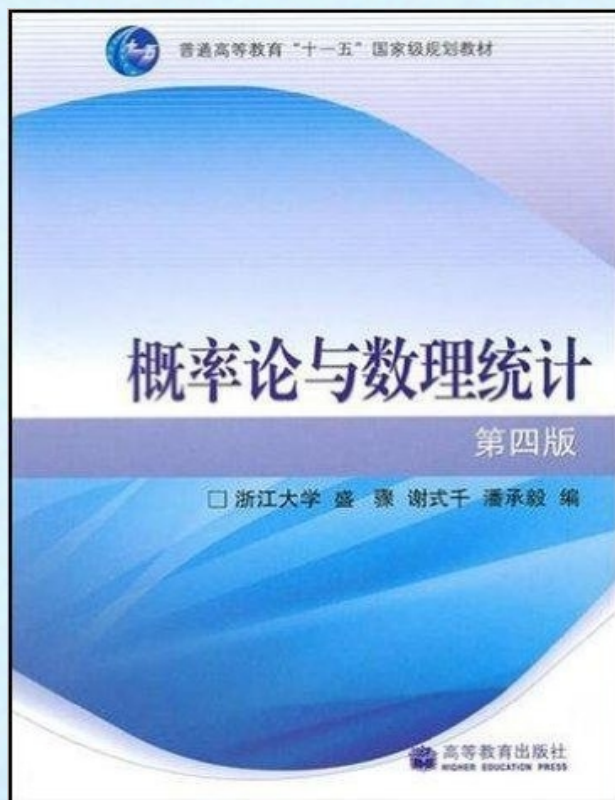


目 录



- 第一章 概率论的基本概念
- 第二章 随机变量及其分布
- 第三章 多维随机变量及其分布
- 第四章 随机变量的数字特征
- 第五章 大数定律及中心极限定理
- 第六章 样本及抽样分布
- 第七章 参数估计
- 第八章 假设检验
- 第九章 方差分析及回归分析

第四节 连续型随机变量及其概率密度

一、概率密度的概念与性质

二、常见连续型随机变量及其概率分布

三、小结

二、常见连续型随机变量及其概率分布

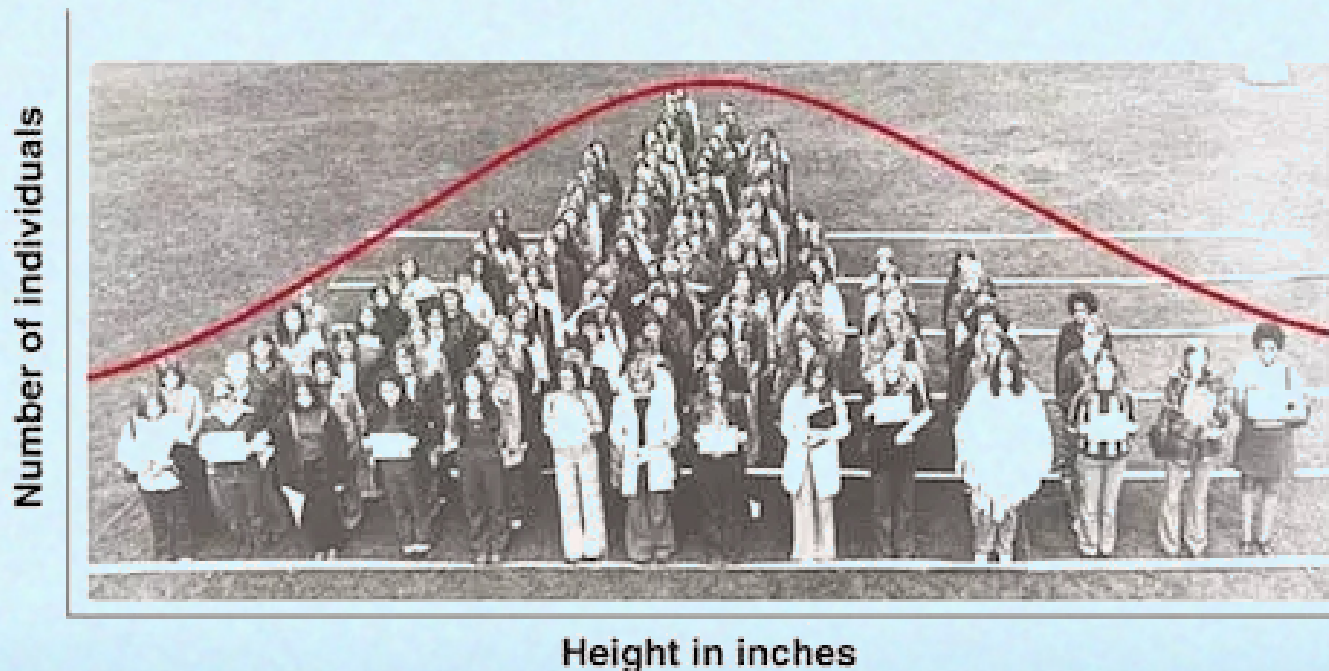
(三) 正态分布

1. 生活中的正态分布现象
2. 正态分布的定义
3. 正态曲线的性质
4. 正态分布的应用举例—— 3σ 原则

1. 生活中的正态分布现象

众所周知，自然界中存在着大量的正态分布，

现象一：人的身高



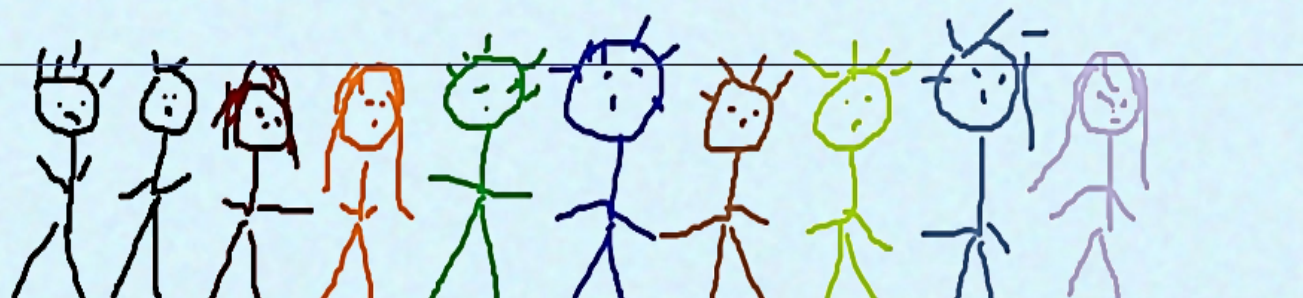
女士身高分布示例

1. 生活中的正态分布现象

现象二： 还记得上学拍合影吗？

宇宙最帅天团

平均身高线



去掉几个特别高的和几个特别矮的，大部分人的身高是差不多的。
不是吗？这也符合**正态分布**的规律。

2.4 连续型随机变量及其概率密度



中国最帅天团

1. 生活中的正态分布现象

现象三：假设在高中的时候，成绩大概是班级15名附近。每次考试结束的时候都会和同桌讨论答案。情况如下图所示：

咱们对一下选择题和判断题吧？大头



第N次模考	正确率	班级排名
1	65%	15
2	60%	18
3	75%	13
4	70%	14
5	70%	15
6	65%	16



好的，呆瓜

大家看，这个情况是不是很有趣，每次考试，不管你对完答案心情如何，结果都差不多，成绩并没有大幅度变化。（当然会有个别猛人逆袭，不过逆袭的都是个别人，这也符合正态分布的规律。）

