第三章 概率与分布

概率

- 概率是0和1之间的一个数目，表示某个事件发生的可能性或经常程度。
- 你买彩票中大奖的机会很小(接近0)
- 但有人中大奖的概率几乎为1
- 你被陨石击中的概率很小(接近0)
- 但每天有陨石击中地球的概率为1
- 你今天被汽车撞上的概率几乎是0
- 但在北京每天发生车祸的概率是1。

- 发生概率很小的事件称为小概率事件(**small probability event**)；
- 小概率事件不那么可能发生，但它往往比很可能发生的事件更值得研究。
- 在某种意义上，新闻媒体的主要注意力大都集中在小概率事件上。

一、概率（Probability）——机会大小的度量

- 概率是表示某事件出现可能性大小的一种指标。记为：P
- 概率的值域：［0，1］
- P（Ω）＝1：必然事件；
- P（φ）＝0：不可能事件。
- 0<P<1：随机事件。
  - 随机事件的特点：（随机——随机会而定）
    - 一次试验前，不能预言其是否会发生，这说明随机事件具有偶然性；
    - 在相同条件下，进行大量次重复试验，呈现出统计规律性，这说明随机事件具有规律性。

试验与样本空间

- 试验：任何可以产生明确定义的结果的过程。
- 在试验的一次重复中，**有且只有一个**可能的试验结果发生。
- 当试验**所有**可能的结果都已经确定时，我们就确定了该试验的**样本空间**。也就是说，***样本空间定义为所有可能的试验结果的集合***。任何一个特定的试验结果都被称为**样本点**，这是样本空间的组成元素。

试　　验	试验结果
抛掷一枚硬币	正面、反面
对某一零件进行检测	合格、不合格
拨打一次销售电话	购买、不购买
投掷一次骰子	1、2、3、4、5、6
进行一场足球比赛	获胜、失利、平局

返回

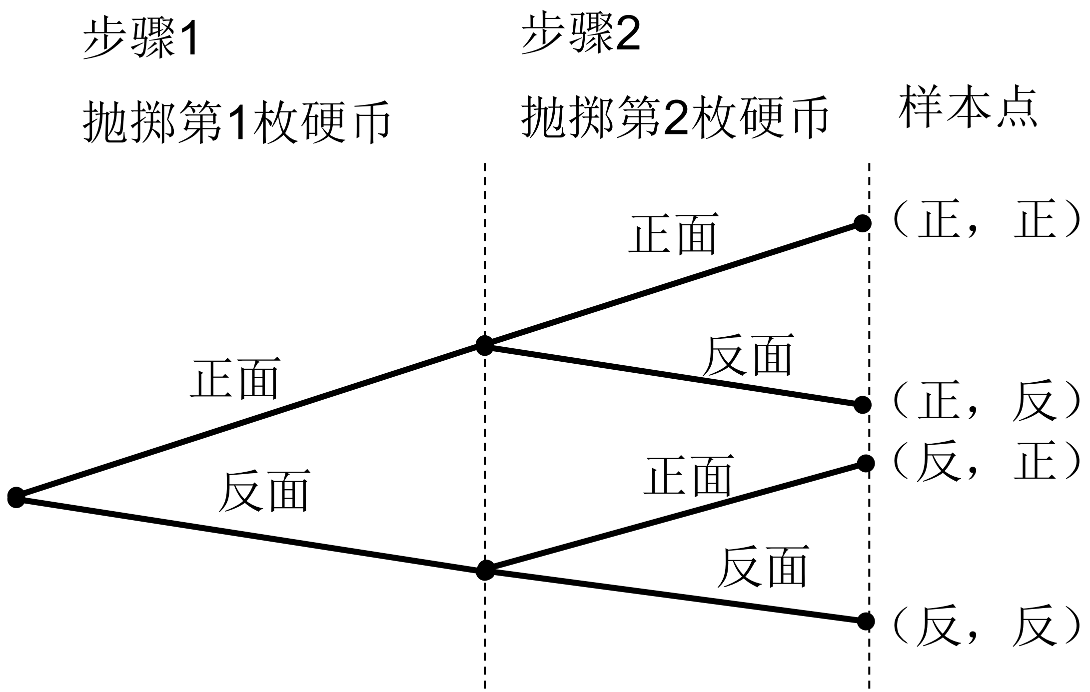
分析试验和列举试验结果的工具——树形图

- 为了理解进行试验时究竟发生了什么，一个重要的步骤就是对该试验的样本点进行确认和计数。
  - 例1。
  - 例2。
- 如果一个试验可以分为连续的k个步骤，在第1步中有 n_1 种试验结果，在第2步中有 n_2 种试验结果，以此类推。那么所有可能的试验结果为 $n_1 n_2 \cdots n_k$ 种。也就是说，整个试验的结果数目等于各个步骤结果数目的连乘积。

返回

树形图分析例1

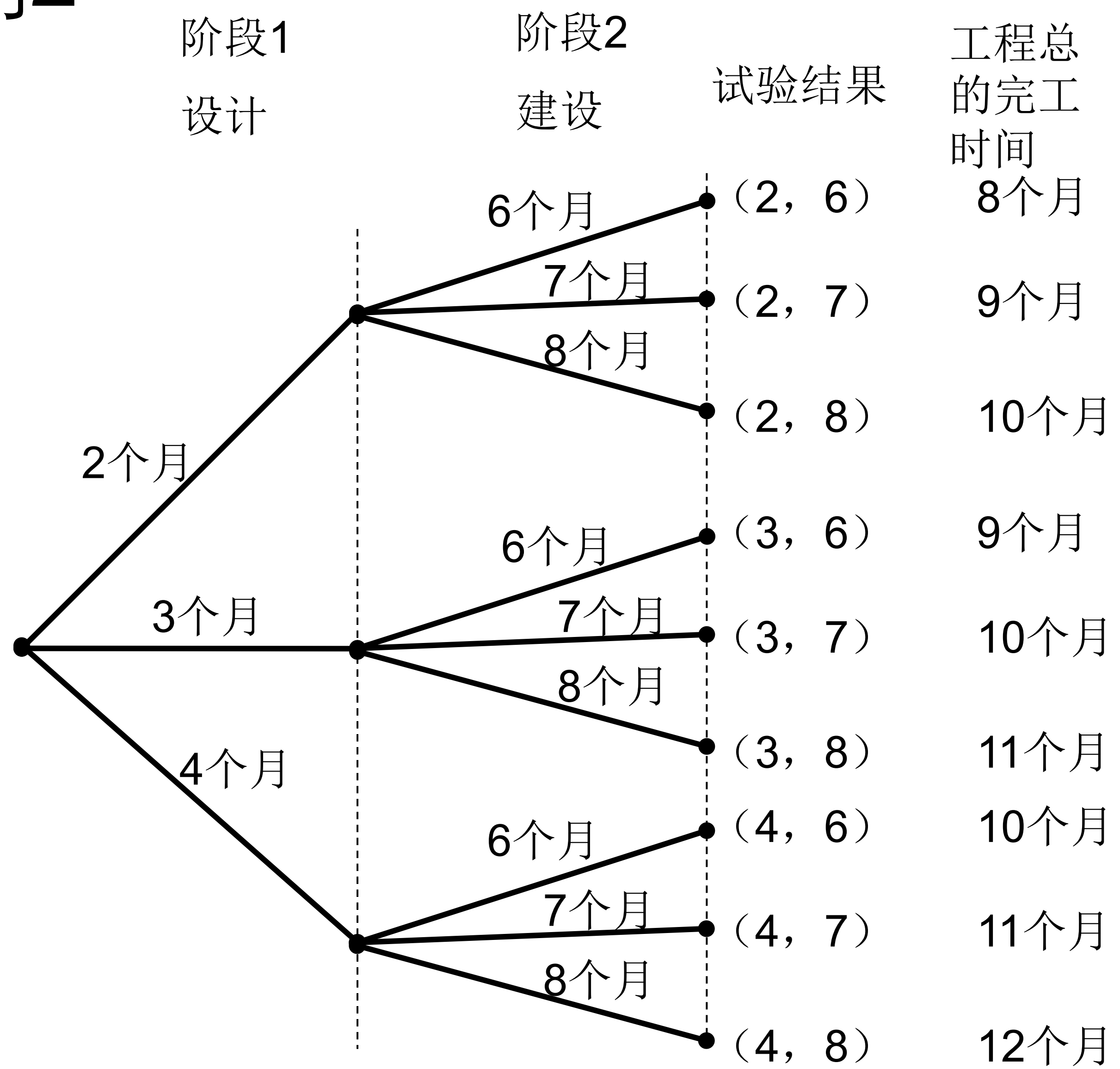
- 考虑抛两枚硬币的试验，有多少种可能的试验结果（样本点）？



返回

树形图分析例2

- 某市欲进行一项工程建设。该工程可以分为两个连续的阶段：阶段1（设计）和阶段2（建设）由于每个阶段都会被尽量紧密地安排和规划，管理人员无法预知该工程的每个阶段完成的确切时间。对类似工程的分析显示，完成计划阶段需要2、3或4个月，而完成建设阶段需要6、7或8个月。由于该项工程的紧迫性，管理人员伴郎确定的完成该工程的时间目标为10个月。



返回

二、后验概率——统计频率

- 设观察了n次试验，而其中事件A出现了m次，m/n称为事件A的频率。
- 当m愈来愈大时，频率m/n虽有些摆动，但幅度愈来愈小而最终会“趋近”于某一介于0与1之间的值p，就把这个p定义为事件A的概率。

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

- 计算统计概率的两个前提条件：
  - 每次试验中某一事件发生的可能性不变；
  - 试验能大量重复，且每次试验相互独立。

三、先验概率——古典概率

- 古典概率模型要求试验满足以下两个条件：
  - 试验的所有可能结果（基本事件）的个数是有限的。
  - 每次试验中每个基本事件出现的可能性相等。
- 定义：在古典概型下，若基本事件总数为n，而事件A包括m个基本事件，则事件A发生的概率为：

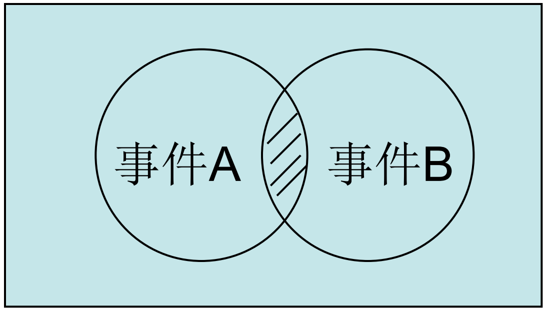
$$P(A)=\frac{m}{n}$$

- 古典的特点：在试验之前，我们就能决定事件发生的概率。故又称这种概率为**先验概率**。

四、概率的定理——加法定理

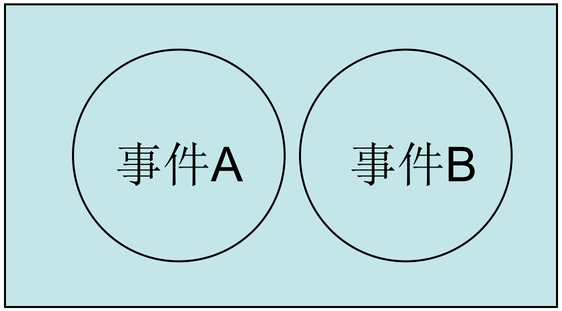
- 加法公式：

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$



- 互不相容事件的情况：

$$P(A\cup B)=P(A)+P(B)$$

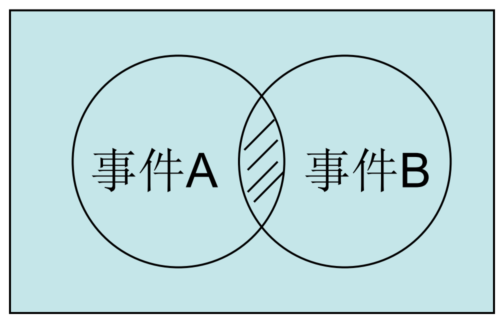


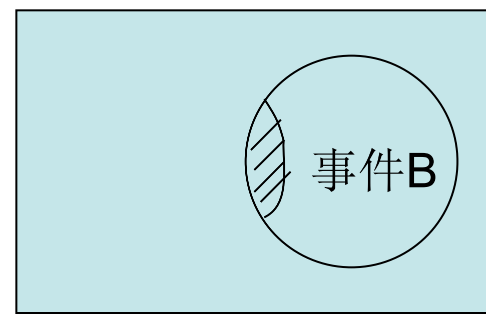
- 互不相容的两事件A、B和的概率等于它们概率的和

五、概率的定理__条件概率

- 某个事件的概率经常会因为某个相关事件的发生与否而受到影响。假设我们有一个事件A，概率为P(A)，如果我们获得了新的信息，即确知另外一个事件，记为B，已经发生了。我们希望利用这一新的信息来对事件A重新计算概率。事件A的新概率记为P(A|B)。符号“|”用来表明我们考虑的是在事件B已经发生的条件下A的概率。因此，符号P(A|B)读作“给定B下A的概率”。

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$





$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$

<u>条件概率计算实例。</u>

六、概率的定理——乘法定理

乘法公式：

$$P(A \cap B) = P(B)P(A \mid B)$$

$$P(A \cap B) = P(A)P(B \mid A)$$

连续发生的独立事件的概率，是事件各次发生的概率相乘之积。

独立事件的乘法公式：

$$P(A \cap B) = P(A)P(B)$$

独立事件的定义：

$P(A \mid B) = P(A)$或$P(B \mid A) = P(B)$

有趣的概率题

- 一个班的40人当中，有没有生日相同的人？

- 有
  10支签，其中放有3支中奖的签。由A先
  抽签，B后抽签。试比较两人的中奖概率。

五、离散型概率分布——泊松分布

- 泊松分布在估计特定时间或空间出现的次数方面非常有用。
  - 例如：感兴趣的随机变量可以是一小时内到达汽车清洗处的汽车数、在10公里长的高速公路上需要修理的汽车数目或100公里长的水管的漏洞个数等。
- 泊松试验的性质：
  - 任意两个相等长度的区间发生一次的概率相等。
  - 任意区间发生或不发生与其他区间发生与否独立。

