关于正态分布分布

正态分布　　t分布

计量资料的统计推断是以正态分布、　标准正态分布　、t分布为理论基础。

正态分布、标准正态分布、　t分布的相互关系是参数估计和假设检验的理论基础。

本课件主要学习正态分布、标准正态分布、t分布的概念、分布特征、相互关系。

正态分布　　t分布

一、正态分布

（一）正态分布的概念

（二）正态分布曲线下的面积分布规律

（三）正态分布曲线的两个参数

（四）标准正态分布

（五）标准正态分布曲线下的面积分布规律

二、 t分布

（一）均数的抽样误差

（二）样本均数的正态分布（中心极限定理）

（三）样本均数的标准正态分布

（四）t值、t分布

（五）t分布特征

一、正态分布

（一）正态分布的概念

正态分布又称高斯分布，是一种很重要的连续型分布，应用甚广。在医学卫生领域中有许多变量的频数分布资料可绘制成直方图而且频数分布是中间（靠近均数处）频数多，两边频数少，且左右对称。

可以设想，如果将观察人数逐渐增多，组段不断分细，图中直条将逐渐变窄，其顶端的中点的连线将逐渐接近于一条光滑的曲线，这条曲线略呈钟型，两头低，中间高，左右对称，近似于数学上的正态分布曲线（图1）