2. 正态分布的概率密度函数

若连续型随机变量 X 的概率密度为

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

其中 $\mu, \sigma (\sigma > 0)$ 为常数, 则称 X 服从参数为 μ, σ 的

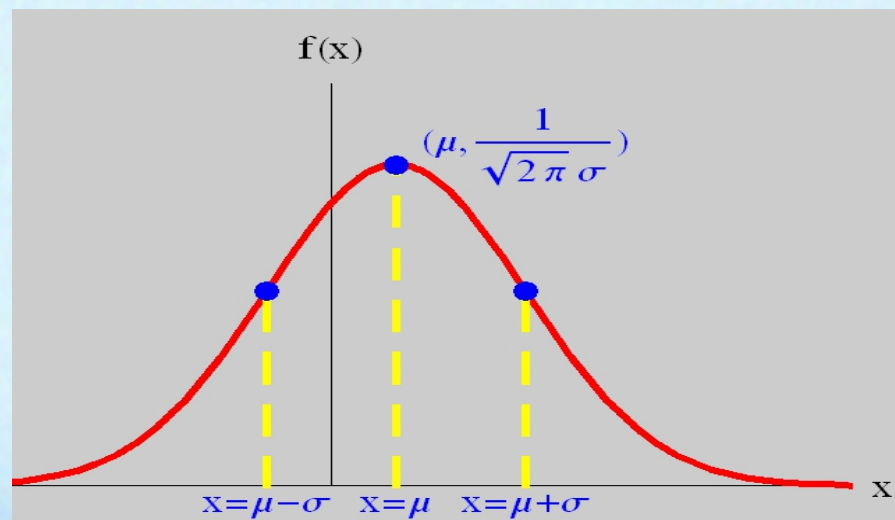
正态分布或高斯分布. 记为 $X \sim N(\mu, \sigma^2)$.

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

根据连续型随机变量概率密度的性质，有：

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x) dx = 1.$$

$f(x)$ 的图形如图所示



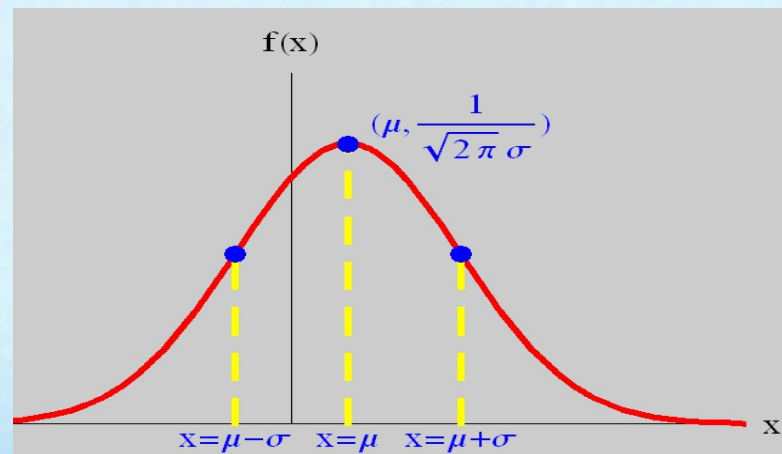
3. 正态分布的性质:

1° 曲线关于 $x = \mu$ 对称 .

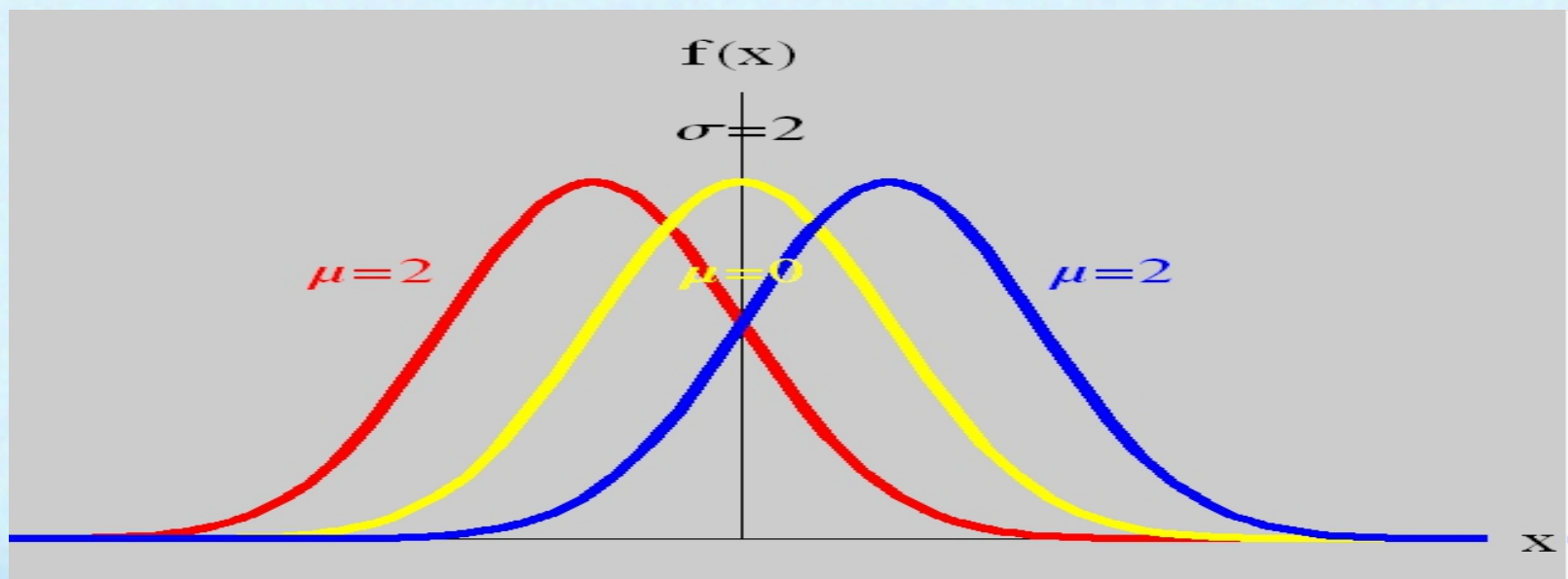
2° 当 $x = \mu$ 时取到最大值 $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$.

