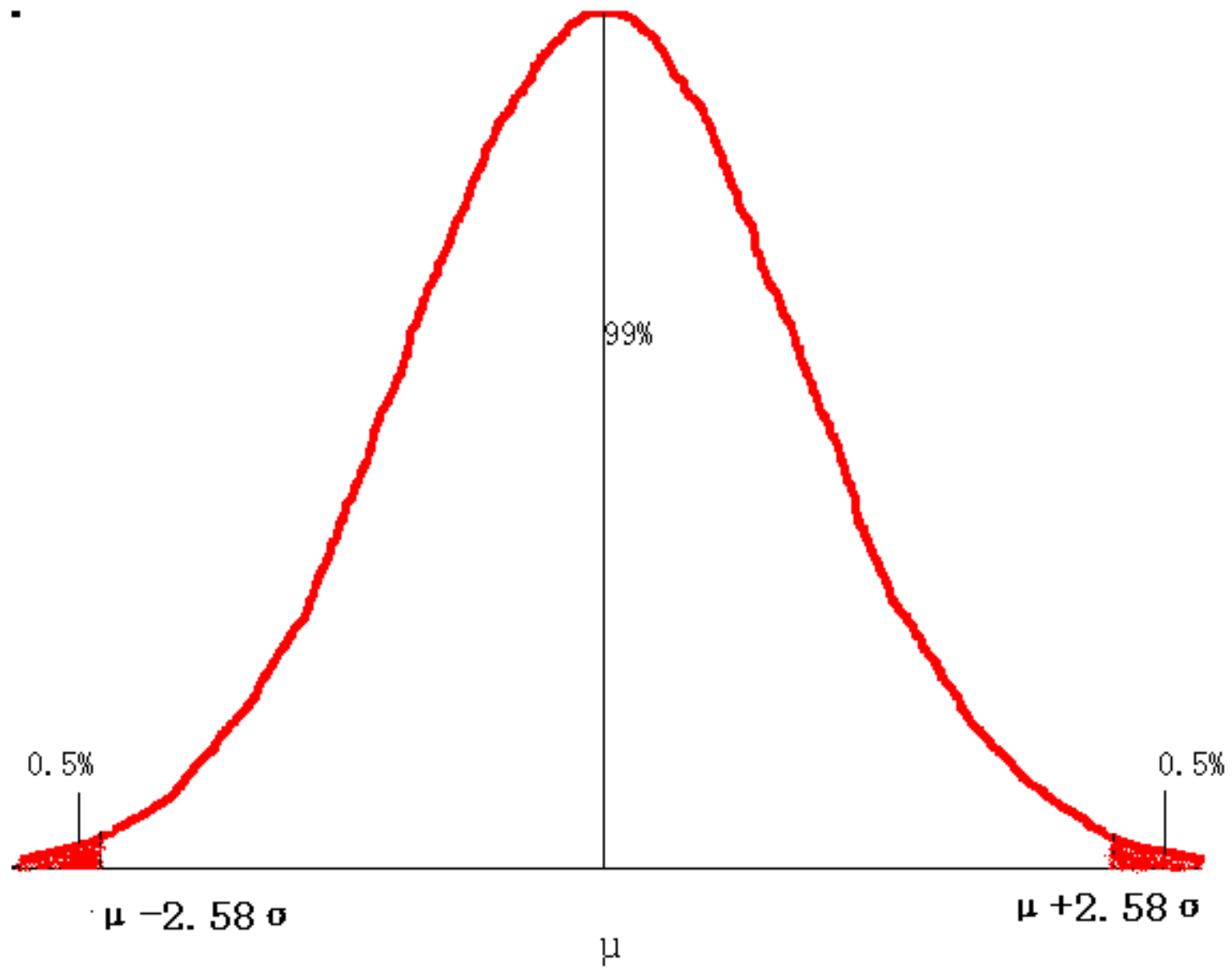
正态分布与标准正态分布的转换



▪



▪



- ● 求某部分面积所对应的变量值
(由 P 求 u 及 X)

- 例上例中, 求中间80%成年男子的红细胞计数范围。

- 解: $P=0.1 \rightarrow u=-1.282 \rightarrow$

根据对称性得区间 $u[-1.282, 1.282] \rightarrow$

$[\bar{X} - 1.282S, \bar{X} + 1.282S] \rightarrow$

$[3.99, 5.44] \times 10^{12}/L$ (实际96人)

- 几个重要区间
双侧

$$u[-1.96,1.96] \rightarrow X[\bar{X}-1.96S, \bar{X} +1.96S]$$
$$\Rightarrow P=0.95$$

$$u[-2.58,2.58] \rightarrow X[\bar{X} -2.58S, \bar{X}+ 2.58S]$$
$$\Rightarrow P=0.99$$

单侧

$$u(-\infty, 1.64] \rightarrow X \leq \bar{X} + 1.64S \\ \Rightarrow P=0.95$$

$$u[-1.64, \infty) \rightarrow X \geq \bar{X} - 1.64S \\ \Rightarrow P=0.95$$

$$u(-\infty, 2.33] \rightarrow X \leq \bar{X} + 2.33S \\ \Rightarrow P=0.99$$

$$u[-2.33, \infty) \rightarrow X \geq \bar{X} - 2.33S \\ \Rightarrow P=0.99$$

120名成年男子RBC的实际分布与理论分布比较

$\bar{X} \pm S$	RBC范围	实际分布		理论分布
		人数	%	%
$\bar{X} \pm 1.00S$	4.15~5.28	83	69.17	68.27
$\bar{X} - 1.64S$	>3.79	113	94.17	95.00
$\bar{X} \pm 1.96S$	3.61~5.83	114	95.00	95.00
$\bar{X} \pm 2.58S$	3.26~6.18	120	100.00	99.00