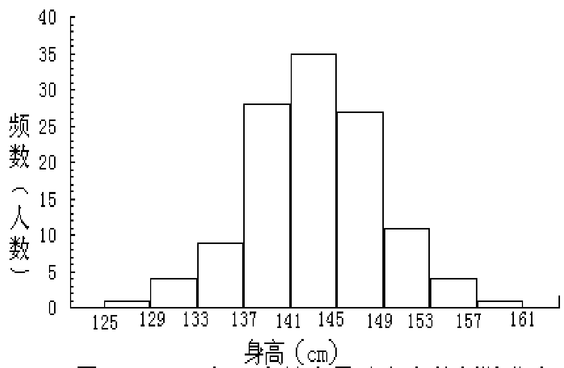


图1.1　120名12岁健康男孩身高的频数分布

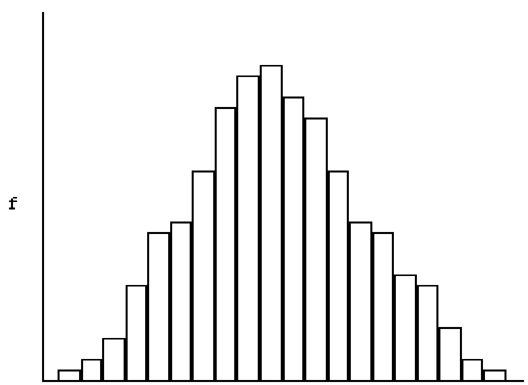


图1.2 不断缩小组距，增加观察例数，则频数分布逐渐接近正态分布

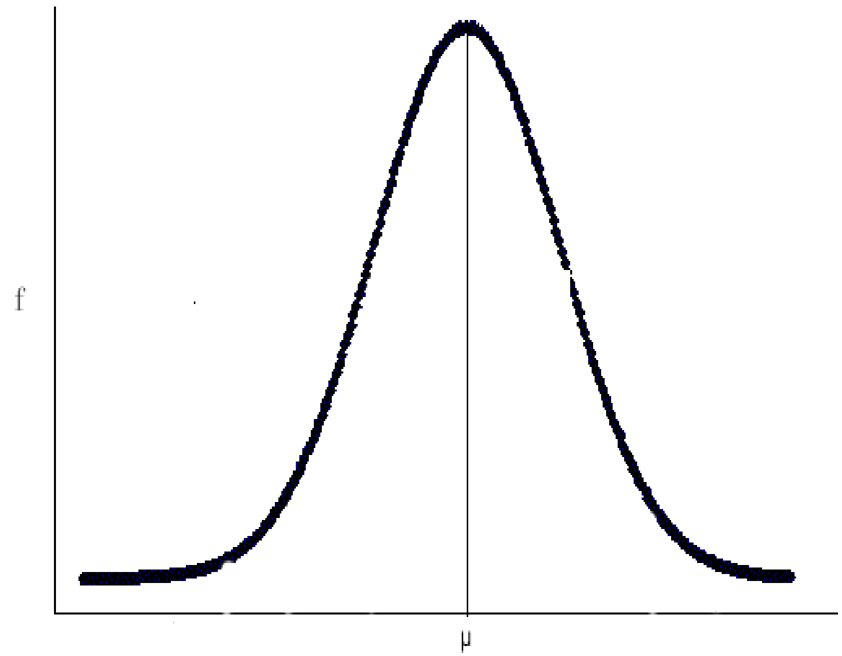


图1.3　正态分布曲线

正态分布的特征

- 正态分布曲线以均数为中心，左右对称。
- 正态分布曲线下的面积分布有一定的规律
- 正态分布曲线在横轴上方均数处最高。
- 正态分布曲线有两个参数：均数μ 为位置参数，标准差σ 为形状参数。

（二）正态分布曲线下的面积分布规律

- 数理统计证明：正态分布曲线下与横轴之间的整体面积为1或100%。以μ为总体均数，σ为总体标准差，则正态分布曲线下面积的分布规律经积分法计算有如下规律（图2）
- $\mu \pm 1\sigma$范围内的面积占正态曲线下总面积的68.27%，即有68.27%的变量值分布在此范围内；
- $\mu \pm 1.96\sigma$范围内的面积占正态曲线下总面积的95.00%，即有95.00%的变量值分布在此范围内；
- $\mu \pm 2.58\sigma$范围内的面积占正态曲线下总面积99.00%，即有99.00%的变量值分布在此范围内