泊松概率函数 $$f(x)=\frac{\mu^{x}e^{-\mu}}{x!}$$

式中：$f(x)$——在一个区间发生x次的概率

μ——在一个区间发生次数的数学期望或均值数

$e=2.71828$

泊松应用实例。

条件概率计算实例

- 美国某大城市警察局，男性和女性警官共有1200名，其中男性960名，女性240人。在过去的两年中，有324名警官得到了升职，情况如下表所示。

$$P(A|M)=\frac{P(A\cap M)}{P(M)}$$

$$=\frac{0.24}{0.80}=0.30$$

$$P(A|W)=\frac{P(A\cap W)}{P(W)}$$

$$=\frac{0.03}{0.20}=0.15$$

	男性（M）	女性（W）	合计
升职人数	288	36	324
未升职人数	672	204	876
合计	960	240	1200

	男性（M）	女性（W）	合计
升职(A)	0.24	0.03	0.27
未升职(B)	0.56	0.17	0.73
合计	0.80	0.20	1.00

返回

40人中至少有2人生日相同的概率

2个人生日不同的概率 $=\frac{364}{365}$

3个人生日不同的概率 $=\frac{364}{365}\times\frac{363}{365}$

3个人生日不同的概率 $=\frac{364}{365}\times\frac{363}{365}\times\frac{362}{365}$

g

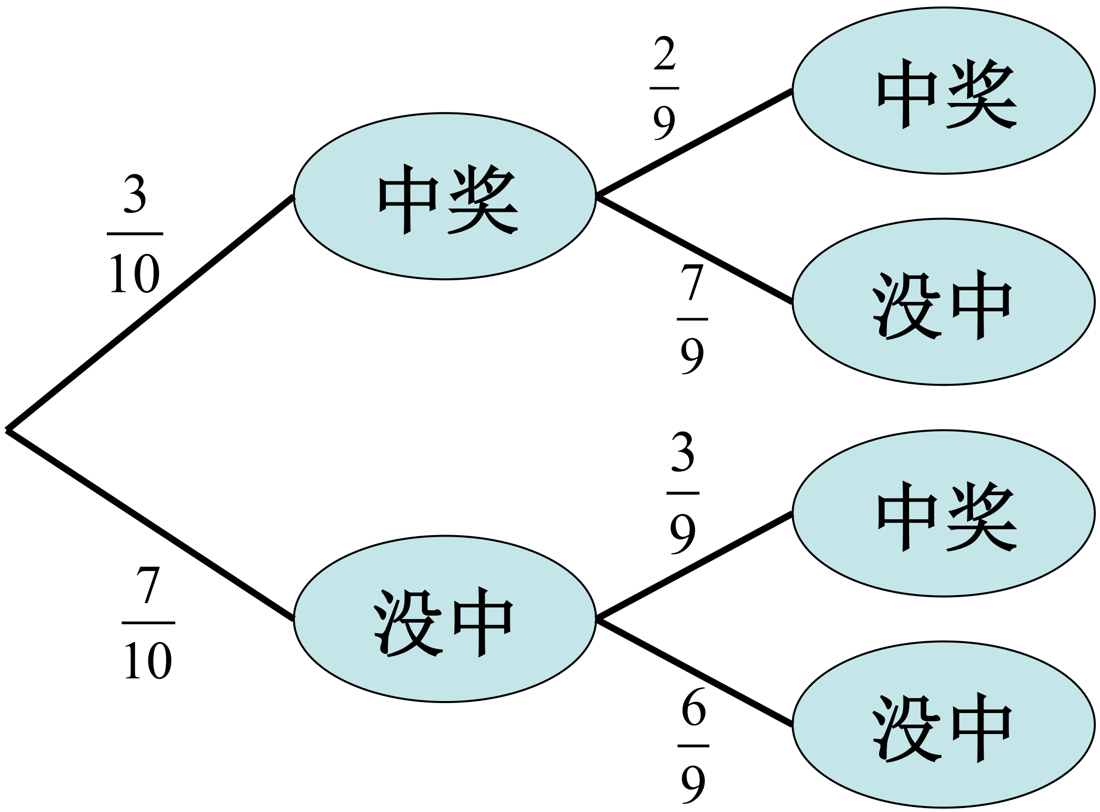
g

g

n个人生日不同的概率 $=\frac{364}{365}\times\frac{363}{365}\times\frac{362}{365}\times\cdots\times\frac{365\text{-}n+1}{365}$

返回

两人的中奖概率问题



$\frac{3}{10}\times\frac{2}{9}=\frac{1}{15}$ 条件概率

$\frac{7}{10}\times\frac{3}{9}=\frac{7}{30}$ 条件概率

$\frac{1}{15}+\frac{7}{30}=\frac{3}{10}$

返回

泊松分布应用实例

- 我们希望研究的是周日早上15分钟内到达某一收费站的汽车数。假定任意两个相等长度时间内汽车到达的概率相等且在任意一段时间汽车到达与否与其他时段汽车到达与否相互独立。根据历史数据的分析表明，15分钟期间平均到达的车辆数为10。管理员想要知道15分钟内恰好有5辆车到达的概率。

$$f(x) = \frac{10^x e^{-10}}{x!}$$

$$f(5) = \frac{10^5 e^{-10}}{5!} = 0.0378$$

返回

分布

- 随机变量取一切可能值或范围的概率或概率的规律称为概率分布(probability distribution，简称分布)。
- 概率分布可以用各种图或表来表示；一些可以用公式来表示。概率分布是关于总体的概念。

分组区间	划记次数	组中值	次数（f）
96-98		97	2
93-95		94	3
90-92		91	4
87-89		88	8
84-86		85	11
81-83		85	17
78-80		79	19
75-77		76	14
72-74		73	10
69-71		70	7
66-68		67	3
63-65		64	1
60-62		61	1
合计			100

第一节 分布概述分布

- 前面介绍过的样本均值、样本标准差和样本方差等样本特征的概念是相应的总体特征的反映。

- 分布的类型:
  - 离散型分布和连续型分布
  - 经验分布和理论分布
  - 基本分布和抽样分布

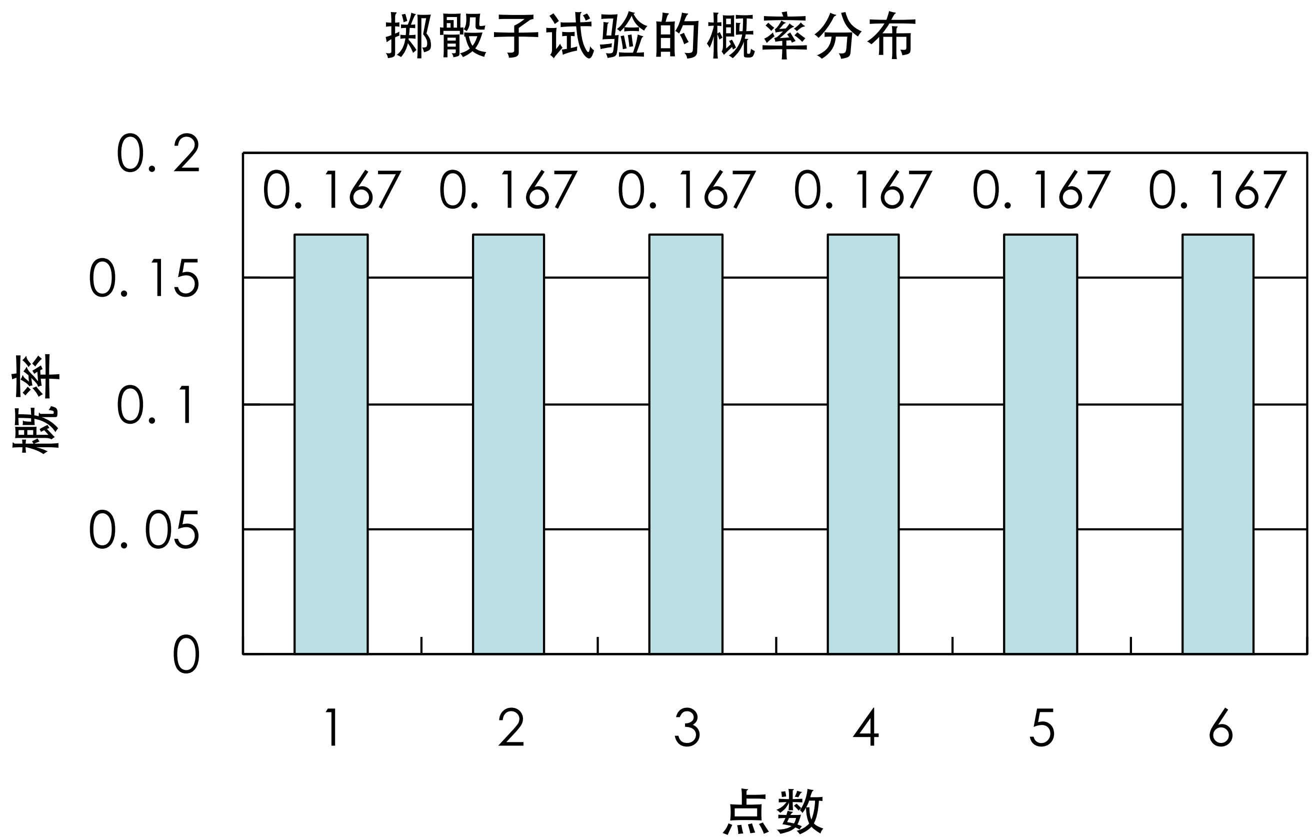
离散变量的分布

- 离散变量只取离散的值，比如骰子的点数、网站点击数、顾客人数等等。每一种取值都有某种概率。各种取值点的概率总和应该是**1**。

- 当然离散变量不不仅仅限于取非负整数值。

- 一般来说，某离散随机变量的每一个可能取值x_i都相应于取该值的概率$p(x_i)$，这些概率应该满足关系

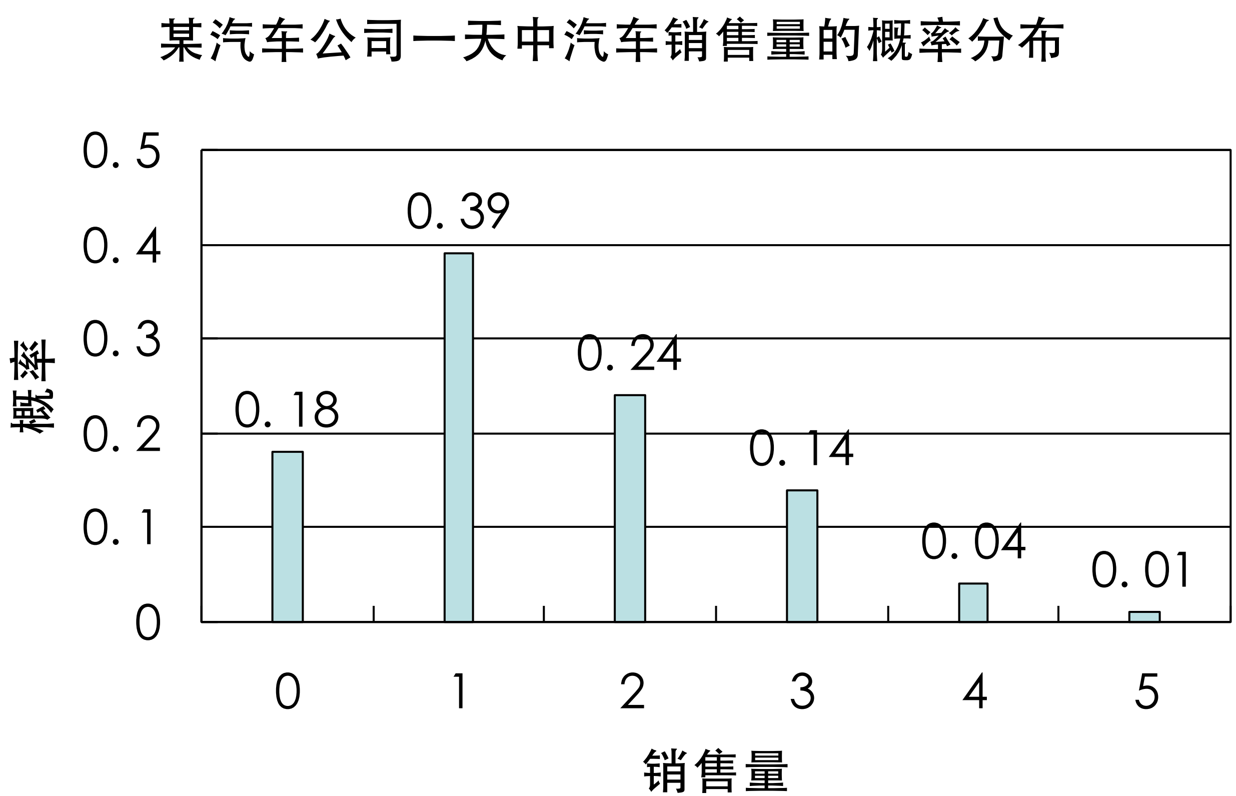
$$\sum_i p(x_i) = 1, \quad p(x_i) \ge 0$$



掷骰子试验的概率分布
0.2
0.15
0.1
0.05
0
概率
0.167
0.167
0.167
0.167
0.167
0.167
1
2
3
4
5
6
点数

某汽车公司一天中汽车销售量的概率分布

X（一天中销售汽车数）	0	1	2	3	4	5
$P(X=x_i)=P(x_i)=p_i$	0.18	0.39	0.24	0.14	0.04	0.01



- 如所关心的是两骰子点数之和，则下表包含了所有36种可能试验结果的搭配和相应的点数和。

两骰子点数和		第一个的点数					
		1	2	3	4	5	6
第二个的点数	1	**2**	**3**	**4**	**5**	**6**	**7**
	2	**3**	**4**	**5**	**6**	**7**	**8**
	3	**4**	**5**	**6**	**7**	**8**	**9**
	4	**5**	**6**	**7**	**8**	**9**	**10**
	5	**6**	**7**	**8**	**9**	**10**	**11**
	6	**7**	**8**	**9**	**10**	**11**	**12**

可以看出，如果我们考虑点数和等于2的事件，则仅有一种可能的试验结果（两个骰子均为一点）；而如果我们考虑点数和等于7的事件，则有六种可能的试验结果。两个骰子点数之和总共有2至12等11种可能，即有11种可能的事件，而这11种事件相应于上面所说的36种可能的试验结果的一些集合。这些事件和试验结果的集合归纳在下面表中：

事件: 两骰子点数和	**集合:** 相应的试验结果（两个数字分别表示第一和第二个骰子的点数）	集合中元素的个数	**事件的概率**
2	(1,1)	1	1/36
3	(1,2) (2,1)	2	2/36
4	(1,3) (2,2) (3,1)	3	3/36
5	(1,4) (2,3) (3,2) (4,1)	4	4/36
6	(1,5) (2,4) (3,3) (4,2) (5,1)	5	5/36
7	(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	6	6/36
8	(2,6) (3,5) (4,4) (5,3) (6,2)	5	5/36
9	(3,6) (4,5) (5,4) (6,3)	4	4/36
10	(4,6) (5,5) (6,4)	3	3/36
11	(5,6) (6,5)	2	2/26
12	(6,6)	1	1/36

例. 从某大学到火车站途中有**6**个交通岗,假设在各个交通岗是否遇到红灯相互独立,并且遇到红灯的概率都是**1/3.**

(1)设**X**为汽车行驶途中遇到的红灯数,求**X**的分布律.

(2)求汽车行驶途中至少遇到**5**次红灯的概率.

解:(1)由题意,X~B(6,1/3),于是,X的分布律为:

$$P\{X=k\}=C_6^k\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{6-k}\qquad k=0,1,\dots,6$$

(2) $$P\{X\geq 5\}=P\{X=5\}+P\{X=6\}$$

$$=C_6^5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)+\left(\frac{1}{3}\right)^6=\frac{13}{729}$$