- 为应用方便，将标准正态分布的双侧尾部面积之和等于 α 时所对应的变量值记为 $z_{\alpha/2}$ ($u_{\alpha/2}$)

$$\text{即 } P(|z| > z_{\alpha/2}) = \alpha$$

- 对数正态分布

原始数据经对数转化后服从正态分布，如正常成人血铅含量，医院病人的住院天数等

参考值范围估计

- 特定的“正常人群” (排除了对所研究指标有影响的疾病和有关的特定人群)的解剖、生理、生化指标及组织代谢含量等数据中大多数个体的取值所在范围 (range of reference value)
- 也称正常值范围 (range of normal value) , 前者较合理。

◆ 参考值范围估计的一般原则与步骤

- ① 确定研究总体，保证研究对象的同质性。
- ② 确定样本量，一般 $n > 100$ ，并判定是否应分层。
- ③ 确定单侧或双侧（根据专业知识）。
- ④ 确定适当百分范围。常取80%，90%，95%，99%等
- ⑤ 选定适当统计方法：
 - 正态分布法
 - 百分位数法

◆ 参考值范围的估计方法

- 正态分布法 (适于正态分布资料或可转化为正态分布的资料) $\bar{X} \pm uS$

- ◆ 双侧95%参考值范围: $\bar{X} \pm 1.96S$

- ◆ 单侧95%参考值范围: $> \bar{X} + 1.64S$
或 $< \bar{X} - 1.64S$

- 百分位数法 (适于非正态分布资料), 所需n大于正态分布法

- ◆ 双侧95%参考值范围: $[P_{2.5}, P_{97.5}]$

- ◆ 单侧95%参考值范围: $> P_5$ 或 $< P_{95}$

例 某地调查正常成年男子144人的红细胞数，均数 $5.38 \times 10^{12}/L$ ，标准差为 $0.44 \times 10^{12}/L$ 。试估计该地成年男子红细胞数的95%参考值范围。

● 下限 $\bar{X} - 1.96s = 5.38 - 1.96(0.44)$

$$= 4.52(10^{12} / L)$$

● 上限 $\bar{X} + 1.96s = 5.38 + 1.96(0.44)$

$$= 6.24(10^{12} / L)$$

【例】某市1974年为了解该地居民发汞的基础水平，调查了留住该市一年以上，无汞作业接触史的健康居民238人的发汞含量如下表，试估计该市健康居民发汞值的95%参考值范围。

解：求单侧95%上限，用百分位数法，即求 P_{95}

$$P_r = L_r + \frac{i_r}{f_r} (n \times r\% - f_{r-})$$

$$P_{95} = 2.3 + \frac{0.4}{16} (238 \times 95\% - 212) = 2.65$$

该市健康居民发汞值的95%参考值范围应小于
2.65 $\mu\text{g/g}$ 。

某市238名健康人发汞含量

发汞值 ($\mu\text{g/g}$)	人数	累积频数	累积频率 (%)
0.3-	20	20	8.40
0.7-	66	86	36.13
1.1-	60	146	61.34
1.5-	48	194	81.51
1.9-	18	212	89.08
2.3-	16	228	95.80
2.7-	6	234	98.32
3.1-	1	235	98.74
3.5-	0	235	98.74
3.9-4.3	3	238	100.00
合计	238		

参考值范围与医学决定水平

- 医学决定水平是一个界限值。它是决定人体健康与否的一条界限，来源于大量的医学实践，它的升高和降低决定着某种疾病的是与否，在临床治疗过程中带给临床医生的意义是采用什么方式对病人进行治疗和治疗是否见效。它具有诊断和治疗的决定性使用价值，因此被称为医学决定水平。
- 医学决定水平看来更合理、更客观、更有助于临床的应用。当然，真正建立起每一项试验的医学决定水平是一个十分复杂的问题，存在着许多的实际困难

- 白细胞参考值 $(4\sim 10) \times 10^9/L$
- 决定水平临床意义及措施
 - $0.5 \times 10^9/L$: 低于此值, 病人有高度易感染性, 应采取相应的预防性治疗及预防感染措施。
 - $3 \times 10^9/L$: 低于此值为**白细胞减少症**, 应再作其他试验, 如**白细胞分类计数**、观察外周血涂片等, 并应询问用药史。
 - $11 \times 10^9/L$: 高于此值为**白细胞增多**, 此时作白细胞分类计数有助于分析病因和分型, 如果需要应查找感染源。
 - $30 \times 10^9/L$: 高于此值, 提示可能为白血病, 应进行**白细胞分类**, 观察外周血涂片和进行骨髓检查

小结

- 正态分布的分布特征
- 正态分布变量的 $z(u)$ 变换
- 正态分布曲线下面积的分布规律
- 如何用正态分布法和百分位数法计算参考值范围