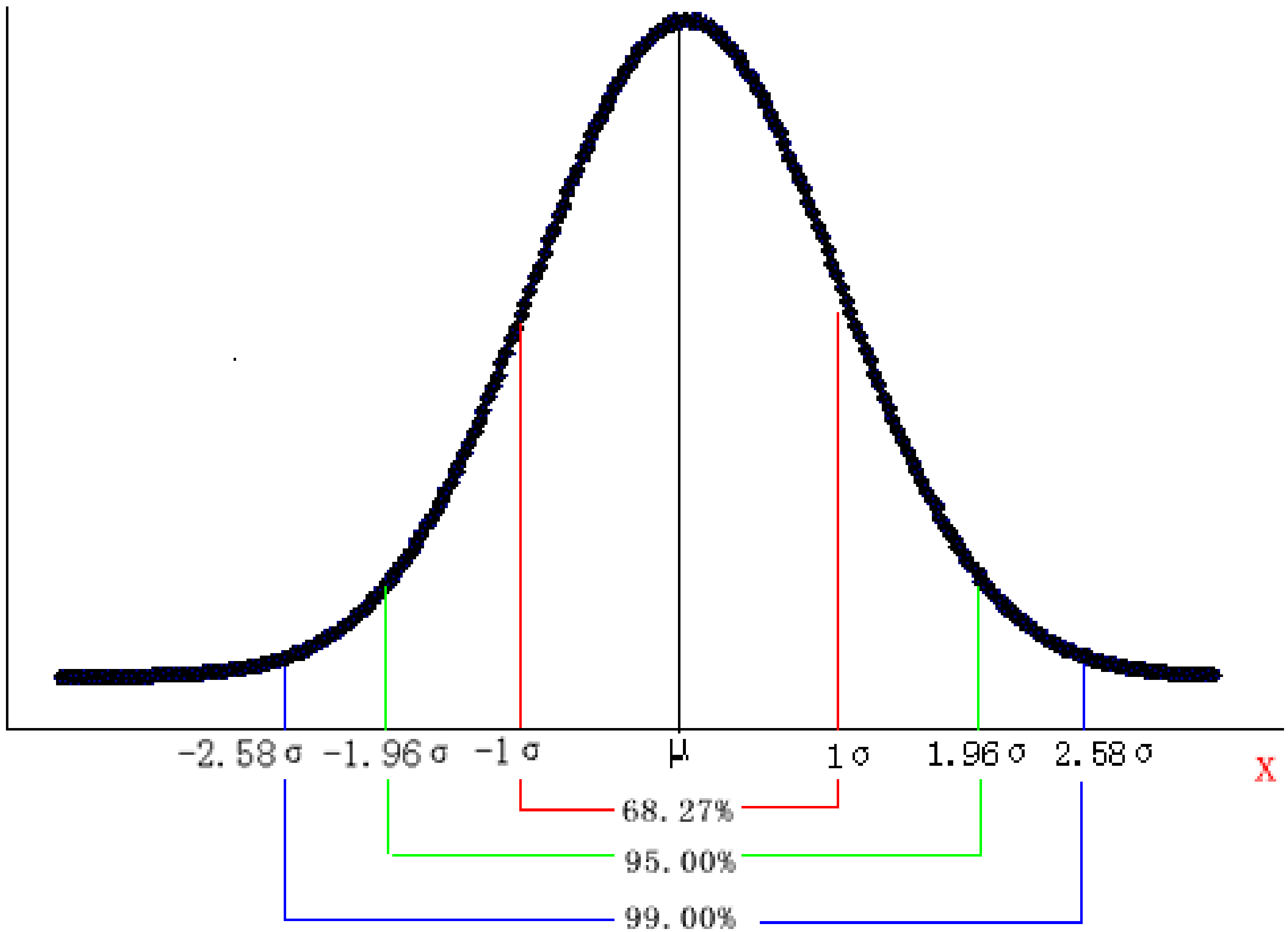


图2 正态曲线下面积的分布规律

（三）正态分布曲线的两个参数

- 均数μ决定曲线在横轴上的位置是正态分布曲线的位置参数（图3.1）。

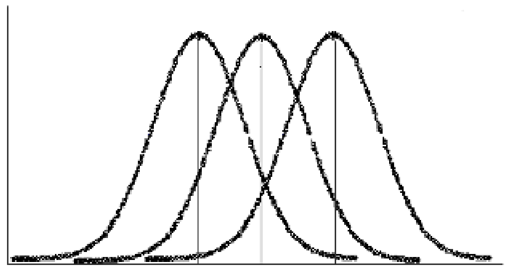


图3.1 标准差相同 $\sigma=1$ 而均数 μ 不同的三条曲线

- 标准差σ决定曲线的形状是正态分布曲线的形状参数（变异度参数）（图3.2）。

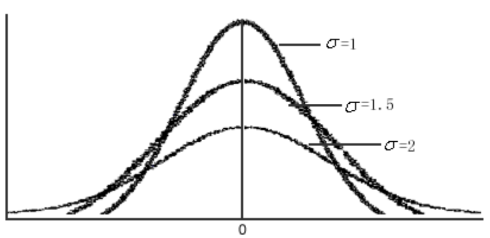


图.2 均数相同 $\mu=0$ 而 σ 不同的三条正态曲线

（四）标准正态分布

对于任何一个均数为μ，标准差为σ的正态分布，都可以通过变换，使之成为μ=0, σ=1的标准正态分布。变换的方法是将变量值x变换为u，u=x- μ / σ，u值的分布就是标准正态分布。