3° 在 $x = \mu \pm \sigma$ 处曲线有拐点 ;

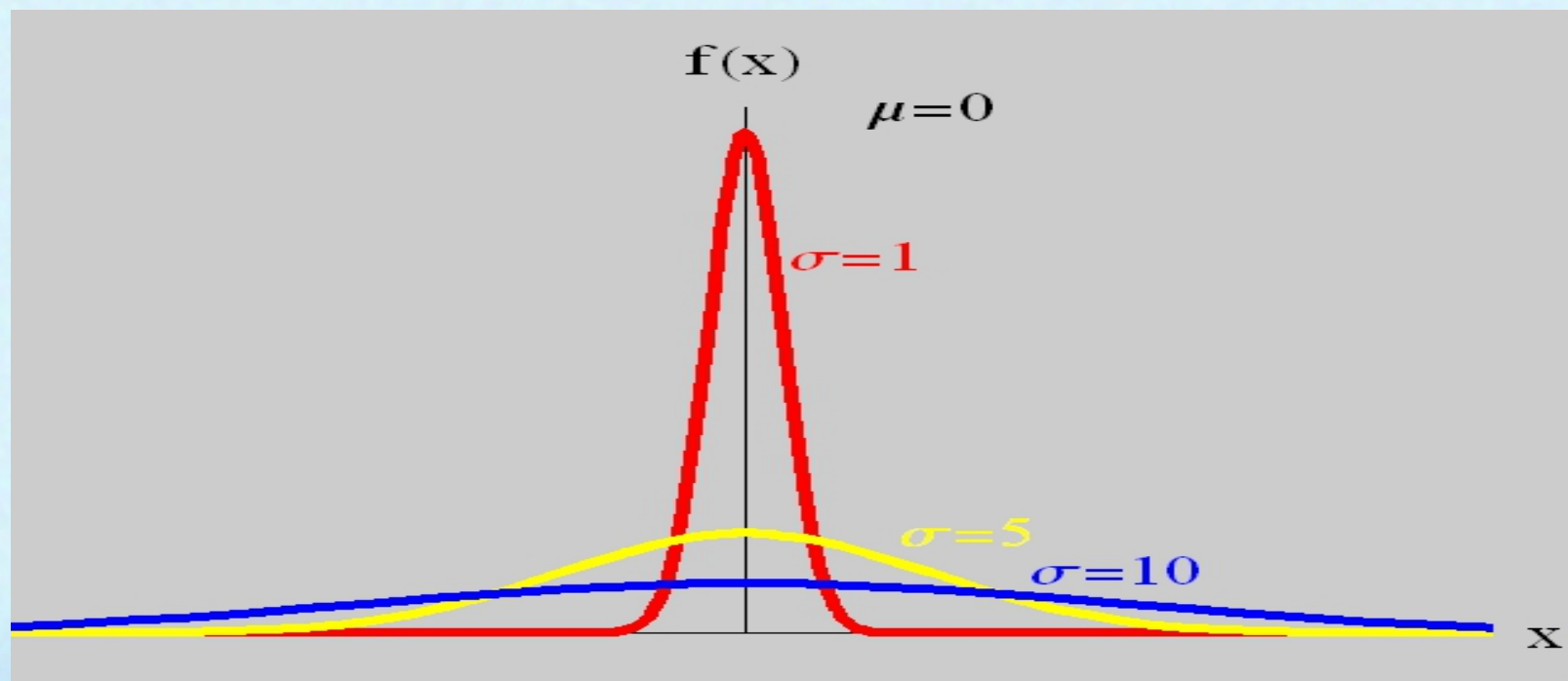
4° 曲线以 x 轴为渐近线 ;



5° 如果固定 σ ，改变 μ 的值，则图形沿着 Ox 轴平移，而不改变其形状，可见正态分布的概率密度曲线 $y = f(x)$ 的位置完全由参数 μ 所确定。 μ 称为位置参数。

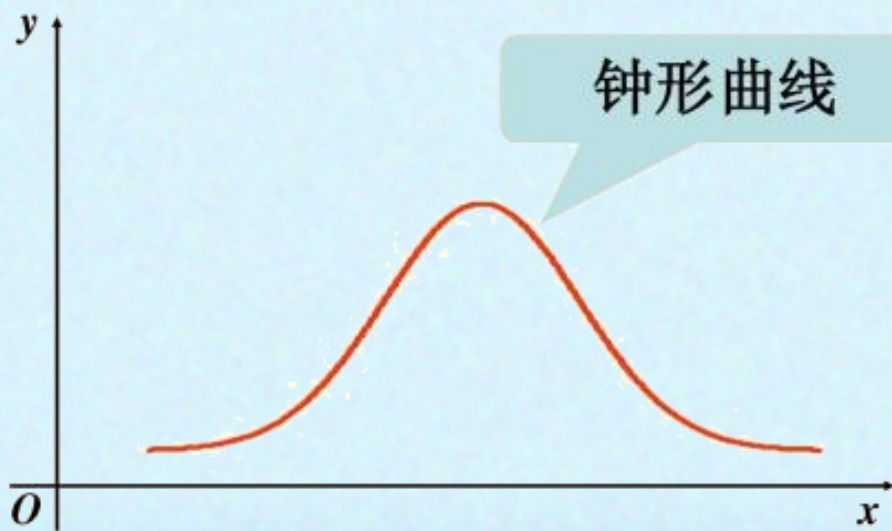


6°当固定 μ , 改变 σ 的大小时, $f(x)$ 图形的对称轴不变, 而形状在改变, σ 越小, 图形越高越尖, σ 越大, 图形越矮越平. σ 称为形状参数,