第二节　二项分布和Poisson分布

一、二项试验

基于二项试验的分布是常见的离散分布之一。

- 二项试验的特点：

（1）**n**次重复试验相互独立；
（2）每次试验有两种可能结果**A**和非**A**；
（3）每次试验中事件**A**发生的概率固定（p）。

- 下面试验可看成为二项试验：
  - 进入某商场的顾客有多少人购买某商品；
  - 被调查对象中有多少人认可某种产品
  - 新生婴儿中有多少个男孩

二、二项分布

- 关于二项试验，最常见的问题是：

如果进行***n***次独立重复试验，每次事件**A**发生的概率均为***p***（**0<P<1**），那么事件**A**恰好发生***k***次的概率是？

- **Bernoulli概型定理：** $P_{\mathrm{n}}(\mathrm{k}) = C_n^k p^k (1-p)^{n-k}$

若以X表示n重贝努里试验事件A发生的次数，则X服从参数为n,p的二项分布。

记作X~B（n,p）

,其分布律为：

$$P\{X=k\}=C_n^k p^k (1-p)^{n-k}, (k=0,1,\ldots,n)$$

$$(0<q=1-p<1)$$

例**4.** 某人射击的命中率为**0.02**，他独立射击**400**次，试求其命中次数不少于**2**的概率。

解　设X表示400次独立射击中命中的次数，

则$X \sim B(400, 0.02)$，故

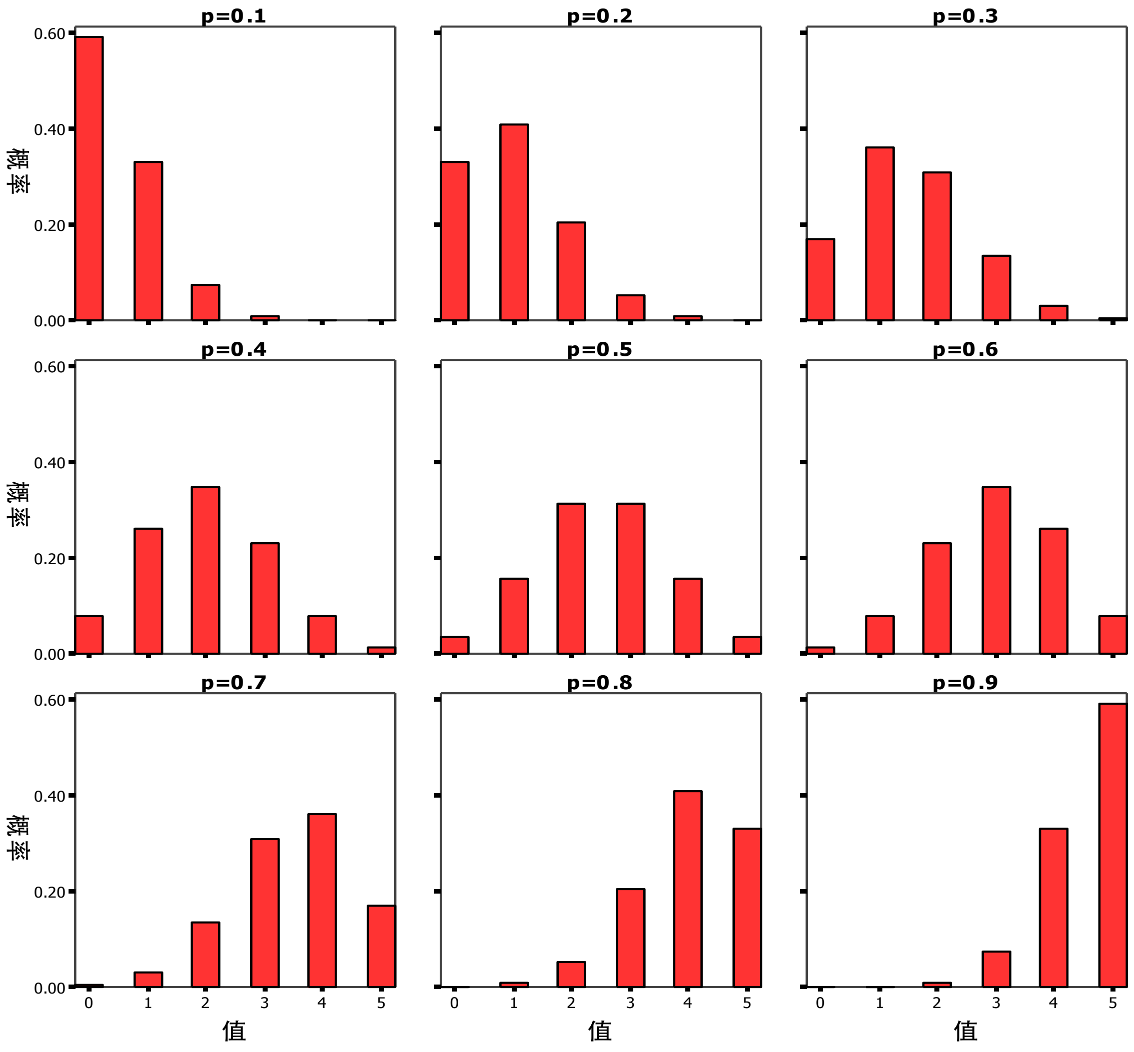
$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\}$$

$$= 1 - 0.98^{400} - (400)(0.02)(0.98^{399}) = \ldots$$

三、二项分布的性质

- 二项分布是离散型分布，概率直条（线）图是越阶式的。
  - p=q时对称；p≠q时偏倾；
  - p<q且np>=5时渐进正态分布。
- 二项分布的数字特征
  - $\mu = np$　　　$\sigma = (npq)^{1/2}$
- 二项分布是一族分布，其形态依赖n、p

图4.1 九个二项分布B(5,p) (p＝0.1到0.9)的概率分布图



四、Poisson分布

- 泊松分布在估计特定时间或空间出现的次数方面非常有用。
  - 例如：一小时内到达汽车清洗处的汽车数;
  - 在10公里长的高速公路上需要修理的汽车数目;
  - 或100公里长的水管的漏洞个数等;
  - 单位时间内到达某一服务柜台请求服务的顾客数;
  - 放射性物质放射出来并到达某区域的粒子数;
- 泊松试验的性质：
  - 任意两个相等长度的区间发生一次的概率相等。
  - 任意区间发生或不发生与其他区间发生与否独立。

参数为λ的Poisson分布变量的概率分布为:

$$P(k)=\frac{\lambda^{k}e^{-\lambda}}{k!},k=0,1,2,\cdots$$

（*p*(k)表示Poisson变量等于k的概率），其中λ为单位区间内发生次数的数学期望。

E(X)= λ

D(X)= λ

例　　　设某企业职工周一请事假的人数**X**近似泊松分布，且设周一请事假的平均数为2.5人,

求:1)**X**的标准差;

2)在给定的某周一正好请事假是5人的概率.

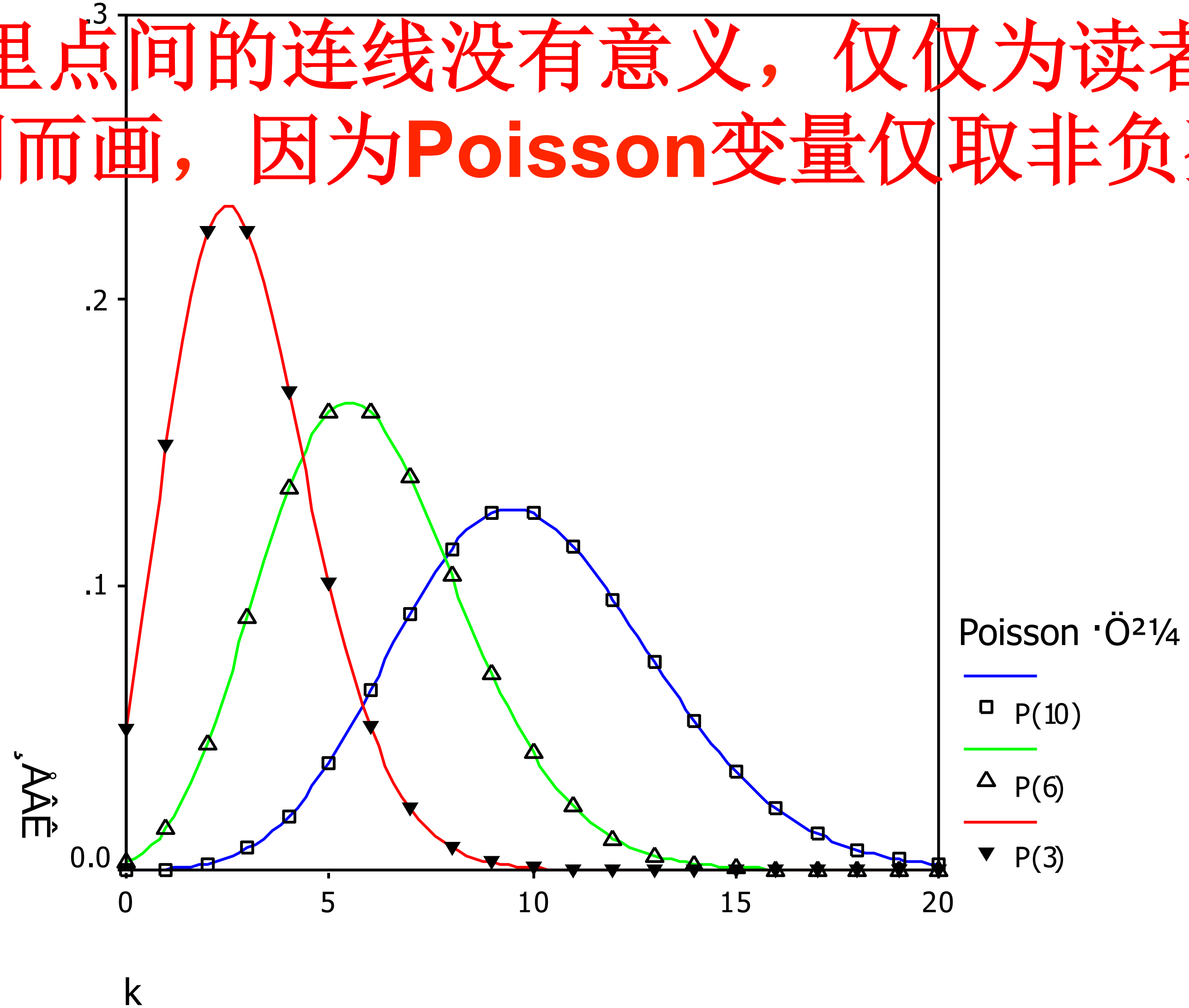
解:**E(X)=2.5, λ=E(X)=D(X)=2.5, σ=D(X)$^{1/2}$=2.5$^{1/2}$=1.581**

$$P(X)=P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$$

$$P(X=5)=\frac{2.5^5 e^{-2.5}}{5!}=0.067$$

参数为3、6、10的Poisson分布（只标出了20之内的部分）

这里点间的连线没有意义，仅仅为读者容易识别而画，因为Poisson变量仅取非负整数值



泊松分布应用实例

- 我们希望研究的是周日早上**15**分钟内到达某一收费站的汽车数。假定任意两个相等长度时间内汽车到达的概率相等且在任意一段时间汽车到达与否与其他时段汽车到达与否相互独立。根据历史数据的分析表明，**15**分钟期间平均到达的车辆数为**10**。管理员想要知道**15**分钟内恰好有**5**辆车到达的概率。

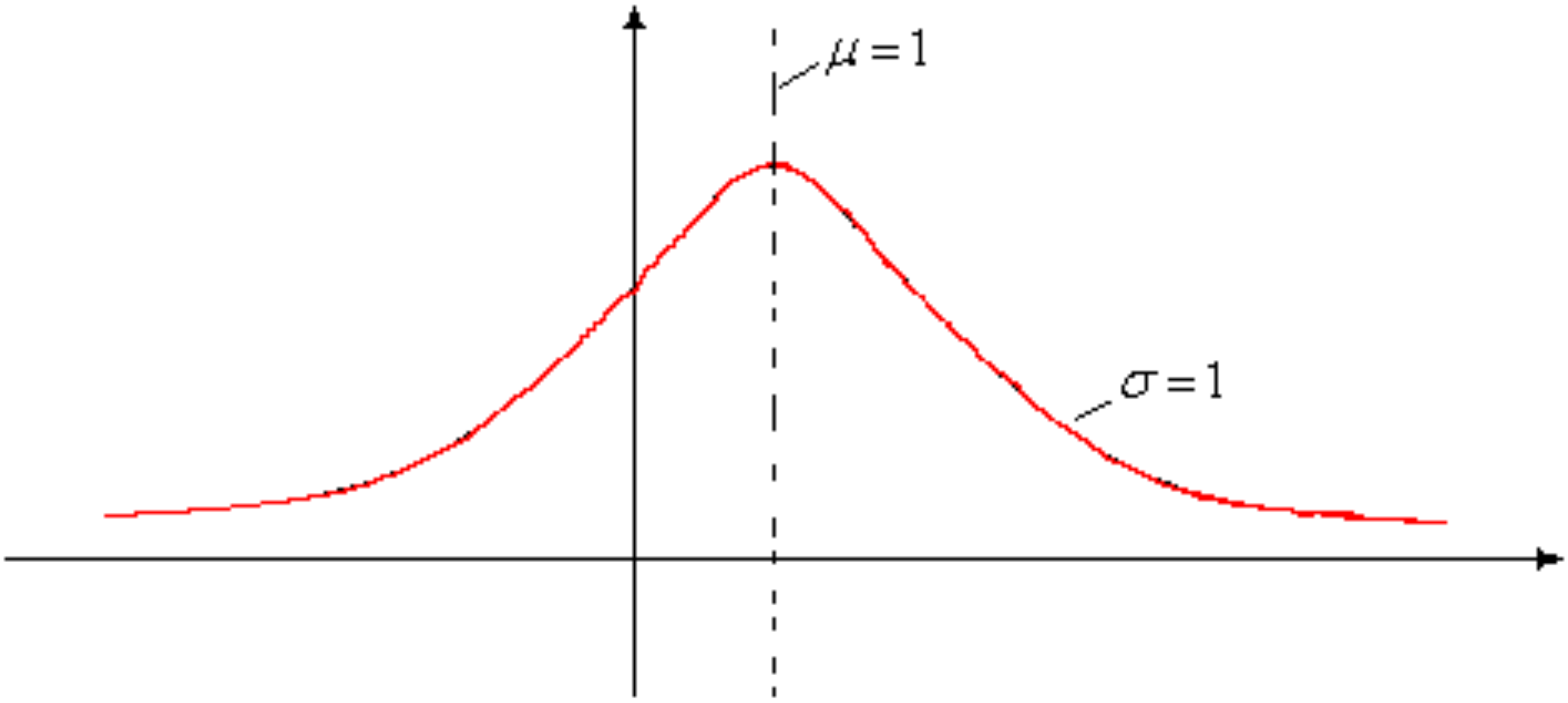
$$f(x) = \frac{10^x e^{-10}}{x!}$$

$$f(5) = \frac{10^5 e^{-10}}{5!} = 0.0378$$

连续变量的分布

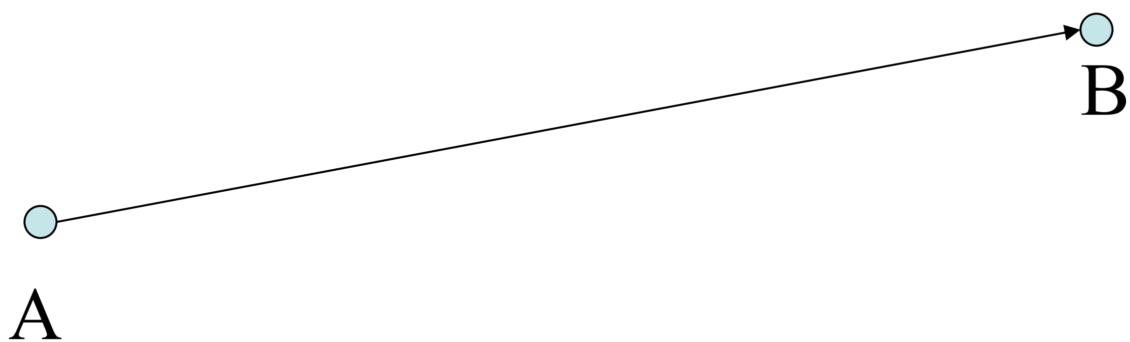
- 取连续值的变量，如高度、长度、重量、时间、距离等等；它们被称为连续变量 **(continuous variable)**。
- 换言之，一个随机变量如果能够在一区间（无论这个区间多么小）内取任何值，则该变量称为在此区间内是连续的，其分布称为连续型概率分布。
- 它们的概率分布很难准确地用离散变量概率的条形图表示。

正态分布



一、概述

正态分布是实践中应用最为广泛，在理论上 研究最多的分布之一，故它在概率统计中占有特 别重要的地位。

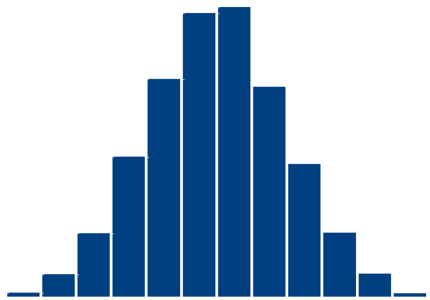


A，B间真实距离为μ，测量值为X。
X- μ的概率分布应该是什么形态？

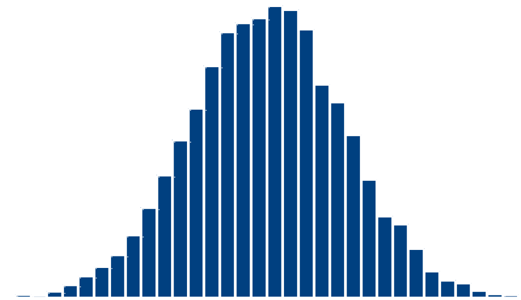
<table>
<tr><th>分组区间</th><th>划记次数</th><th>组中值</th><th>次数（f）</th></tr>
<tr><td>96-98</td><td>||</td><td>97</td><td>2</td></tr>
<tr><td>93-95</td><td>|||</td><td>94</td><td>3</td></tr>
<tr><td>90-92</td><td>||||</td><td>91</td><td>4</td></tr>
<tr><td>87-89</td><td>|||| |||</td><td>88</td><td>8</td></tr>
<tr><td>84-86</td><td>|||| |||| |</td><td>85</td><td>11</td></tr>
<tr><td>81-83</td><td>|||| |||| |||| ||</td><td>85</td><td>17</td></tr>
<tr><td>78-80</td><td>|||| |||| |||| ||||</td><td>79</td><td>19</td></tr>
<tr><td>75-77</td><td>|||| |||| ||||</td><td>76</td><td>14</td></tr>
<tr><td>72-74</td><td>|||| ||||</td><td>73</td><td>10</td></tr>
<tr><td>69-71</td><td>|||| ||</td><td>70</td><td>7</td></tr>
<tr><td>66-68</td><td>|||</td><td>67</td><td>3</td></tr>
<tr><td>63-65</td><td>|</td><td>64</td><td>1</td></tr>
<tr><td>60-62</td><td>|</td><td>61</td><td>1</td></tr>
<tr><td>合计</td><td></td><td></td><td>100</td></tr>
</table>