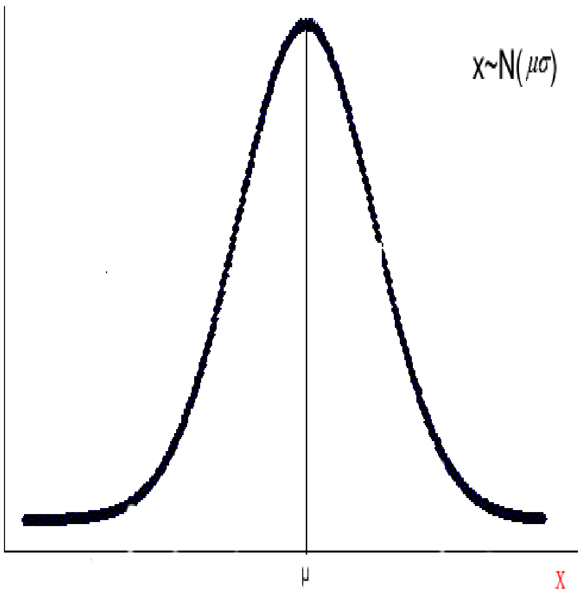


图4.1　正态分布曲线

$$u=x-\mu/\sigma \longrightarrow$$

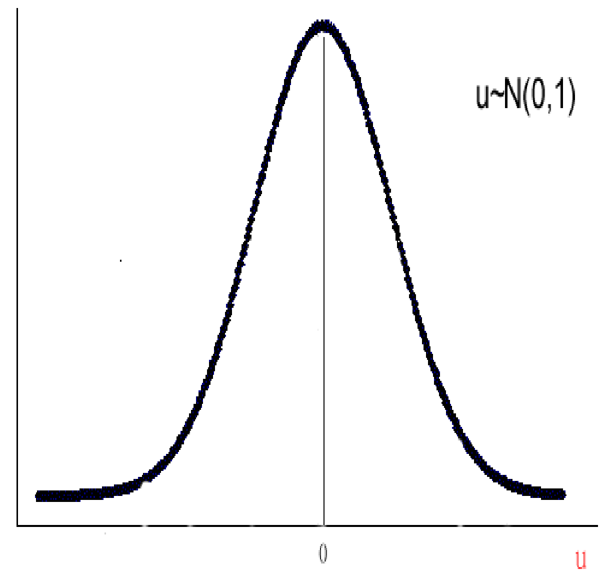


图4.2　标准正态分布曲线

（五）标准正态分布曲线下的面积分布规律

- 标准正态分布曲线以u值为横轴变量，位置参数μ=0，形状参数σ=1，标准正态分布曲线与横轴之间的整体面积为1或100%。标准正态分布曲线下面积的分布规律有如下规律（图5）
- u=-1,u=1范围内的面积占正态曲线下总面积的68.27%，即有68.27%的变量值分布在此范围内；
- u=-1.96,u=1.96 范围内的面积占正态曲线下总面积的95.00%， 即有95.00%的变量值分布在此范围内；
- u=-2.58,u=2.58范围内的面积占正态曲线下总面积99.00%， 即有99.00%的变量值分布在此范围内。