正态分布的图形— 钟形分布

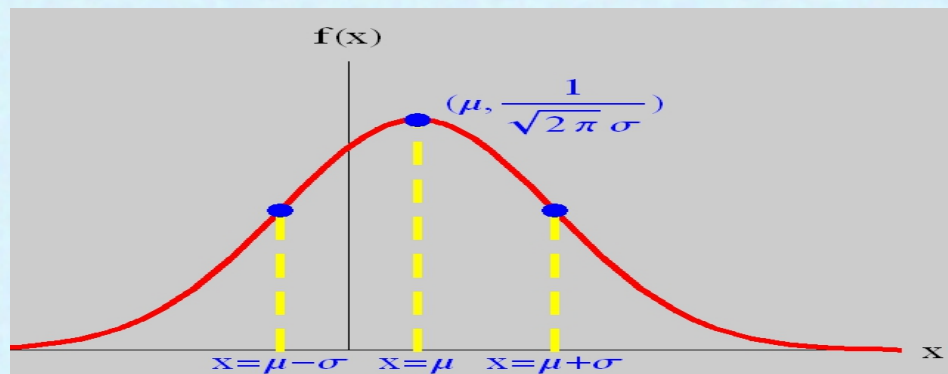
钟形分布的特征是“两头小，中间大”，即靠近中间的变量值分布的次数多，靠近两边的变量值分布的次数少，其曲线图宛如一口钟，如图所示。



$$X \sim N(\mu, \sigma^2).$$

X 的概率密度为

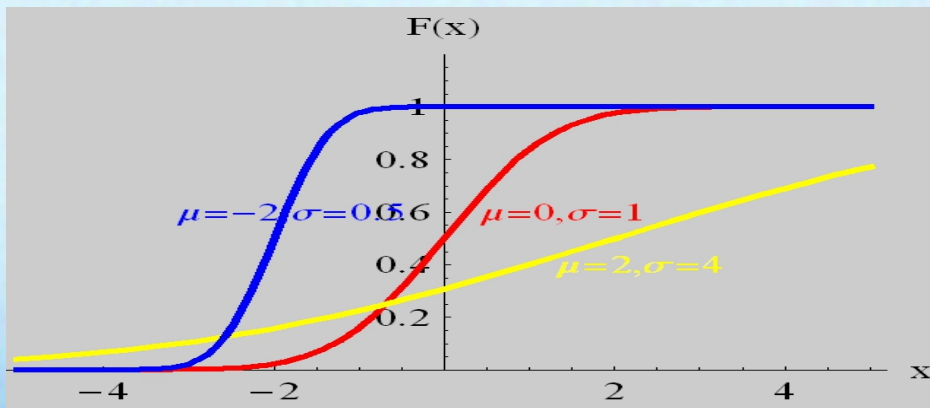
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$



X 的概率分布函数为

$$F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t)dt$$

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



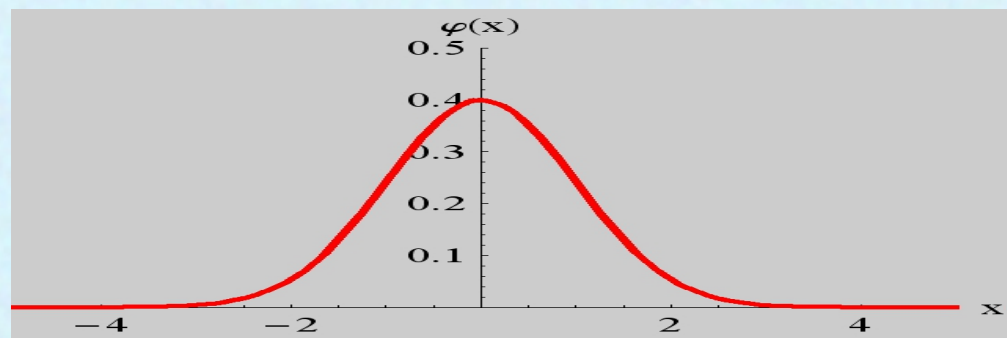
$$X \sim N(\mu, \sigma^2).$$

当 $\mu = 0$, $\sigma = 1$ 时称 X 服从标准正态分布 .

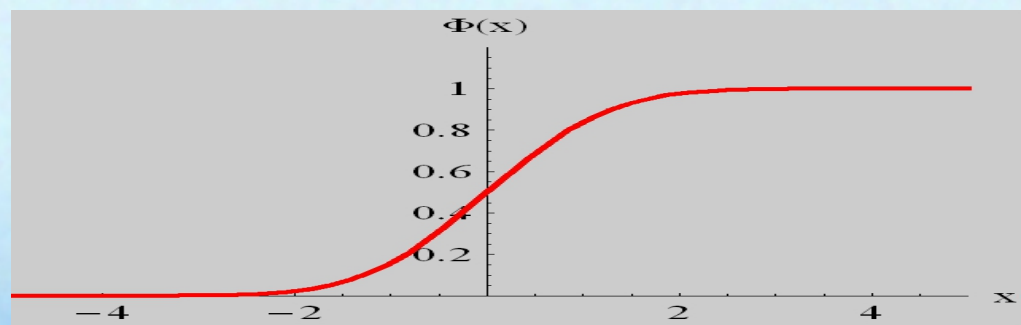
$$\text{记: } X \sim N(0,1)$$

其概率密度和分布函数 分别用 $\varphi(x), \Phi(x)$ 表示 , 即有:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

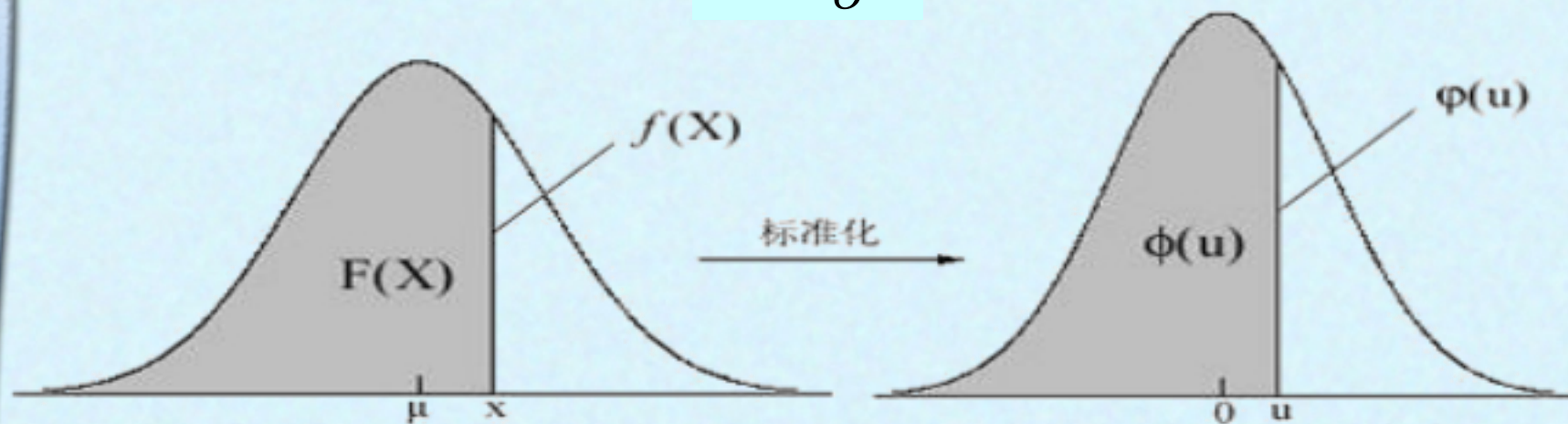


$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$



正态分布和标准正态分布的关系:

$$X \sim N(\mu, \sigma^2) \xrightarrow{\quad} Z = \frac{X - \mu}{\sigma} \xrightarrow{\quad} X \sim N(0, 1)$$