逐渐增加矩形条数目的直方图和一个形状类似的密度曲线。

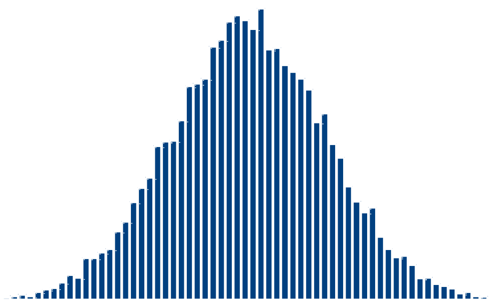
(1)



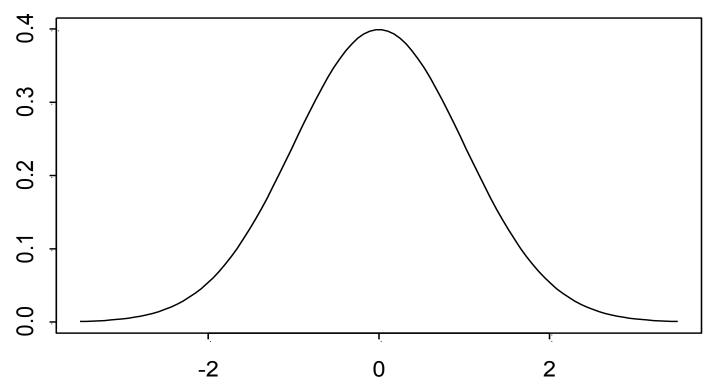
(2)



(3)



(4)

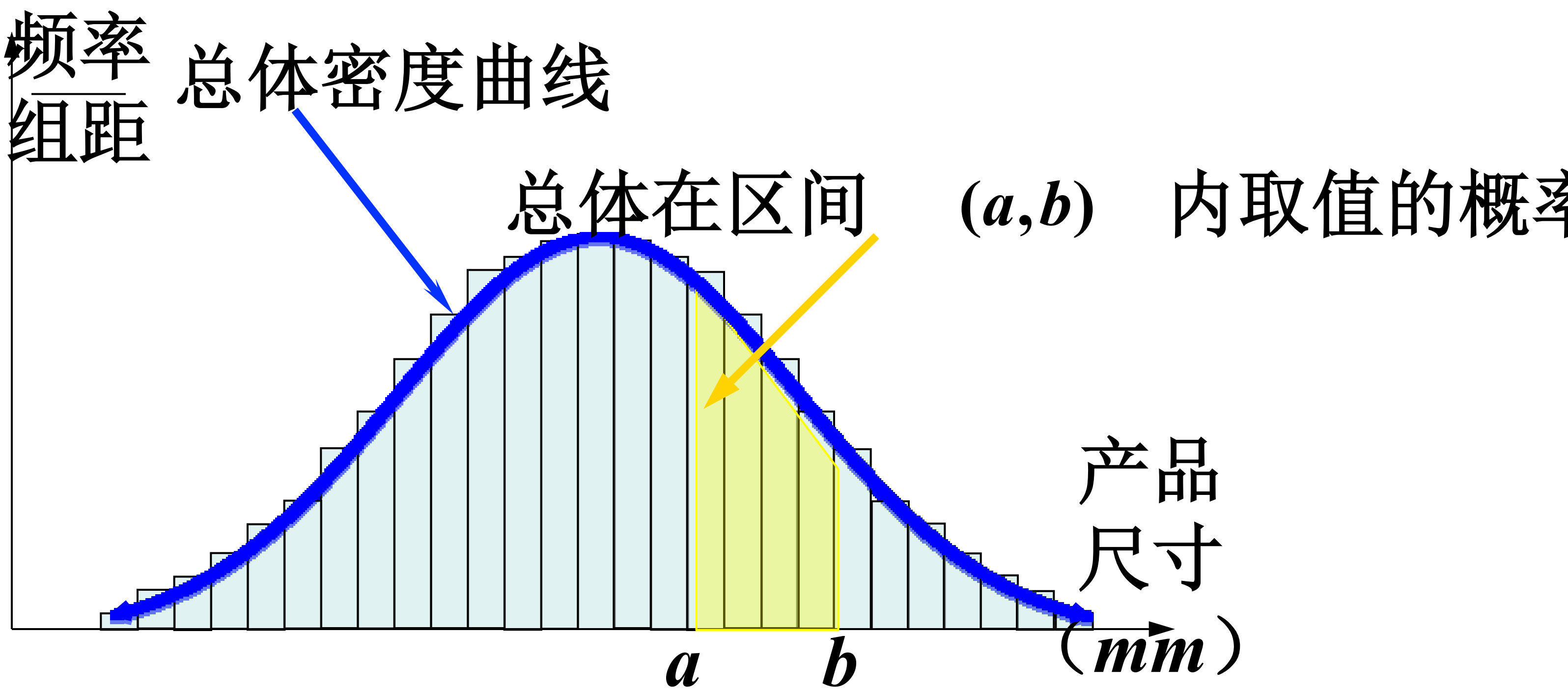


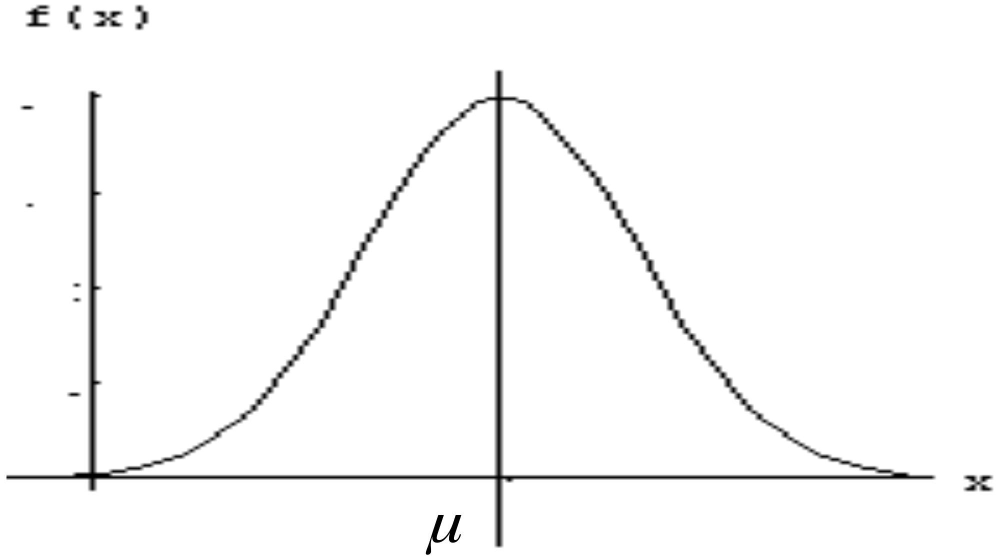
1. 样本的频率分布与总体分布之间的关系 .

2. 频率分布直方图 与总体密度曲线.

3. 总体密度曲线的形状特征.

中间高，两头低





若随机变量X的概率密度函数为

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

其中 μ 、σ为实数， σ>0 ， 则称X服从参数为μ ,σ^2

的正态分布(记为N(μ, σ^2))，可表为**X～N(μ, σ^2).**

二、正态分布曲线

密度函数

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

- 简记 $N(\mu,\sigma^2)$

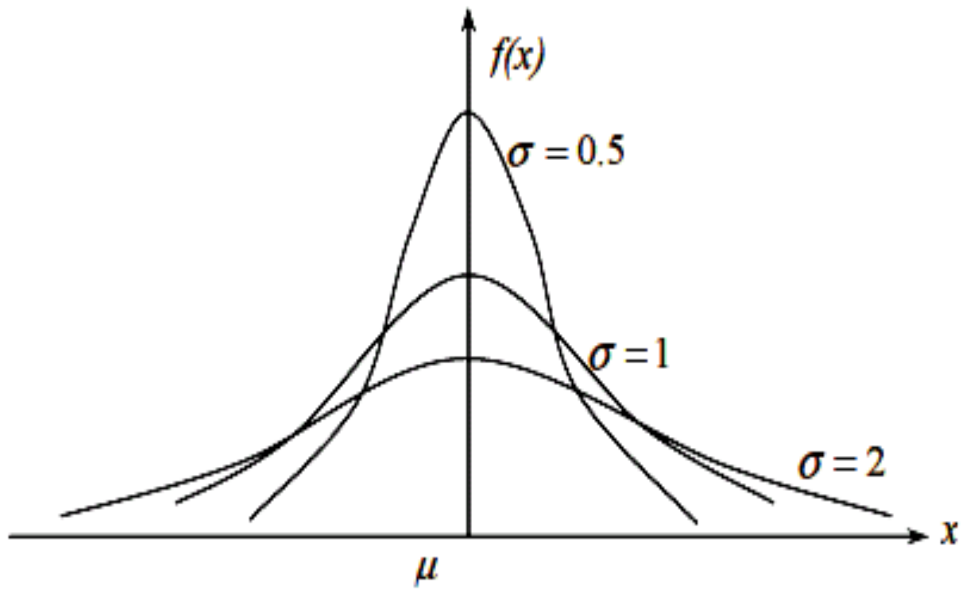


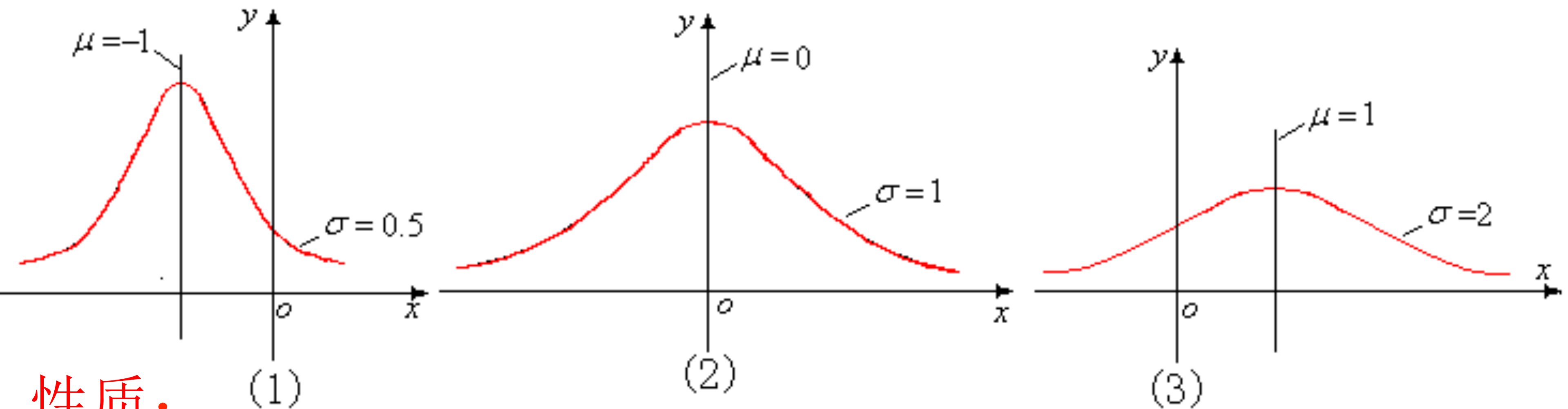
图 4-47　σ 变而 μ 不变

- **特征：**
- **单峰、对称；**

μ决定分布的中心位置，σ决定峰高和宽窄。

（3）正态曲线的性质

观察：

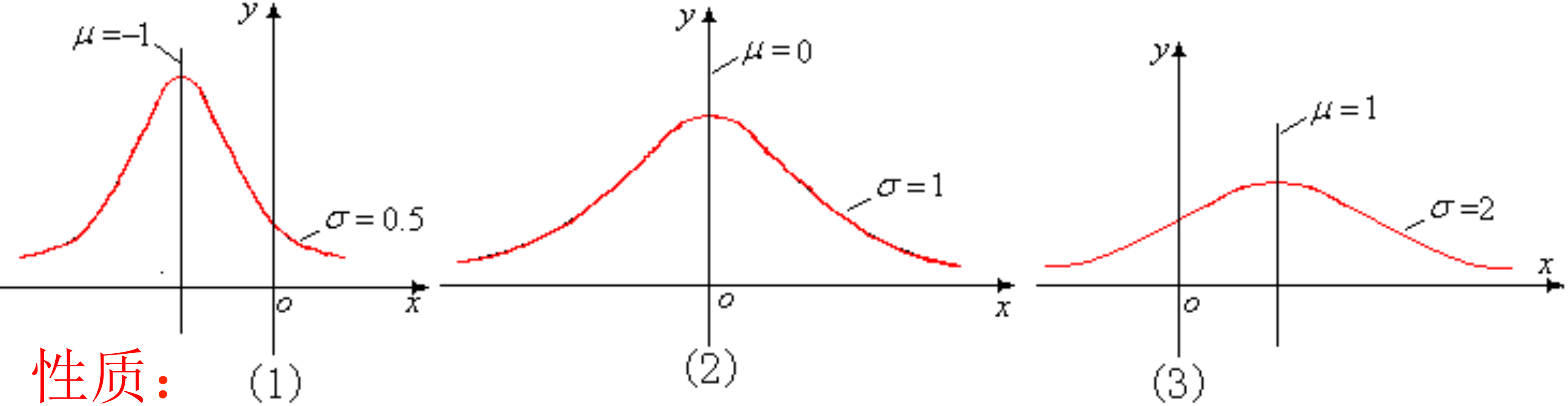


性质：

①曲线在x轴的上方，与x轴不相交；

②曲线关于直线$x=\mu$对称，且在$x=\mu$时位于最高点；

③当$x<\mu$时，曲线上升；当$x>\mu$时，曲线下降。并且当曲线向左、右两边无限延伸时，以x轴为渐进线，向它无限靠近。



性质：

④当μ一定时，曲线的形状由σ确定。

σ越大，曲线越“矮胖”，表示总体的分布越分散

σ越小，曲线越“瘦高”，表示总体的分布越集中

特征

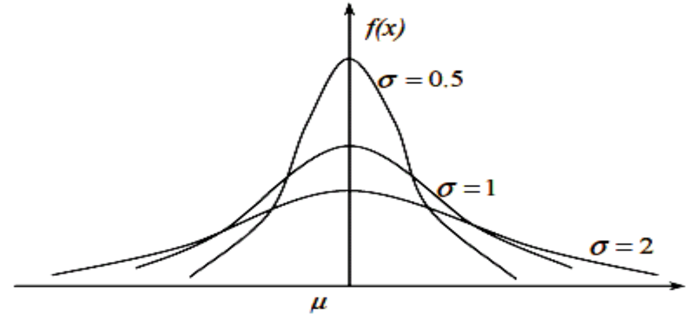
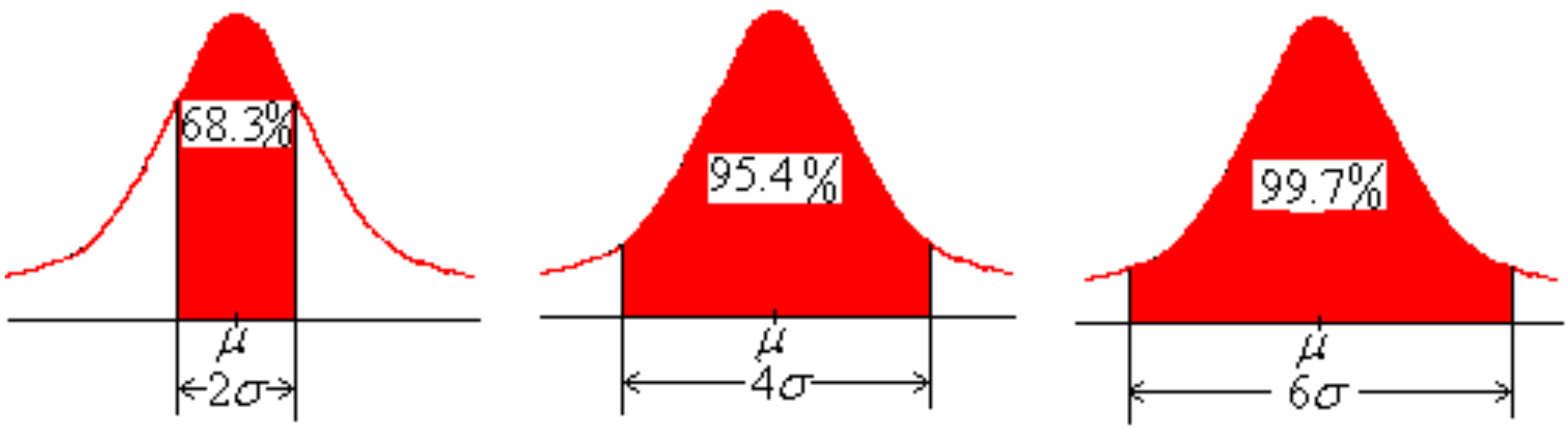