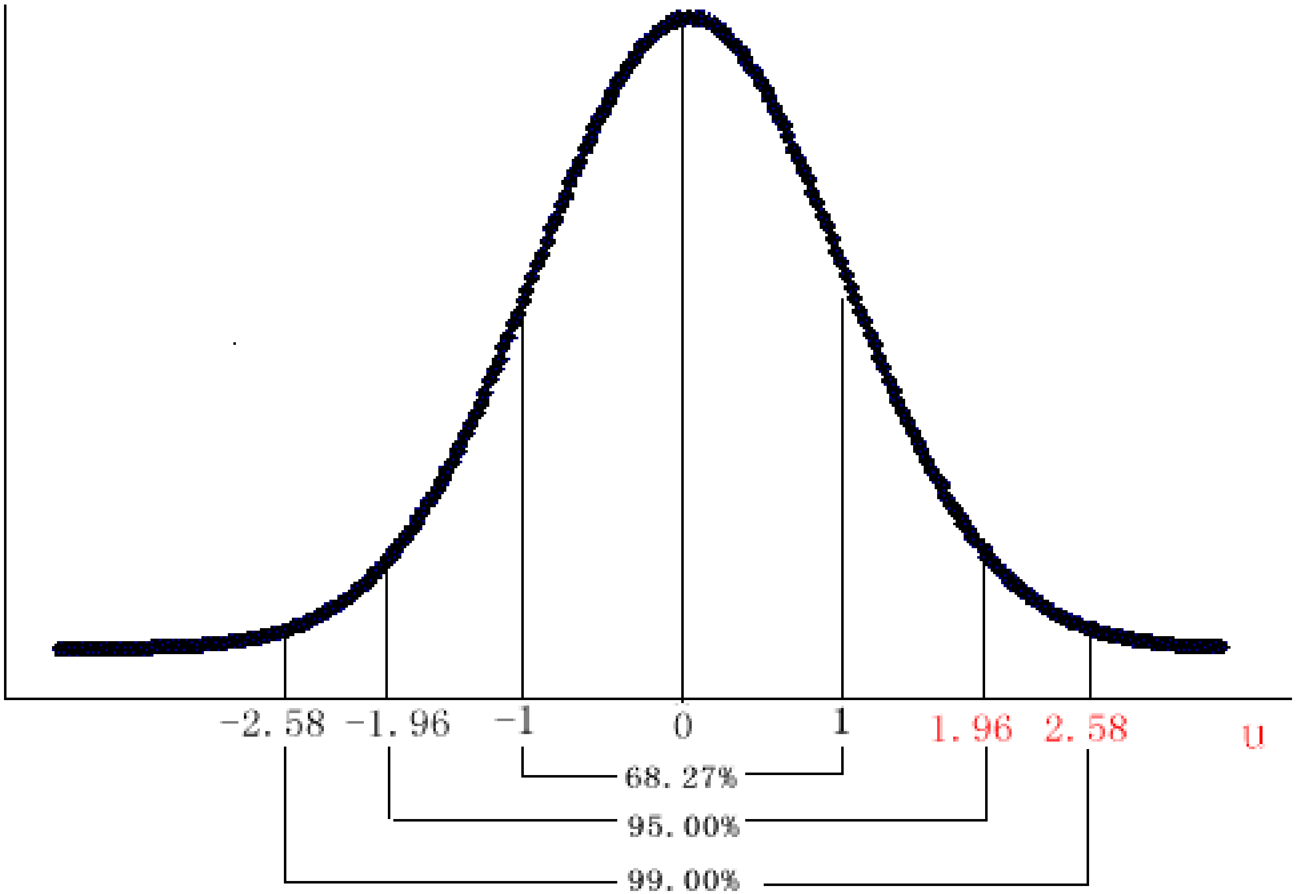


图5 标准正态分布曲线下的面积分布规律

二、t 分布

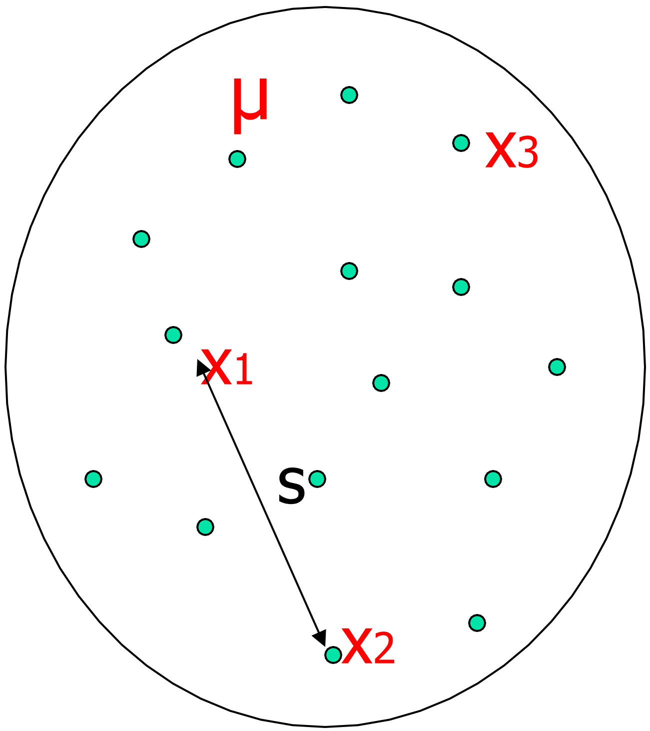
(一)均数的抽样误差　标准误

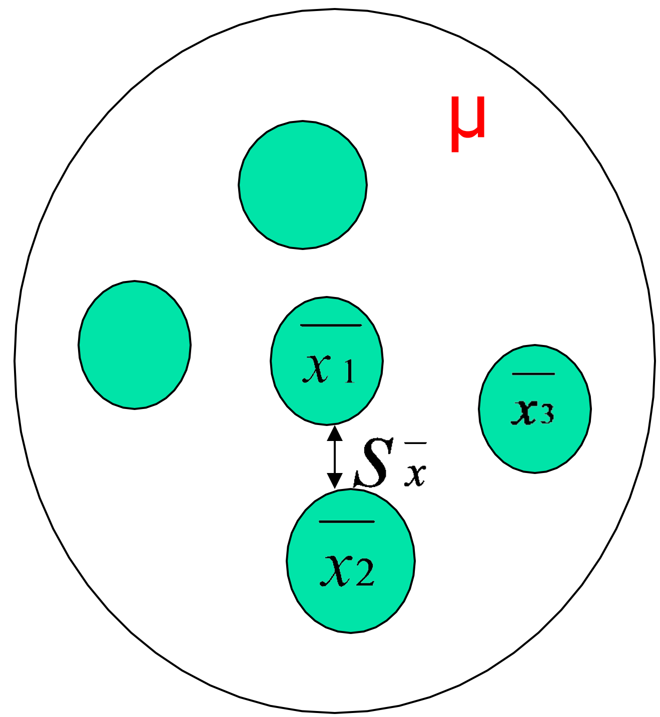
- 在总体中随机抽取一部分个体作为样本，进行调查研究以推论总体的方法，称为抽样研究方法。
- 由抽样而引起的样本均数与总体均数之间的差别及样本均数与样本均数之间的差别称为抽样误差。
- 从正态分布的同一总体中随机抽取例数相等的若干个样本，分别计算它们的均数，这些样本均数的标准差称为标准误。