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

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

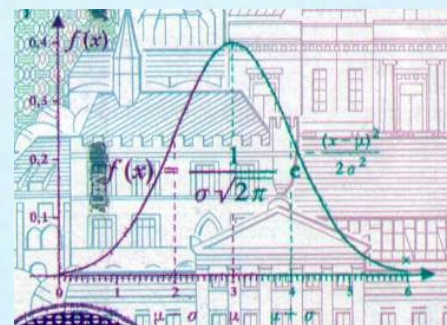
$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

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

可查表

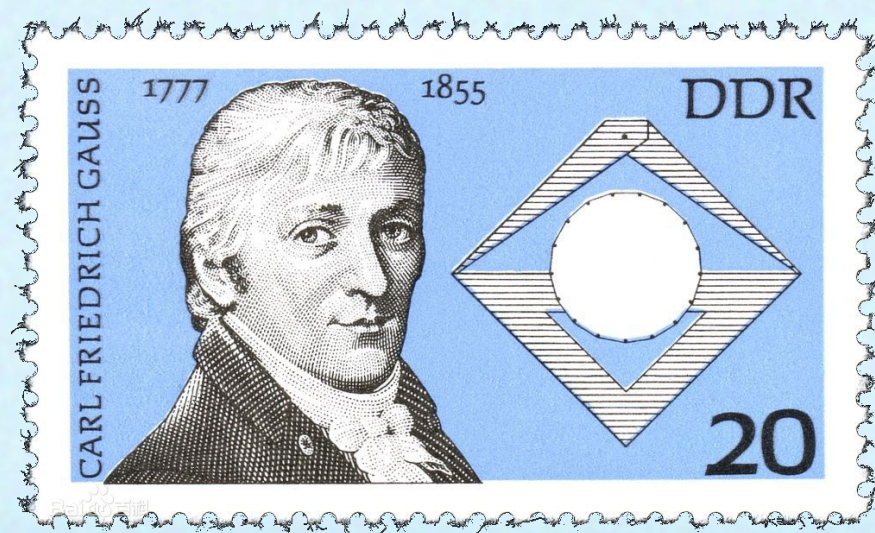
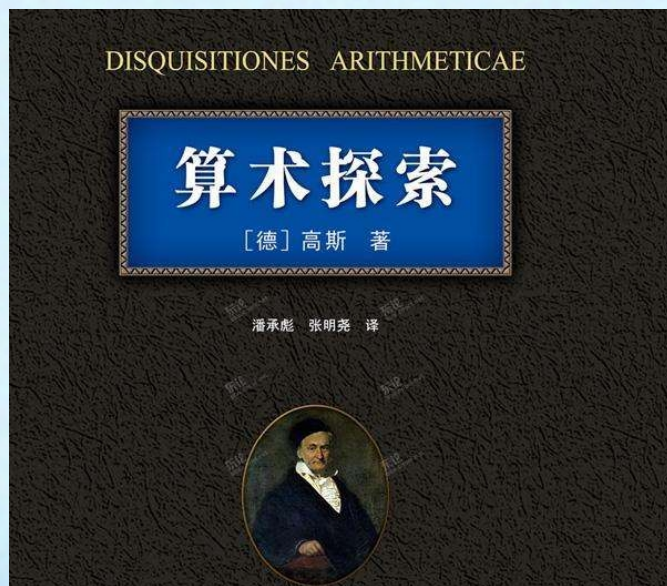
统计小知识：“数学王子”——高斯