图 4-47 σ变而μ不变

- 正态曲线的中央点（平均数对应点）最高，然后逐渐向两侧下降，先向内弯，再向外弯，拐点位于$\pm 1\sigma$处。
- 正态曲线下的面积为1，过平均数的垂线将其等分
- 正态分布是一族分布，它随变量的平均数、标准差不同而有不同的形态。
- 正态曲线下，标准差与概率（面积）有一定数量关系。如：$\pm 1\sigma$间，包含了总面积的68.26%；$\pm 1.96\sigma$间，包含了总面积的95%；$\pm 2.58\sigma$间，包含了总面积的99%。（对照切比雪夫定理）

上述计算结果可用下表和图来表示：

区间	取值概率
$(\mu-\sigma,\mu+\sigma)$	68.26%
$(\mu-2\sigma,\mu+2\sigma)$	95.44%
$(\mu-3\sigma,\mu+3\sigma)$	99.72%



100名男生的身高分布

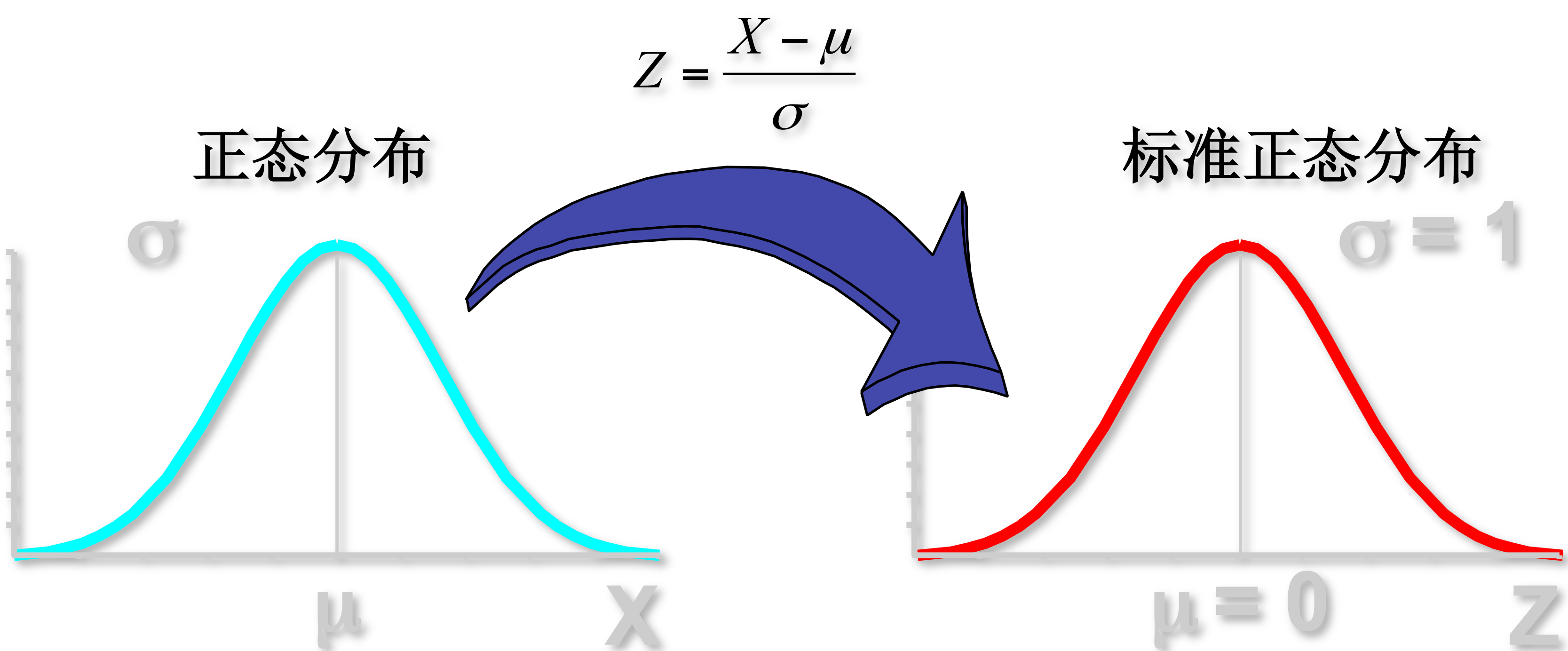
分布 x+-s	身高范围（cm）	实际分布人数	实际分布百分数（%）	理论分布（%）
X+-1s	168.69～176.71	67	67.00	68.27
X +-1.96s	164.84～180.56	95	95.00	95.00
X+-2.58s	162.35～183.05	99	99.00	99.00

$$\left(\mu = 172.7, \sigma = 4.01\right)$$

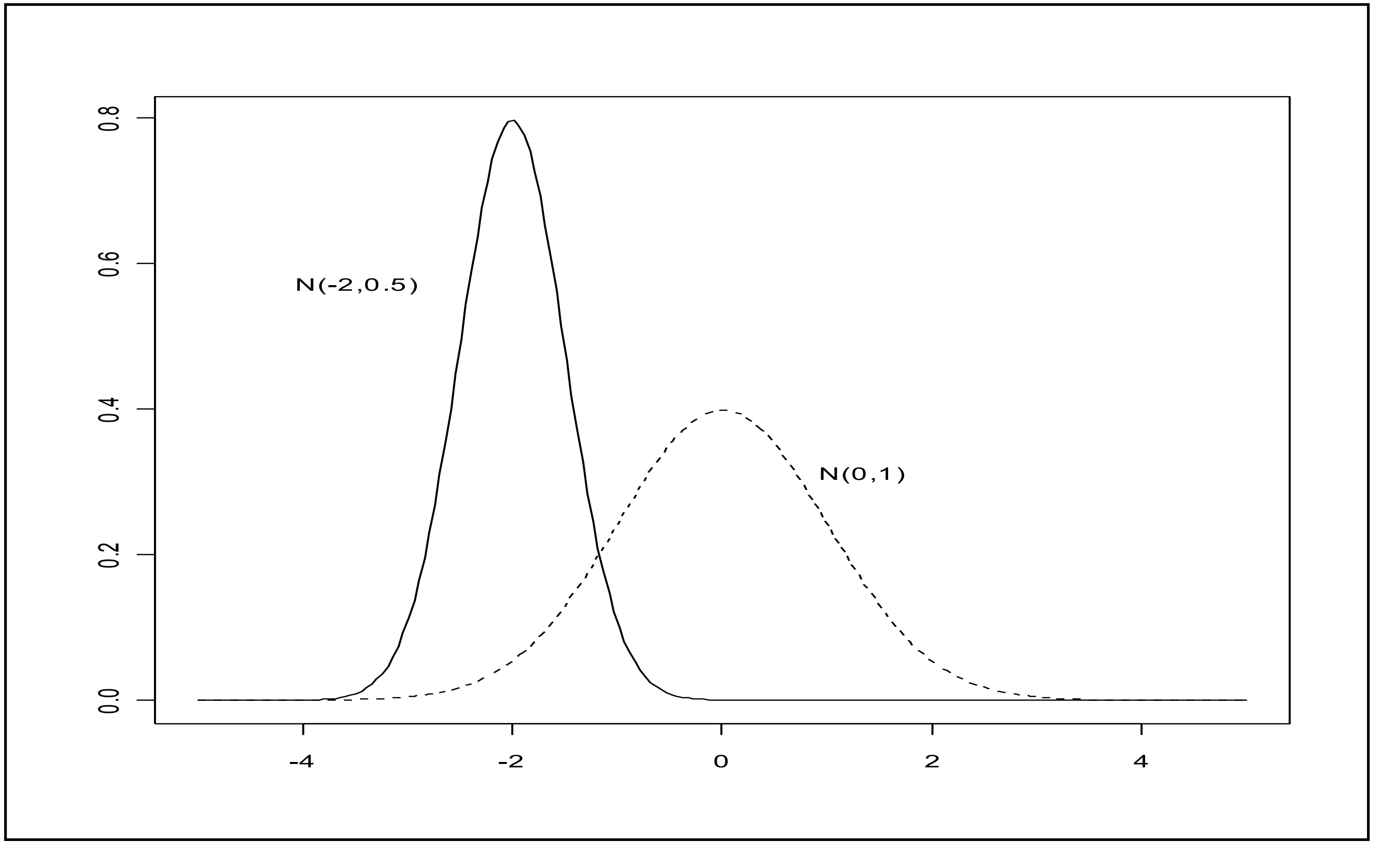
任何一个一般的正态分布都可以通过线性变换转化为标准正态分布.

设 $X \sim N(\mu, \sigma^2)$, 则

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$



两条正态分布的密度曲线。左边是N(-2,0.5)分布，右边是N(0, 1)分布



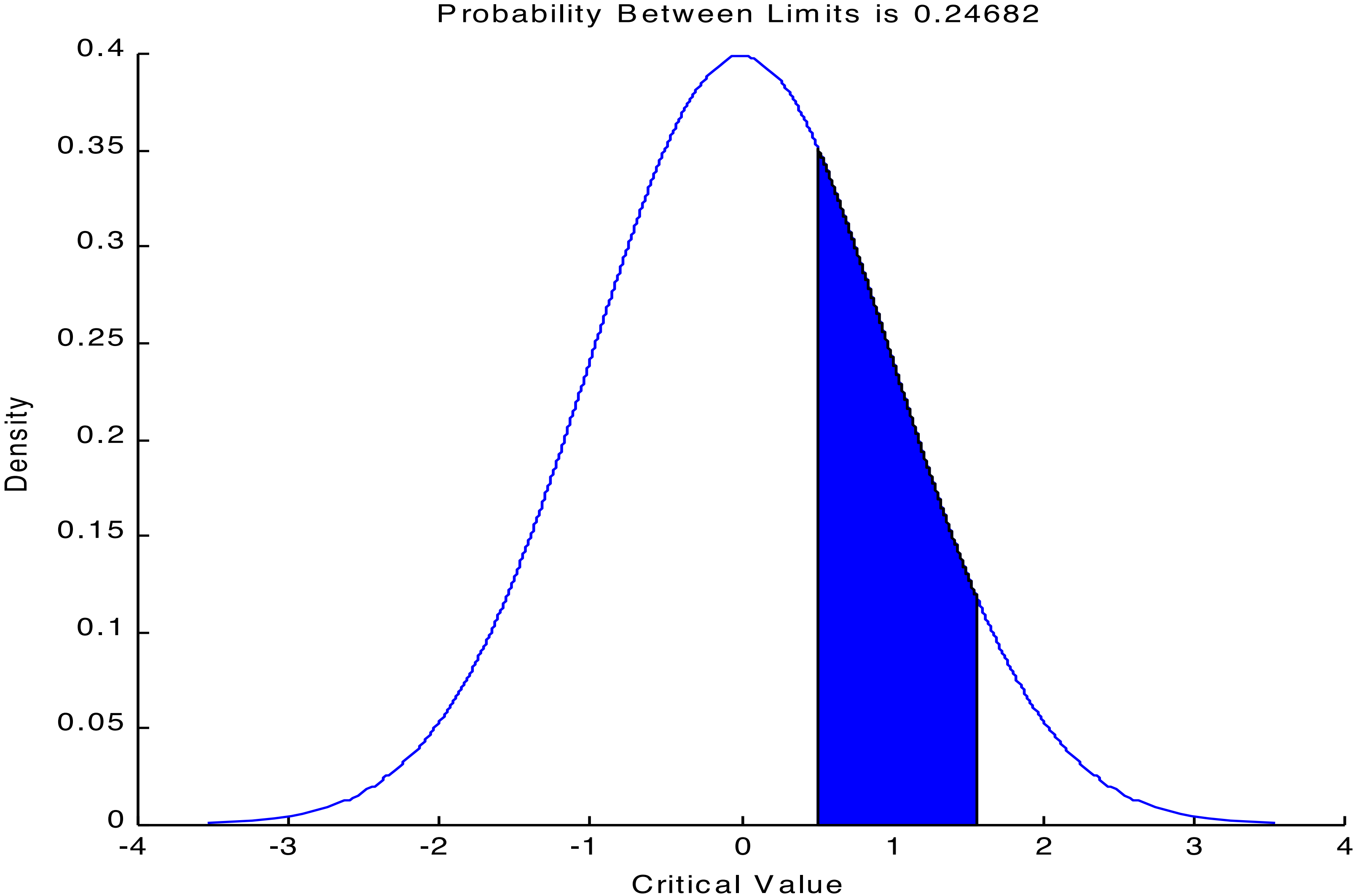
其密度函数表示为

$$\varphi(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}},-\infty<z<+\infty.$$