标准误与标准差的区别

- 标准差描述个体变量值间的变异程度。凡同性质的资料，标准差大表示个体变量值变异大，样本均数对个体的代表性差。标准差小表示个体变量值变异小，样本均数对个体的代表性好。
- 标准误是样本均数的标准差，即描述样本均数的抽样误差。凡同性质的资料，标准误大说明抽样误差大，用样本均数估计总体均数的可靠性小；而标准误小，说明抽样误差小，用样本均数估计总体均数的可靠性大。

标准误与标准差的区别





$\bar{x} \pm s$

$\bar{x} \pm s_{\bar{x}}$

(二)样本均数的正态分布（中心极限定理）

从一个呈正态分布的总体中随机抽取样本含量相等的许多样本，分别计算出它们的样本均数。这些样本均数的频数分布仍是以总体均数为中心的正态分布。

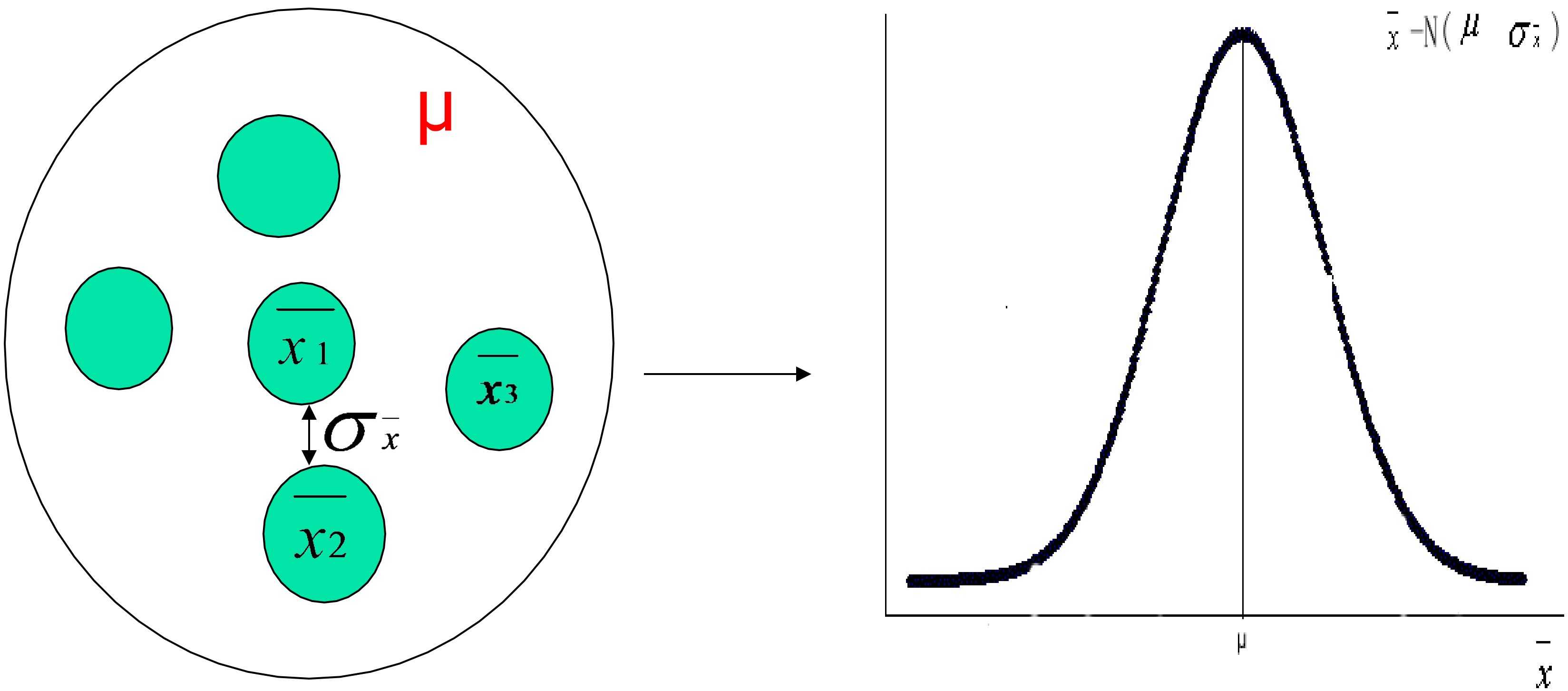


图6 均数的正态分布曲线

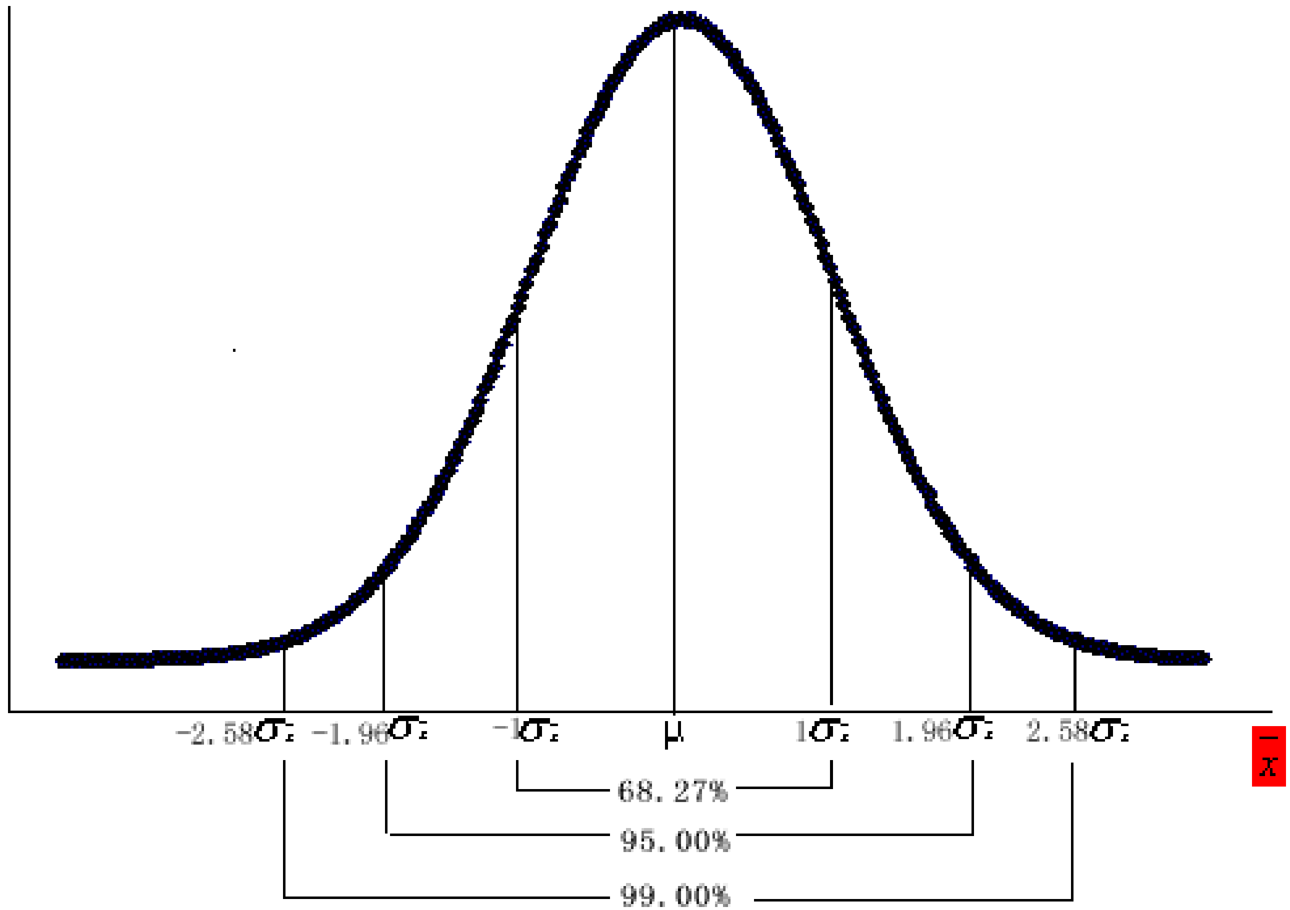


图7 样本均数的正态曲线下面积的分布规律

(三)样本均数的标准正态分布

对于任何一个横轴变量为$\bar{x}$ 均数为μ，标准误为$\sigma_{\bar{x}}$ 的正态分布，都可以通过变换，使之成为μ=0 $\sigma_{\bar{x}}$ =1的标准正态分布。变换的方法是将变量值$\bar{x}$ 变换为u，u=$\bar{x}$ - μ/$\sigma_{\bar{x}}$ ，u值的分布就是标准正态分布。