高斯(1777-1855),德国数学家。近代数学奠基者之一,在历史上的影响,可以和阿基米德、牛顿、欧拉并列,有“**数学王子**”之称。



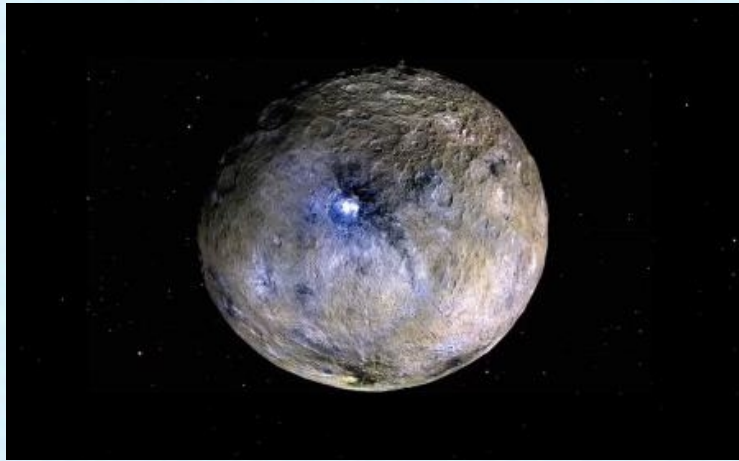
严肃的高斯与美丽的正态分布曲线出现在10元马克和纪念币

- 高斯少年就表现出超人的数学天才。
- 高斯的研究几乎遍及所有数学领域，在数论、代数学、非欧几何、复变函数和微分几何等方面都做出了开创性的贡献。

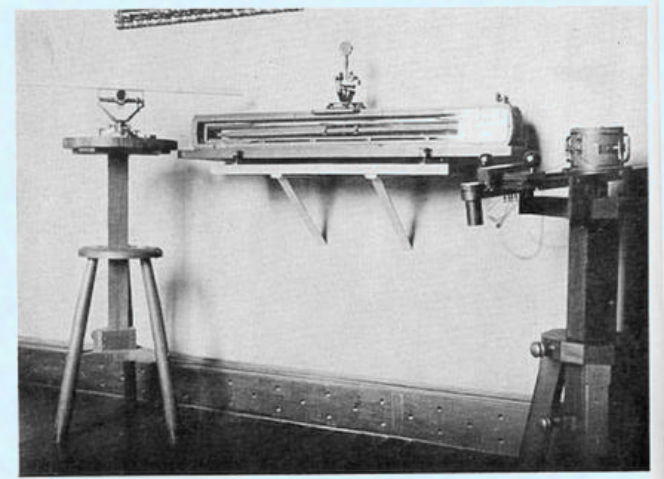


正十七边形

- 高斯还把数学应用于天文学、大地测量学和磁学的研究；



谷神星的轨道

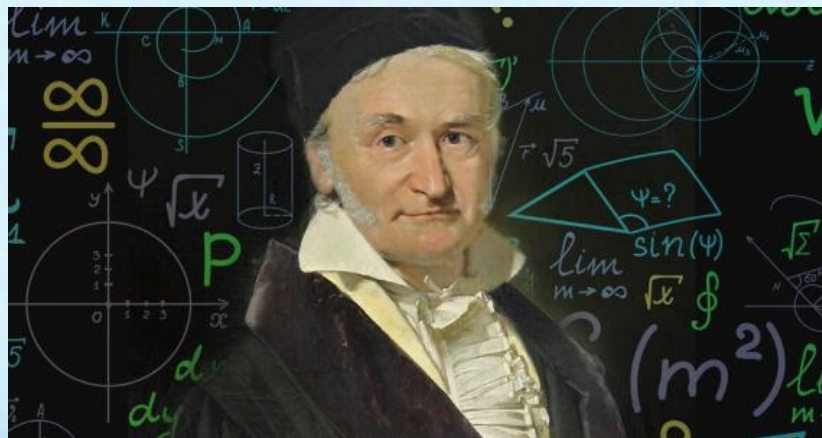


第一台电报机



日光反射镜

- 高斯一生共发表155篇论文，他对待学问十分严谨，只是把他自己认为是十分成熟的作品发表出来，这一点值得今天的学者好好学习。
- 由于高斯在数学、天文学、大地测量学和物理学中的杰出研究成果，“数学王子”的称号是对他一生恰如其分的总结。

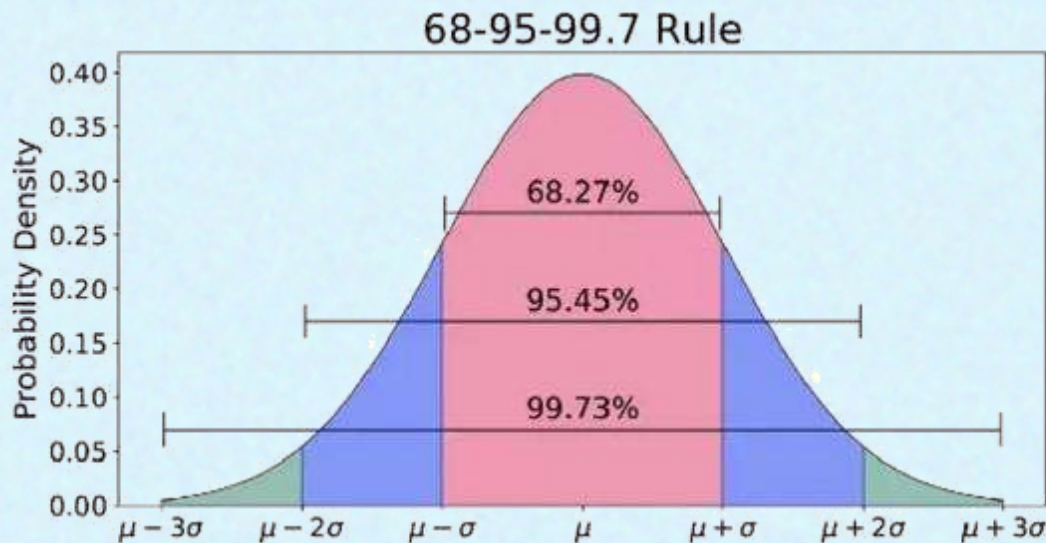


4. 正态分布的应用

正态分布是最常见最重要的一种分布，例如测量误差，人的生理特征尺寸如身高、体重等；正常情况下生产的产品尺寸、直径、长度、重量高度等都近似服从正态分布。



3 σ 原则 (68-95-99.7或 经验法则)



数值分布在 $(\mu - \sigma, \mu + \sigma)$ 中的概率为0.6827 ;

数值分布在 $(\mu - 2\sigma, \mu + 2\sigma)$ 中的概率为0.9545 ;

数值分布在 $(\mu - 3\sigma, \mu + 3\sigma)$ 中的概率为0.9973 ;

可以认为, Y 的取值几乎全部集中在 $(\mu - 3\sigma, \mu + 3\sigma)$ 区间内,
超出这个范围的可能性仅占不到0.3%.

3 σ 原则 (68-95-99.7或 经验法则)

在统计上, 68-95-99.7原则是在正态分布中, 距平均值小于一个标准差、二个标准差、三个标准差以内的百分比, 更精确的数字是68.27%、95.45%及99.73%。

$$P\{\mu - \sigma \leq X \leq \mu + \sigma\} \approx 0.6827$$

$$P\{\mu - 2\sigma \leq X \leq \mu + 2\sigma\} \approx 0.9545$$

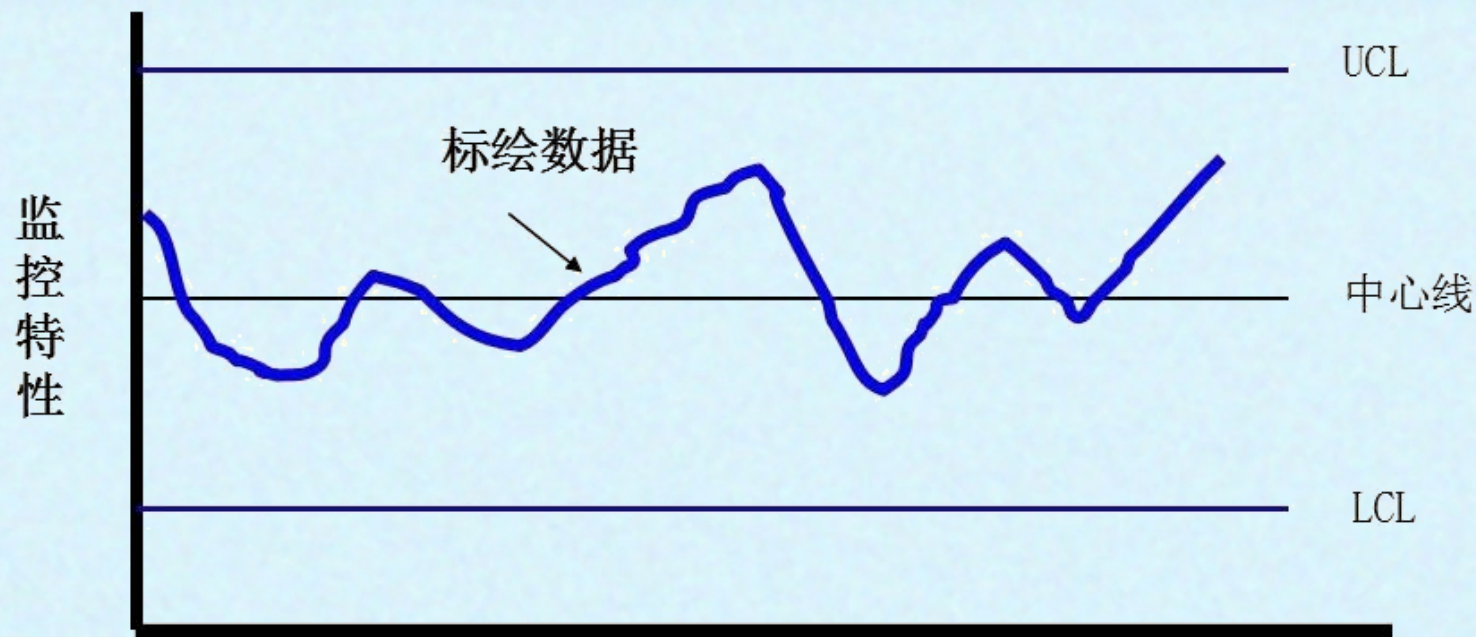
$$P\{\mu - 3\sigma \leq X \leq \mu + 3\sigma\} \approx 0.9973$$

在质量管理领域, 一般都使用 σ 来表示产品的合格程度, 表示质量控制水平.