分布函数表示为

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

标准正态变量在区间(0.51, 1.57)中的概率



- 当然，和所有连续变量一样，正态变量落在某个区间的概率就等于在这个区间上，密度曲线下面的面积。
- 比如，标准正态分布变量落在区间**(0.51,1.57)**中的概率，就是在标准正态密度曲线下面在**0.51**和**1.57**之间的面积。
- 很容易得到这个面积等于**0.24682**；也就是说，标准正态变量在区间**(0.51,1.57)**中的概率等于**0.24682**。如果密度函数为$\phi(x)$，那么这个面积为积分

$$\int_{0.51}^{1.57} \phi(x)dx = 0.24682$$

四、正态分布表的编制和使用

（一）编制方法

（二）结构

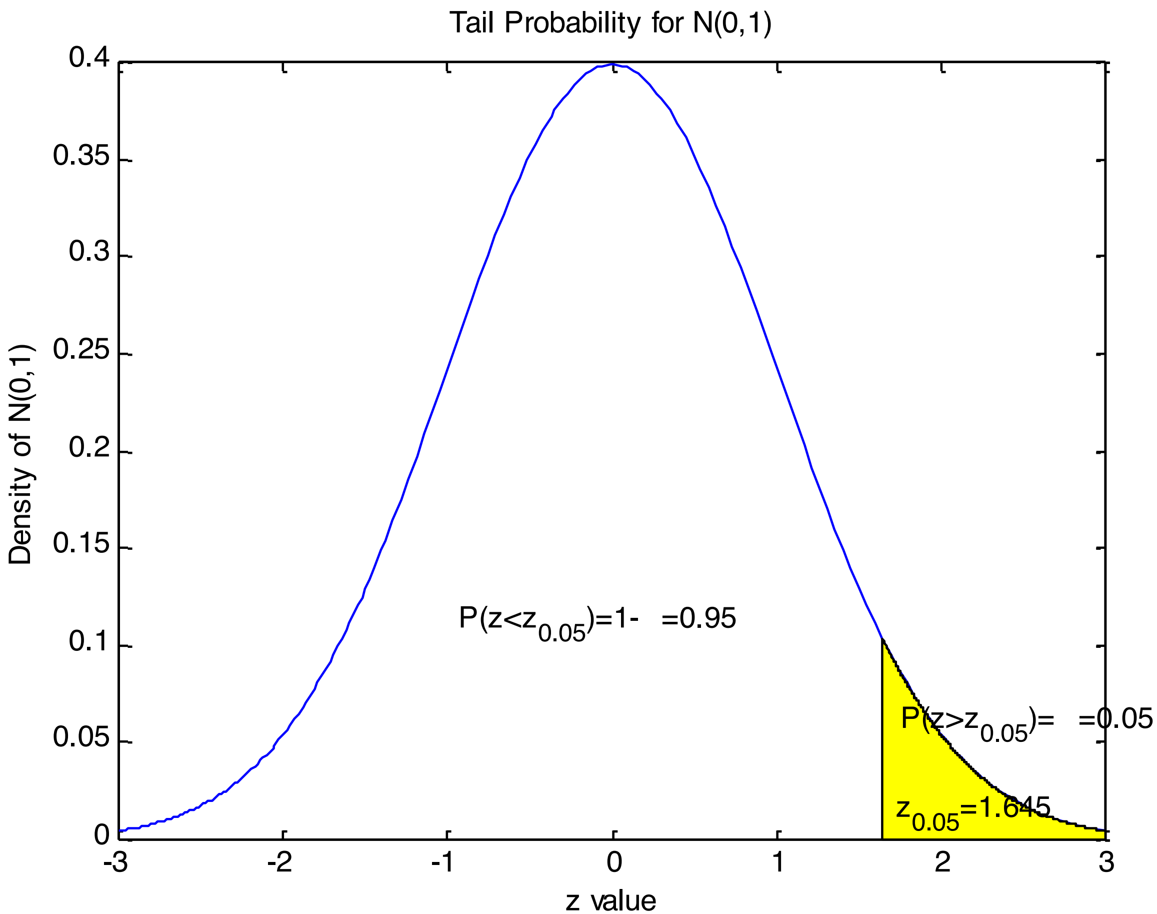
1栏：Z值；　2栏：密度函数Y值；

3栏：概率（面积）值；

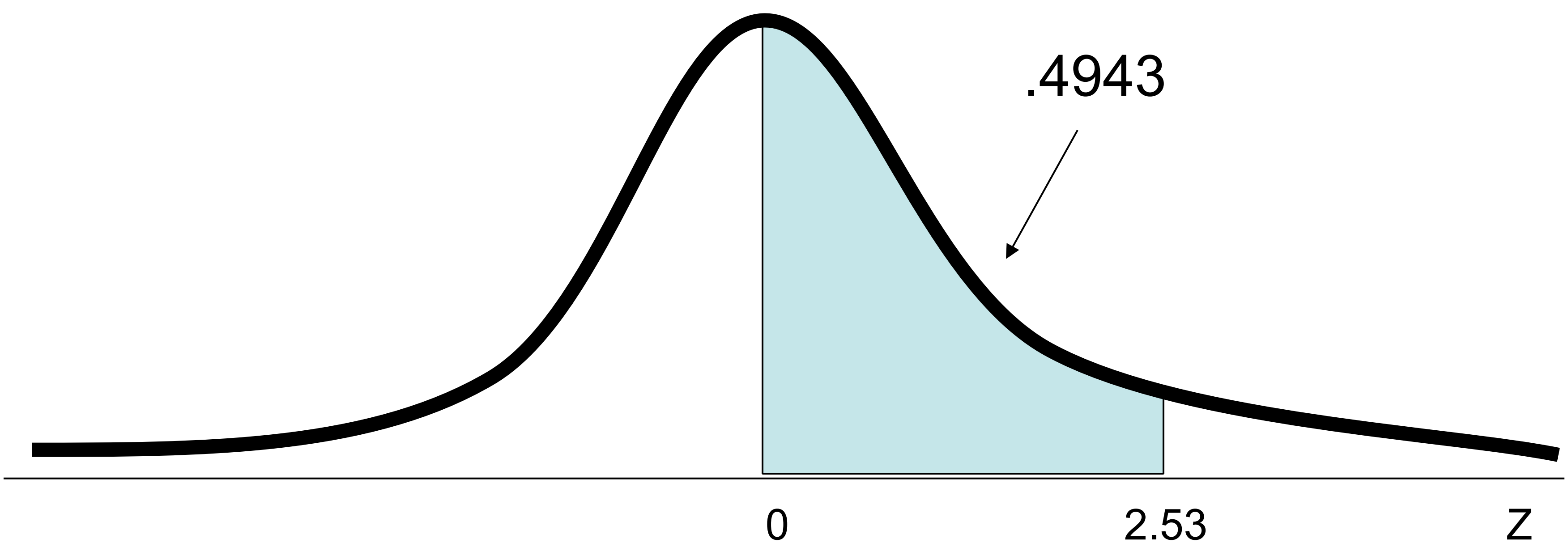
五、数据正态性的检验方法

六、应用

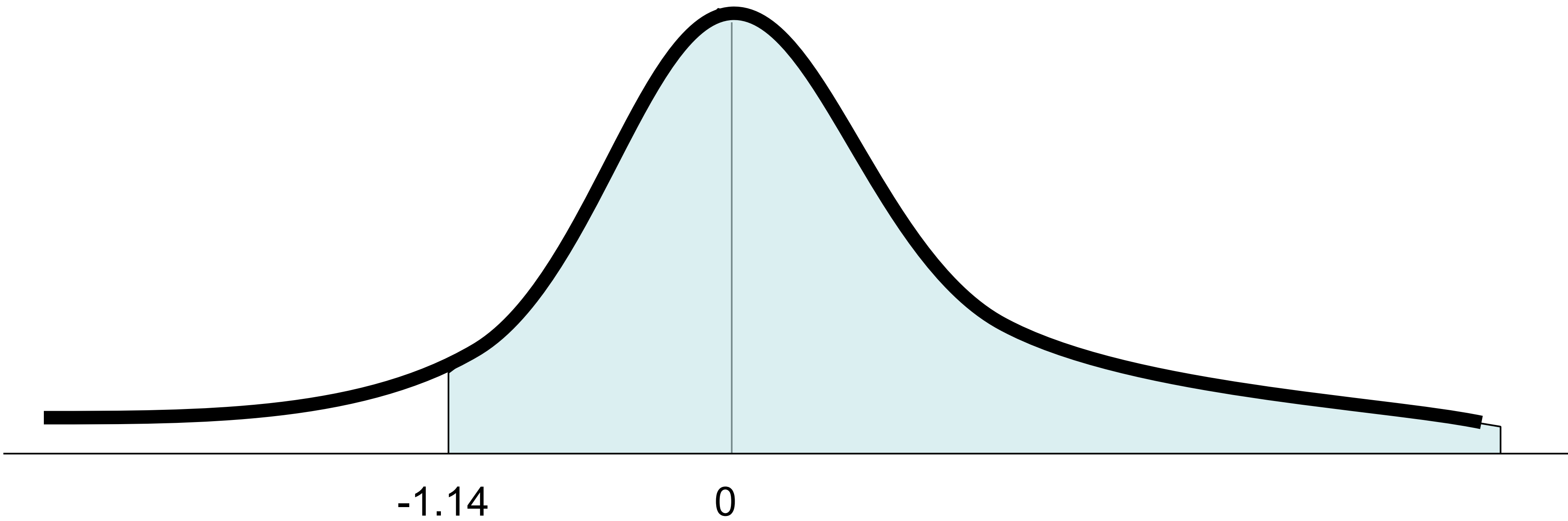
- 通常用z_α表示标准正态分布的α上侧分位数，即对于标准正态分布变量Z，有$P(Z>z_\alpha)=\alpha$。



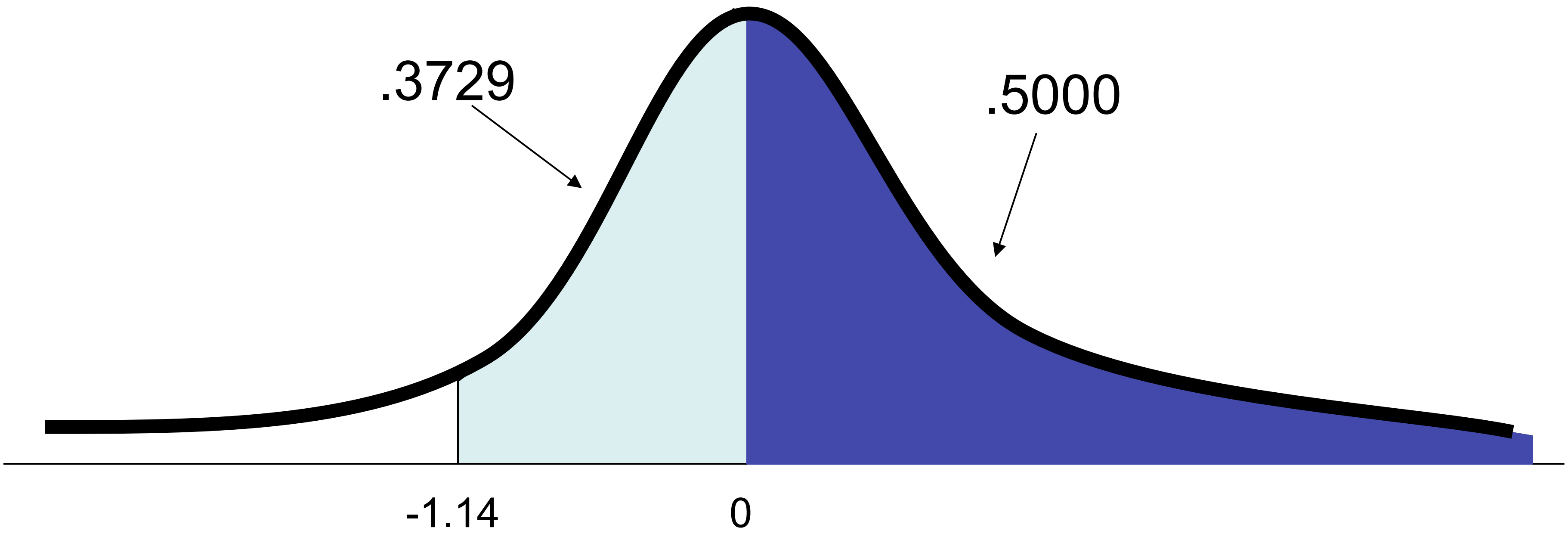
求 $P(0<z<2.53)=?$



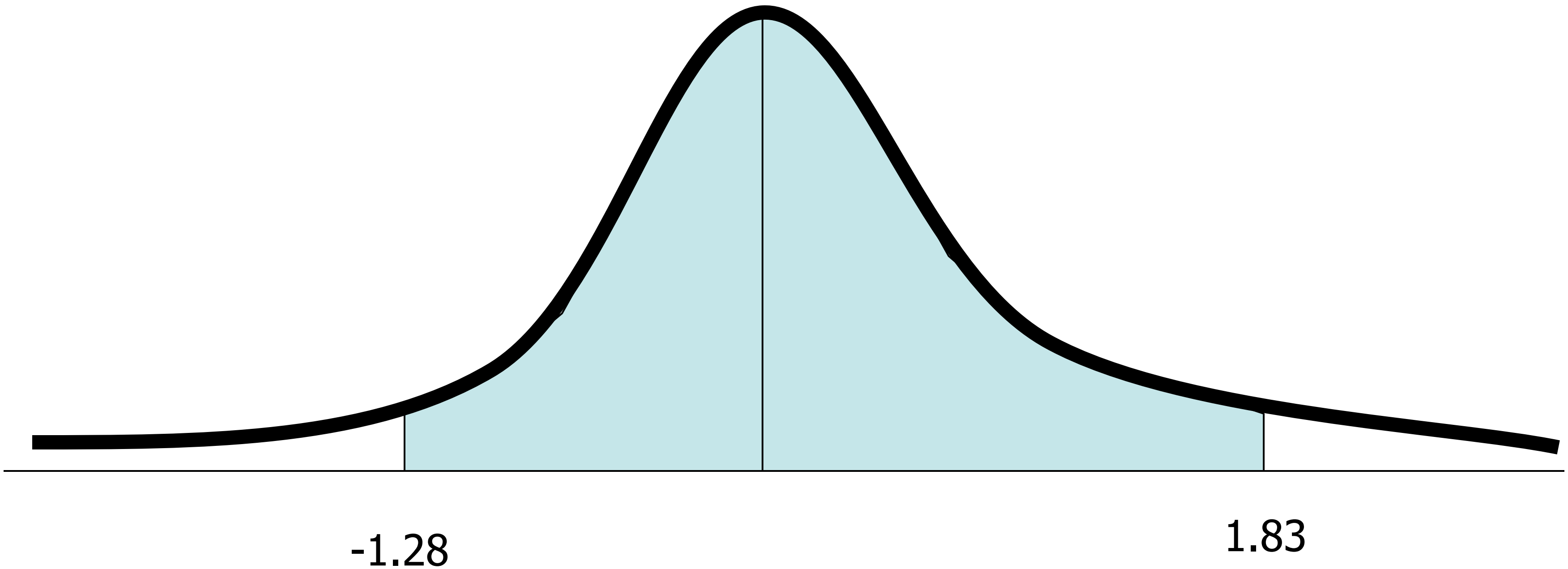
求 $P(z>-1.14)$.



$$P(z>-1.14)=.3729+.5000$$
$$=.8729$$

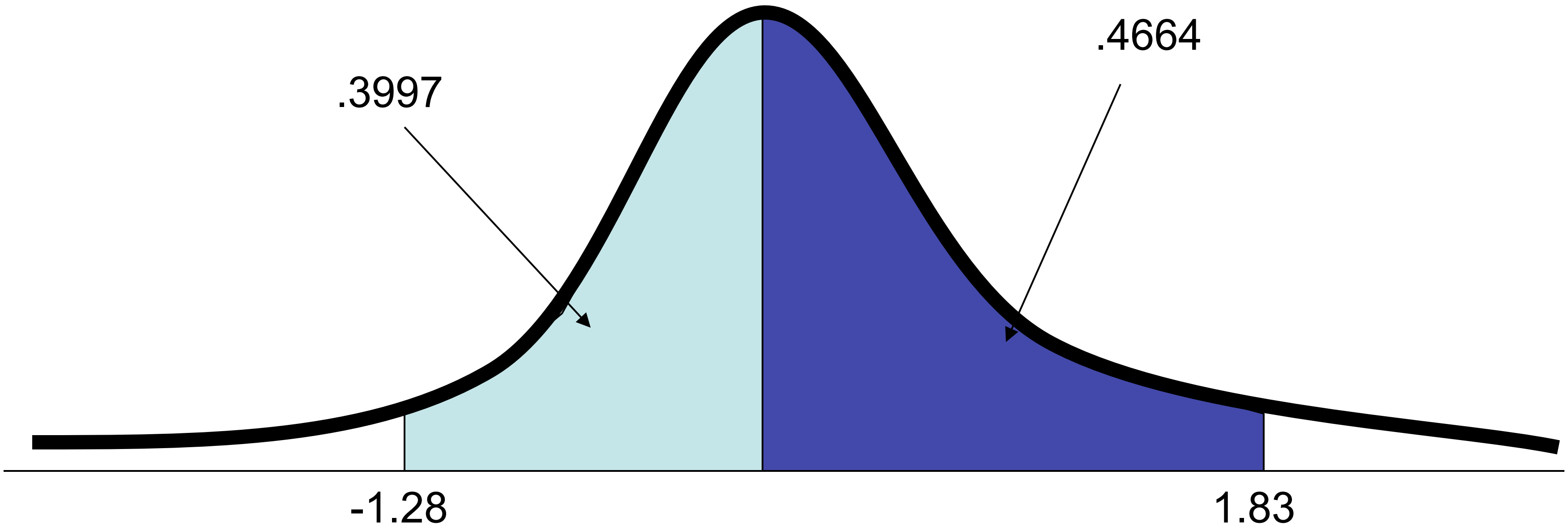


求 $P(-1.28<z<1.83)$.