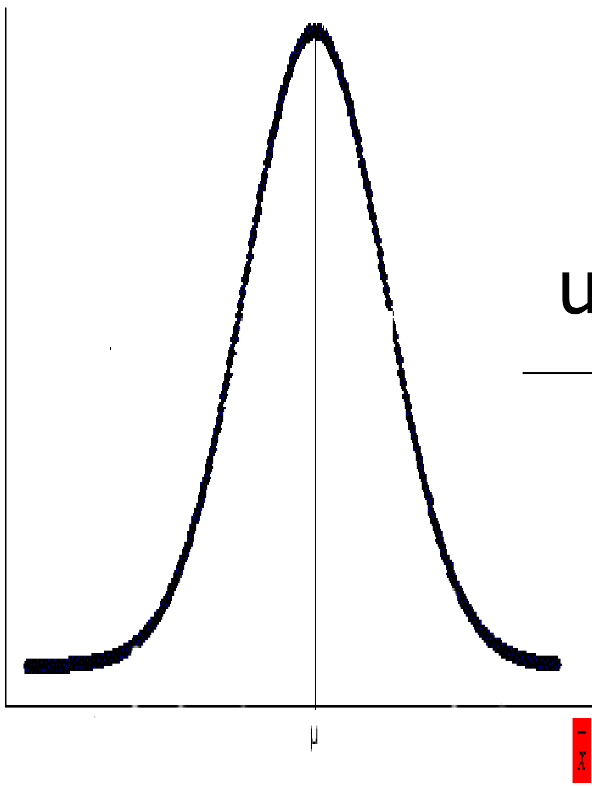


图8.1 均数的正态分布曲线

$$u=\bar{x}\ -\mu/\sigma_{\bar{x}}$$

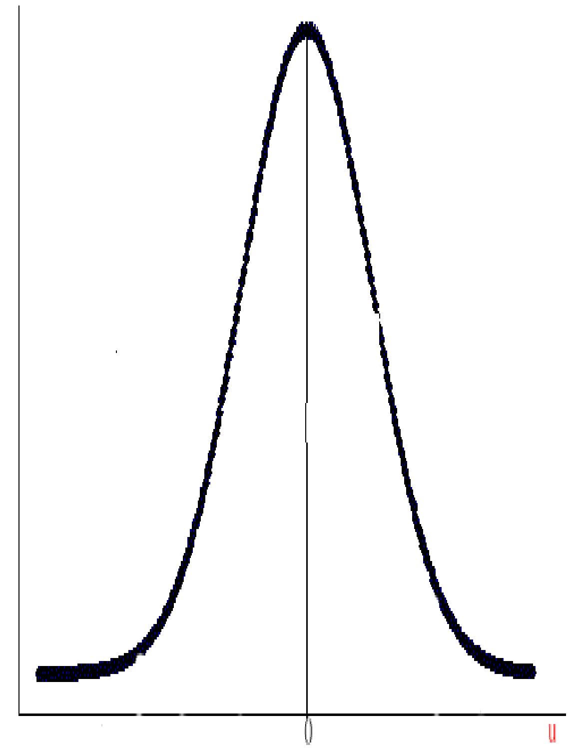


图8.2 均数的标准正态分布曲线

(四) t值　t分布

对于任何一个横轴变量为 $\bar{x}$ 均数为μ，标准误为 $\sigma_{\bar{x}}$ 的正态分布，都可以通过变换，使之成为μ=0, $\sigma_{\bar{x}}$ =1的标准正态分布。变换的方法是将变量值 $\bar{x}$ 变换为u，u= $\bar{x}$ - μ/ $\sigma_{\bar{x}}$ ，u值的分布就是标准正态分布。实际工作中 $\sigma_{\bar{x}}$ 常用 $S_{\bar{x}}$ 估计，t值就是样本均数 $\bar{x}$ 与总体均数μ的差数除以 $s_{\bar{x}}$ 所得之商 $t = \bar{x} - \mu / s_{\bar{x}}$