3 σ 原则 (68-95-99.7或 经验法则)

- 在生产中，很多随机变量的概率分布都可以近似地用正态分布来描述，管理上的3 σ 原则是正态分布。
- 如果生产过程中是稳定可控的，产品良品率达到3 σ 水平(即99.73%)，这批产品品质优良；
- 如果有批产品良品率达不到3 σ 水平，这说明了在生产过程中某部分环节出现异常，造成产品数据发生波动，这就需要从生产找原因了。

SPC控制图构成



数据依时间而变化

UCL = Upper Control Limit / LCL = Lower Control Limit

为了更容易理解八大判异原则，先来看一下图表说明：上、下A/B/C六个区间，分别代表样本值的3倍、2倍、1倍标准差区间。

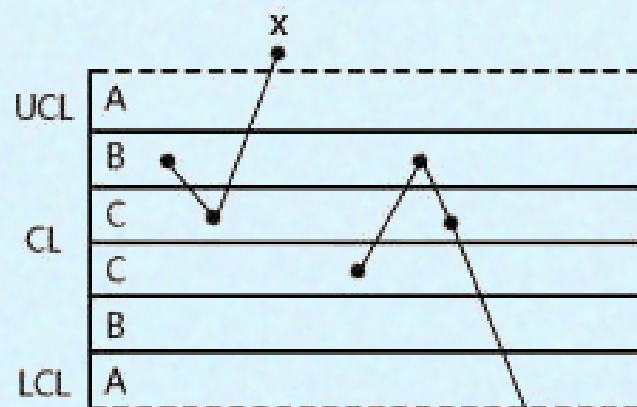
图表说明

UCL	A 3倍标准差区间
	B 2倍标准差区间
	C 1倍标准差区间
CL	C 1倍标准差区间
	B 2倍标准差区间
	A 3倍标准差区间
LCL	

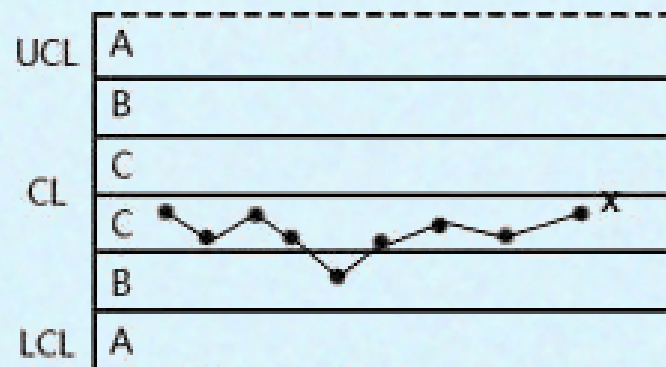
控制图判异的八个准则：

- 准则1：1个点子落在A区以外（点子越出控制界限）
- 准则2：连续9点落在中心线同一侧
- 准则3：连续6点递增或递减
- 准则4：连续14点中相邻点子总是上下交替
- 准则5：连续3点中有2点落在中心线同一侧B区以外
- 准则6：连续5点中有4点子落在中心线同一侧C区以外
- 准则7：连续15点落在中心线同两侧C区之内
- 准则8：连续8点落在中心线两侧且无1点在C区中

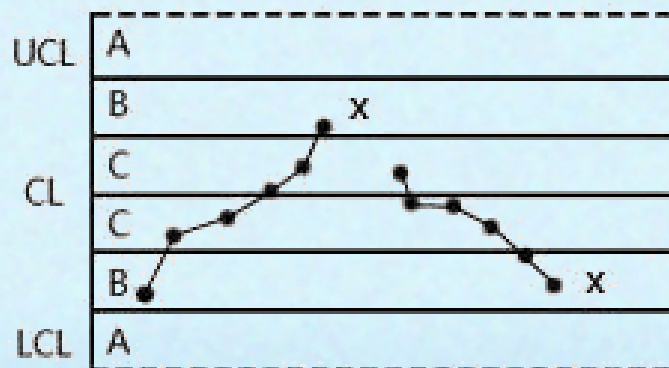
如何从SPC控制图中判段异常?