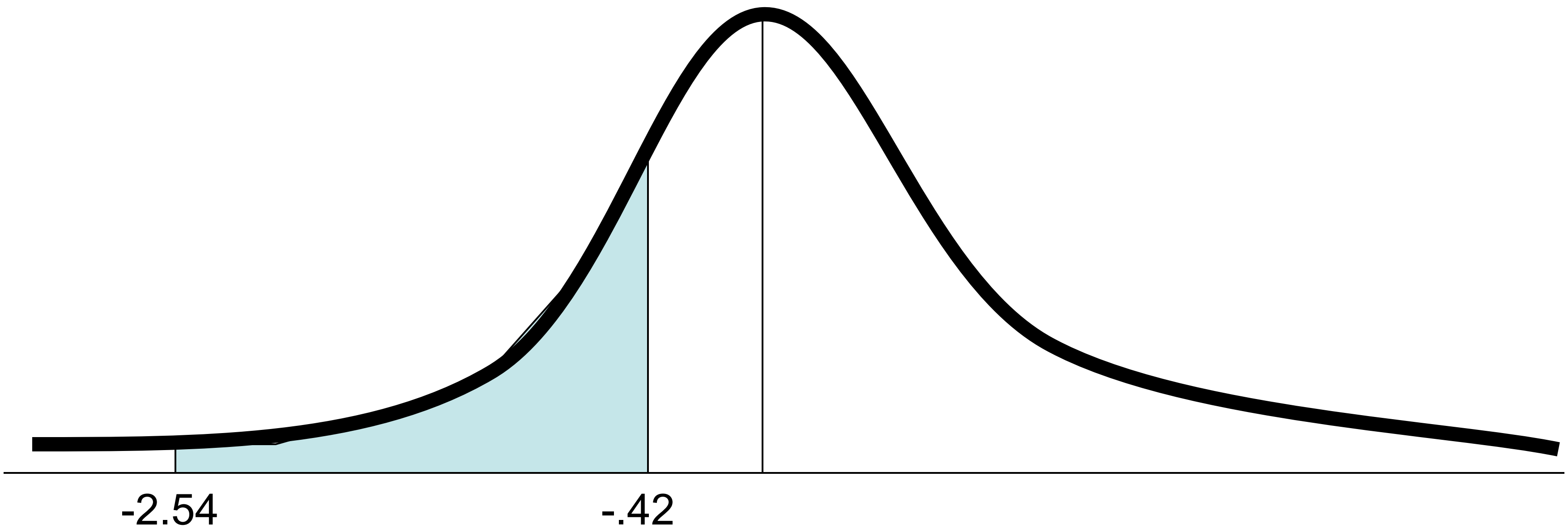


$$P(-1.28<z<1.83)=.3997+.4664$$

$$=.8661$$



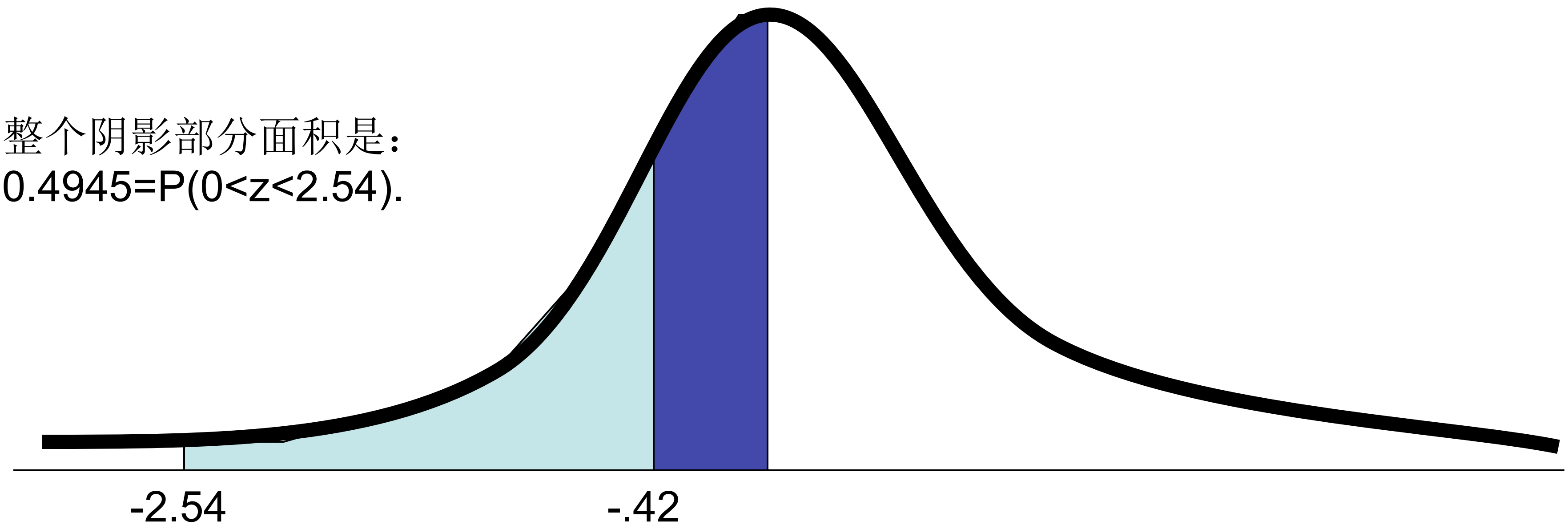
求 $P(-2.54<z<-.42)$.



$$P(-2.54<z<-.42) = .4945 - .1628$$
$$= .3317$$

黄色部分面积是：$0.1628=P(0<z<.42)$.

整个阴影部分面积是：
$0.4945=P(0<z<2.54)$.



例5.23 $X \sim N(0,1)$,求以下概率。

(1) $P(X<1.5)$ (2) $P(X>2)$

(3) $P(-1<X\leq 3)$ (4) $P(|X|\leq 2)$

例5.24、$X \sim N(100, 15^2)$, 求$P(x < 115)$

例5.23 $X \sim N(0,1)$,求以下概率。

(1) $P(X<1.5)$　　(2) $P(X>2)$

(3) $P(-1<X\le 3)$　　(4) $P(|X|\le 2)$

解:(1) $P(X<1.5)=0.5+P(0<X<1.5)=0.5+0.4332=0.9332$

(2) $P(X>2)=0.5-P(0<X\le 2)=0.5-0.4773=0.0227$

(3) $P(-1<X\le 3)=P(0<X\le 3)+P(-1<X\le 0)$

$=P(0<X\le 3)+P(0\le X<1)=0.4987+0.3413)=0.84$

(4) $P(|X|\le 2)=P(-2\le X\le 2)$

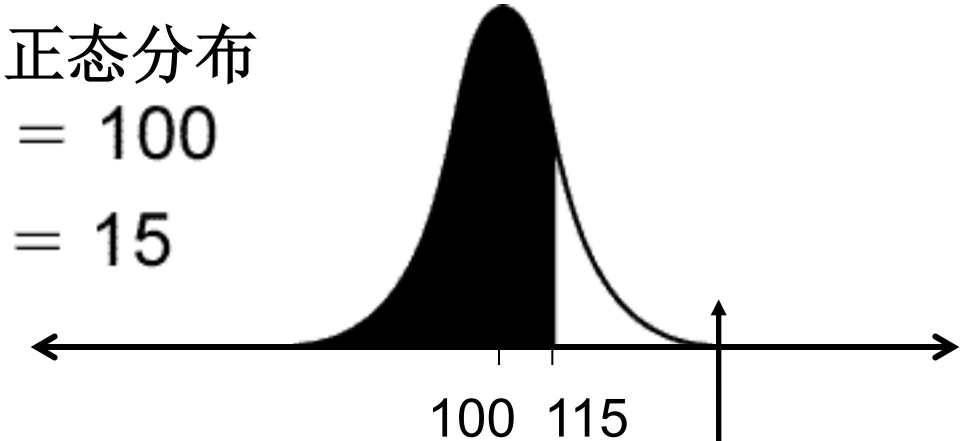
$=2P(0<X\le 2))=2\times 0.4773=0.9545$

例5.24、X～N(100, 15^2), 求P(x < 115)

正态分布

$\mu = 100$

$\sigma = 15$



P(x < 115).

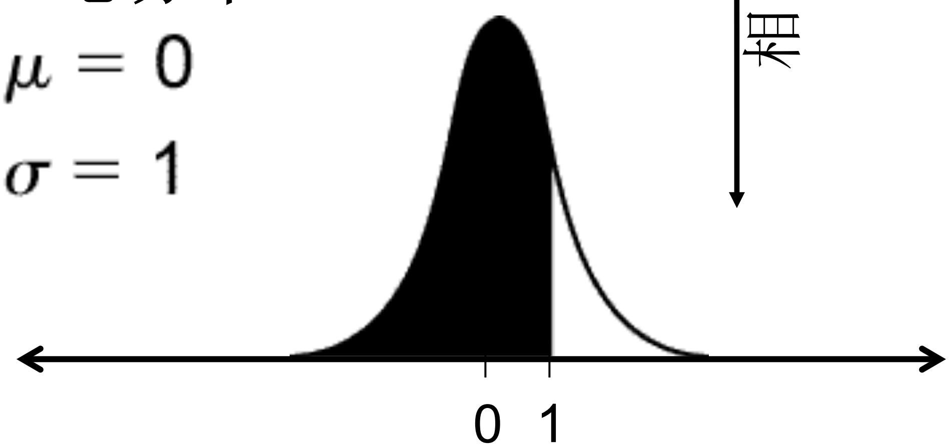
相同

P(z < 1).

标准正态分布

$\mu = 0$

$\sigma = 1$



相同

因为P(z < 1) = 0.8413, 所以, P(x <115) = 0.8413

例5.25 $X \sim N(5,10^2)$

求概率 **(1) $P(5<X<6.2)$**

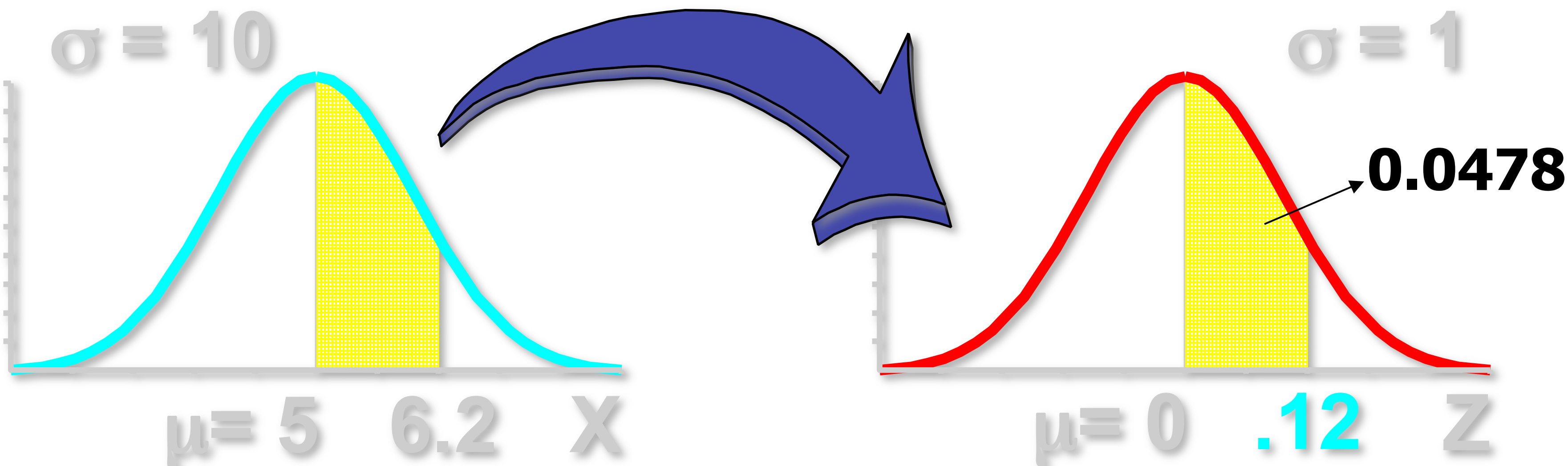
(2) $P(3.8<X<5)$

(3) $P(2.9<X<7.1)$

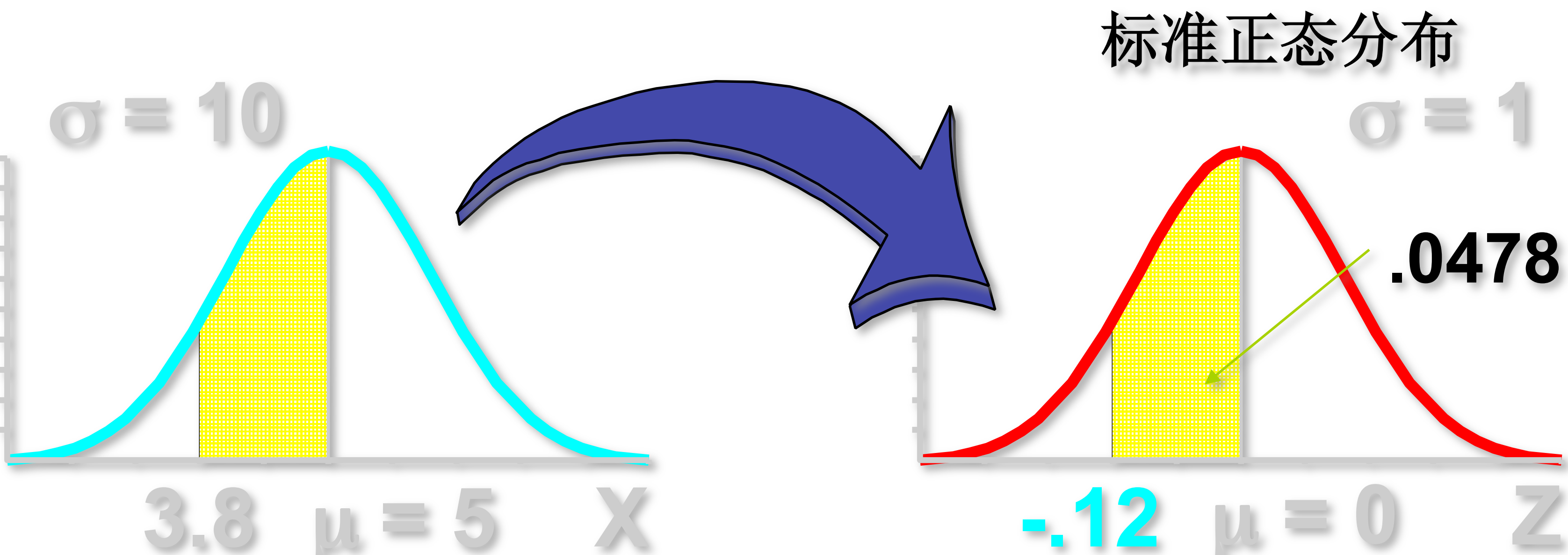
(4) $P(X>8)$

(5) $P(7.1<X<8)$

解: (1) $Z = \frac{X - \mu}{\sigma} = \frac{6.2 - 5}{10} = .12$



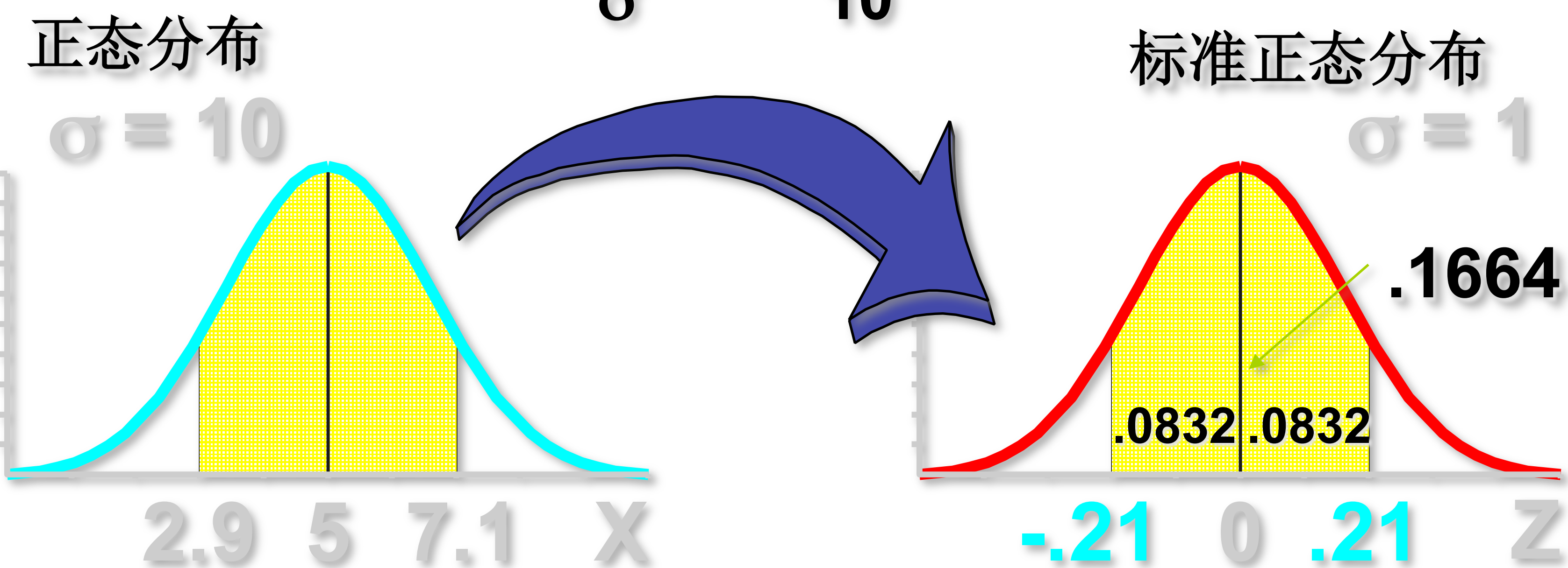
(2) $$Z = \frac{X - \mu}{\sigma} = \frac{3.8 - 5}{10} = -.12$$



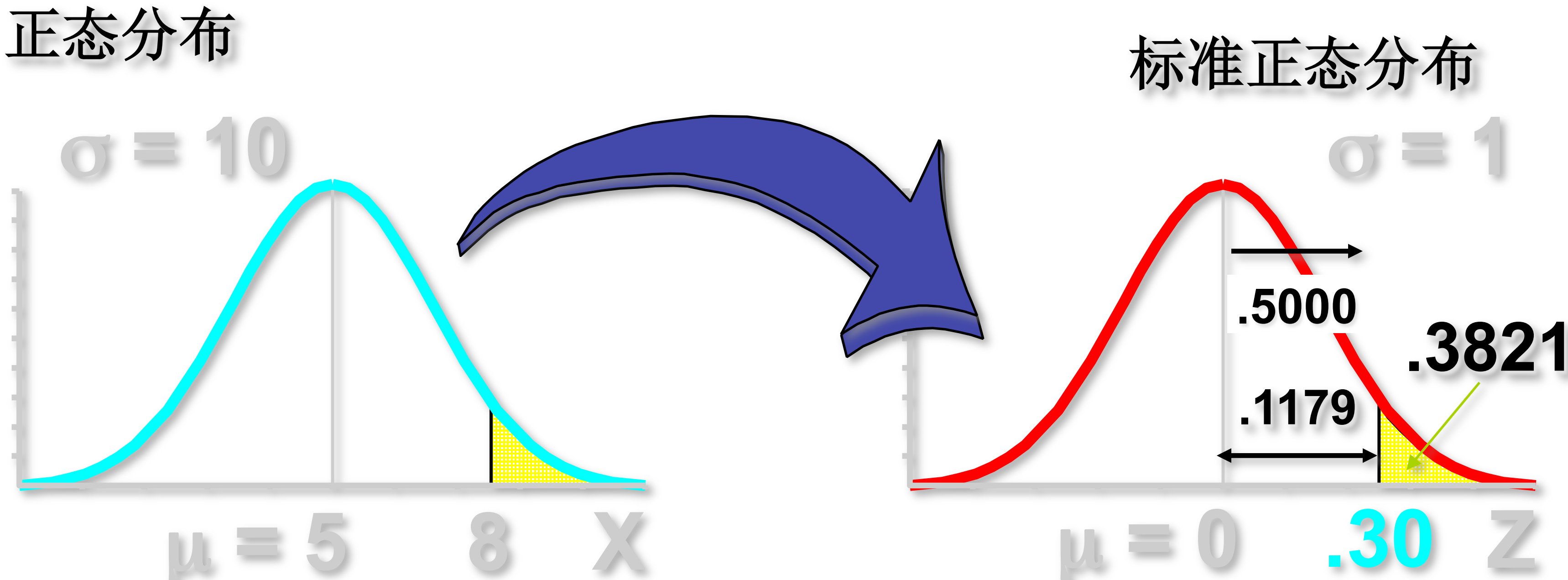
(3)

$$Z=\frac{X-\mu}{\sigma}=\frac{2.9-5}{10}=-.21$$

$$Z=\frac{X-\mu}{\sigma}=\frac{7.1-5}{10}=.21$$



(4) $Z = \dfrac{X - \mu}{\sigma} = \dfrac{8 - 5}{10} = .30$

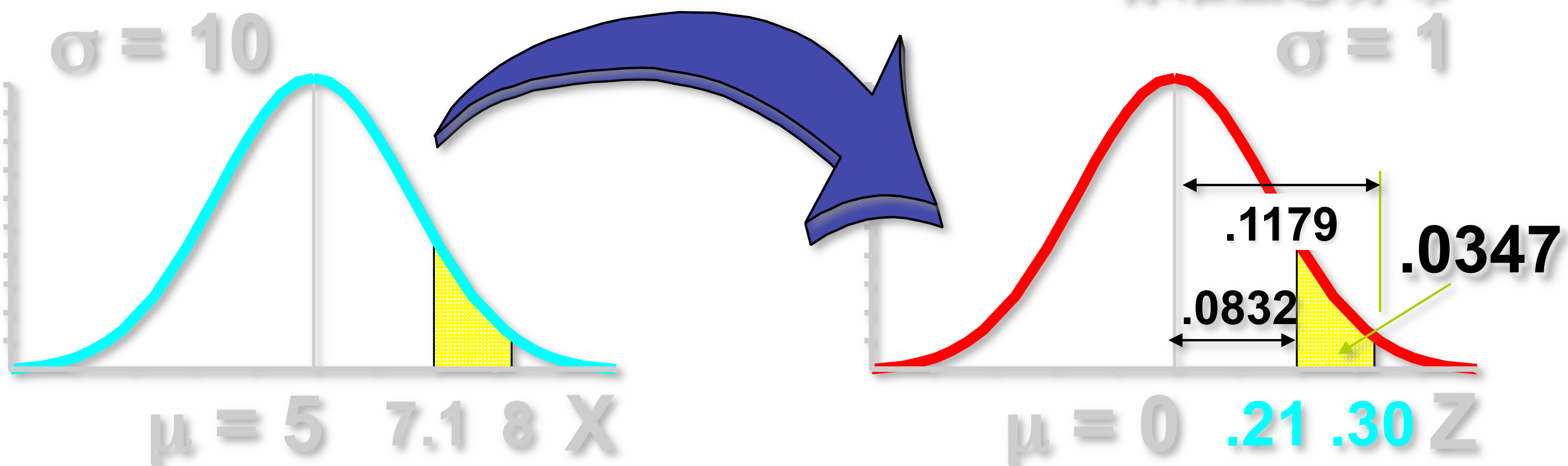


(5)

$$Z = \frac{X - \mu}{\sigma} = \frac{7.1 - 5}{10} = .21$$

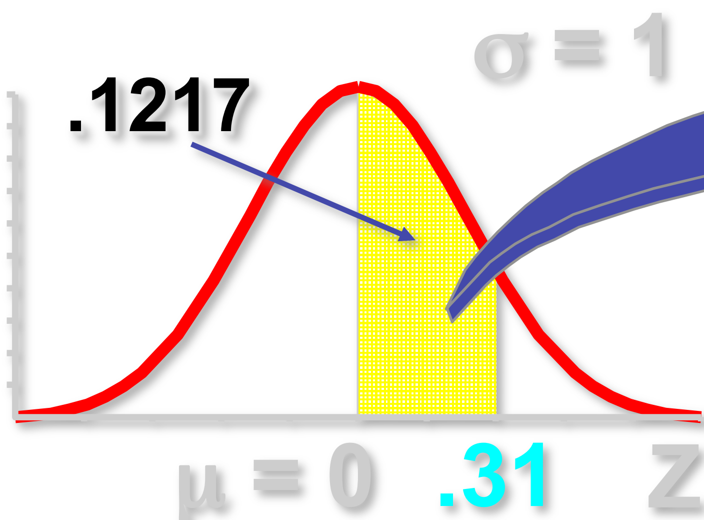
$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 5}{10} = .30$$

标准正态分布



例5.26 已知概率求Z分数

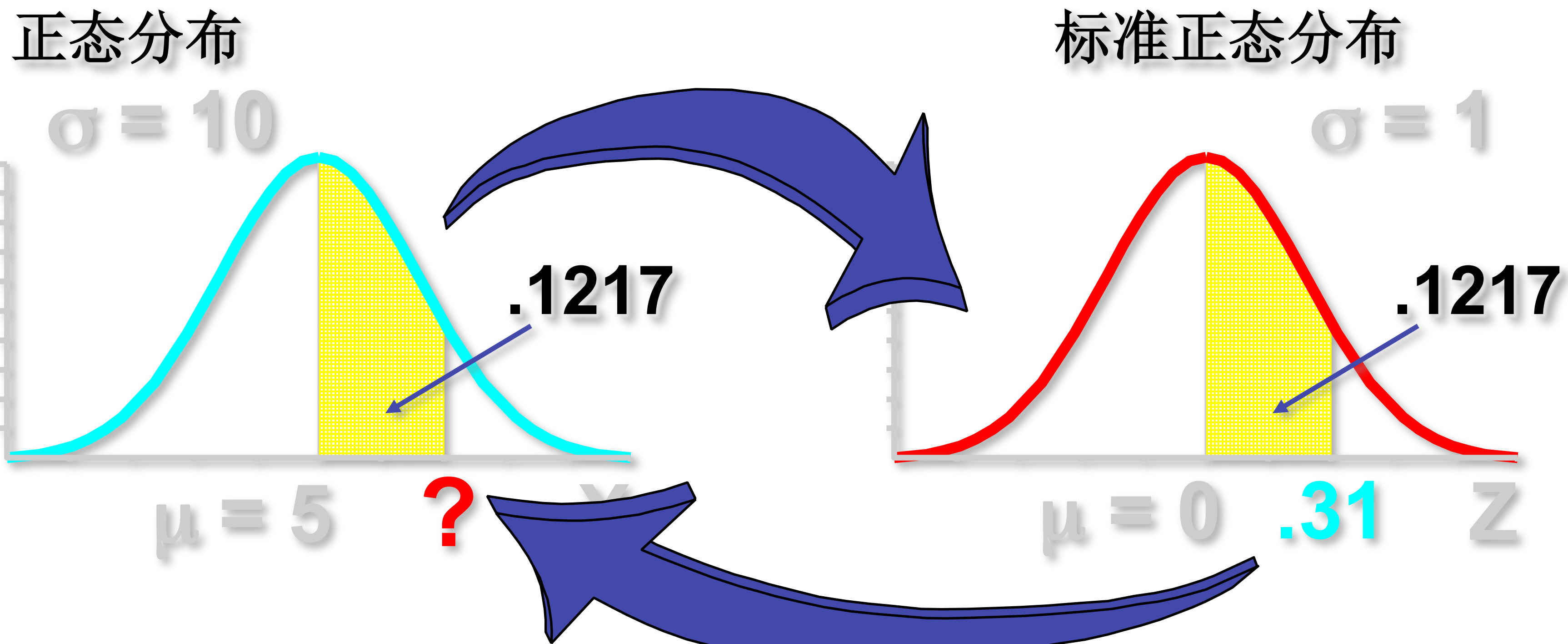
What is Z given P(Z) = .1217?



标准正态分布表（部分）

Z	.00	.01	0.2
[illegible]	000	.0040	.0080
0.[illegible]	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.1179	**.1217**	.1255

例5.27 已知概率求原始分数



$$X = \mu + Z \cdot \sigma = 5 + (.31)(10) = 8.1$$

例5.27 设X~N(5,3^2),求下列概率

(1) $P(X\le 10)$　　(2) $P(2<X<10)$

解:(1)

$$Z=\frac{X-5}{3}\sim N(0,1),$$

$$P(X\le 10)=P\left(Z\le\frac{10-5}{3}\right)=P(Z\le 1.67)$$

$$=P(Z\le 1.67)=P(-\infty<Z\le 0)+P(0<Z\le 1.67)=0.5+0.4525$$

$$=0.9525$$

(2)

$$P(2<X<10)=P\left(\frac{2-5}{3}<Z<\frac{10-5}{3}\right)$$

$$=P(-1<Z<1.67)=P(-1<Z\le 0)+P(0<Z<1.67)$$

$$=P(0\le Z<1)+P(0<Z<1.67)$$

$$=0.3413+0.4525=0.7938$$

例5.28、在某年高考的平均分数为500,标准差为100的正态总体中,某考生得到650分。设当年高考录取率为10％，问该生的成绩能否入围？

解：该生的标准分数为

Z＝(650-500)/100=1.5

查正态分布表,当Z=1.5时, p=0. 433, P(Z<1. 5)=0.5+0.433=0.933,即93. 3%, 从低分到高分的顺序中他处于93. 3%的位置， 那么, 从高分到低分的顺序中他应处于6. 7%的位置。因高考录取是从高分至低分顺序录取的，所以，该生处在录取率10％之内，他的成绩入了围。

例5.29 某市参加数学奥林匹克业余学校入学考试的人数为2800人，只录取学生150人，该次考试的平均分为75分，标准差为8，问录取分数应定为多少？

解：考试成绩服从正态分布，即

$X \sim N(75, 8^2)$, 通过Z转换， $Z=(X-75)/8$， $Z \sim N(0, 1)$。根据题意招生人数的概率为

$P(Z \geq Z_0)=150/2800=0.053$,

$P(0<Z< Z_0)=0.5-0.053 =0.447$

查标准正态分布表,

$Z=1.62$时， $p=0.4474$,　　　可取$Z_0=1.62$

故分数线应定为

$X_0=8Z_0+75=8\times1.62+75=88$(分)

课堂练习：

1、一本书排版后一校时出现错误处数X服从正态分布N(200,400), 求：

（1）出现错误处数不超过230的概率？

（2）出现错误处数在190～210之间的概率？

2、在美国人们第一次结婚的平均年龄是26岁。假设第一次结婚的年龄为正态分布，标准差为4年。问：

（1）一个人第一次结婚时的年龄小于23岁的概率多大?

（2）一个人第一次结婚时的年龄在20到30岁之间的概率多大?

（3）90％的人在什么年龄以前第一次结婚?

(4)3 σ准则

当X~N(0,1)时有

$P(|X|\le 1)=2P(0<X<1)=0.6826$,

$P(|X|\le 2)=2P(0<X<2)=0.9545$,

$P(|X|\le 3)=2P(0<X<3)=0.9973$,

X的取值几乎全部集中在[-3, 3]区间内, 超出这个范围的可能性仅占不到3%

当$X \sim N(\mu, \sigma^2)$时, 有

$P(|x-\mu| \le \sigma)=0.6826$,

$P(|x-\mu| \le 2\sigma)=0.9545$,

$P(|x-\mu| \le 3\sigma)=0.9973$

如果某个值在$|x-\mu| \le 3\sigma$之外, 可以判定为异常值.

例22 、$X \sim N(0.2, 0.05^2)$, 由3σ准则测量值应在0. 2−0. 05×3——0. 2+0. 05×3 之间， 即[0. 05， 0. 35]。有一个值为0. 367>0. 35， 所以为异常值, 应删除。