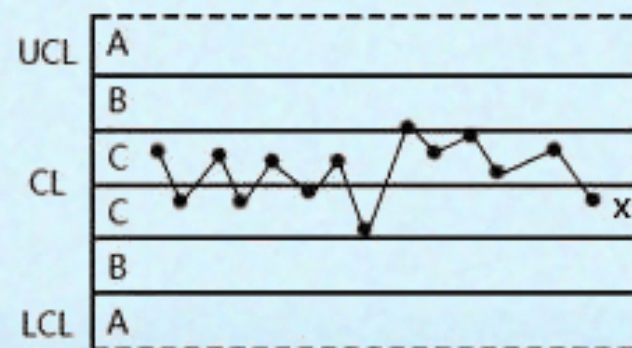
检验1:1个点落在A区以外^x



检验2:连续9点落在中心线同一侧

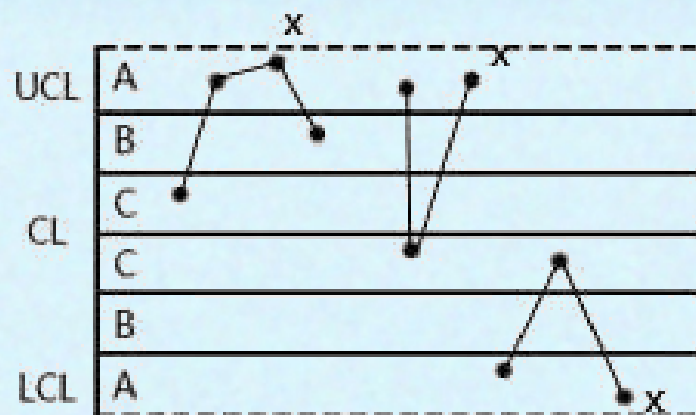


检验3:连续6点递增或递减

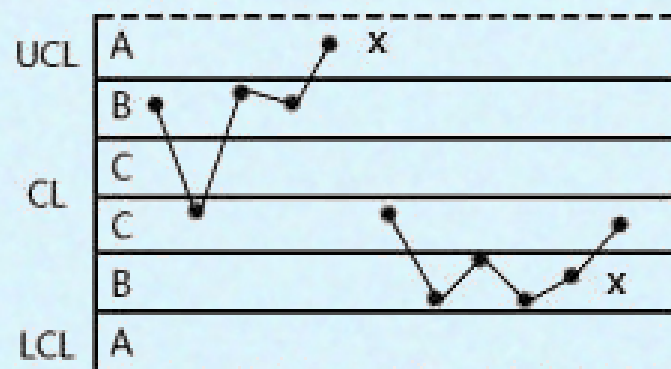


检验4:连续14点中相邻点交替上下

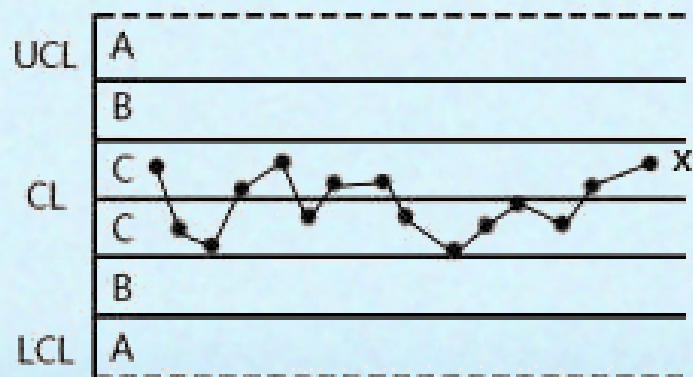
如何从SPC控制图中判段异常?



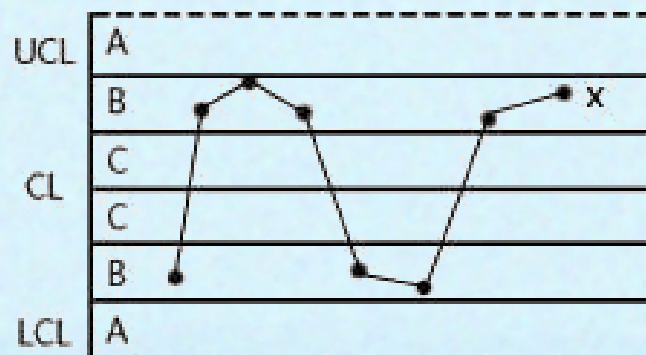
检验5:连续3点中有2点落在中心线同一侧的B区以外



检验6:连续5点中有4点落在中心线同一侧的C区以外

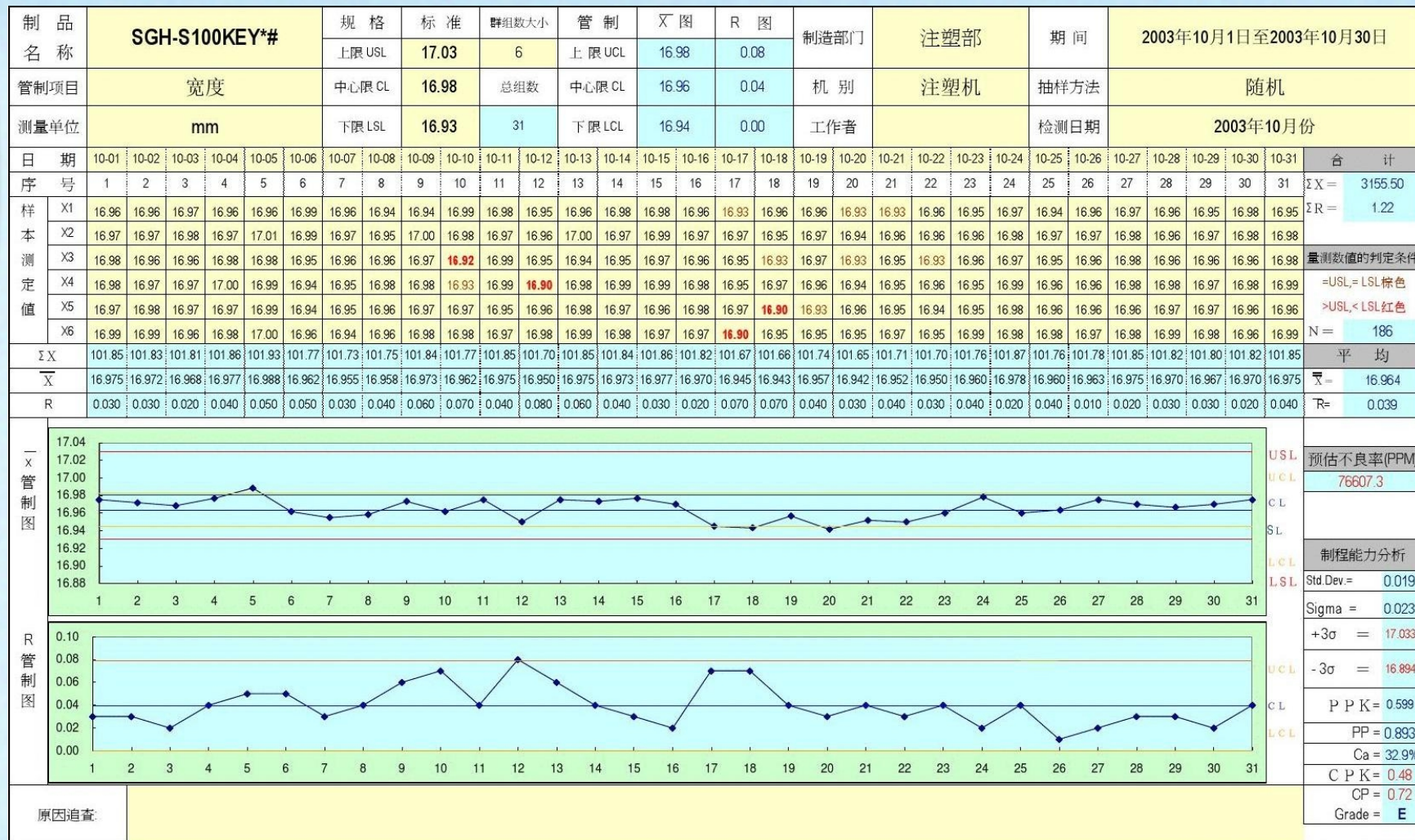


检验7:连续15点落在中心线两侧的C区内



检验8:连续8点落在中心线两侧且无一在C区内

2.4 连续型随机变量及其概率密度



联运塑胶制品厂有限公司SPC模版之X-R管制图

小结:

正态分布是一个在数学、物理及工程等领域都非常重要的概率分布，在统计学的许多理论知识方面有着重大的影响力。

谢 谢!