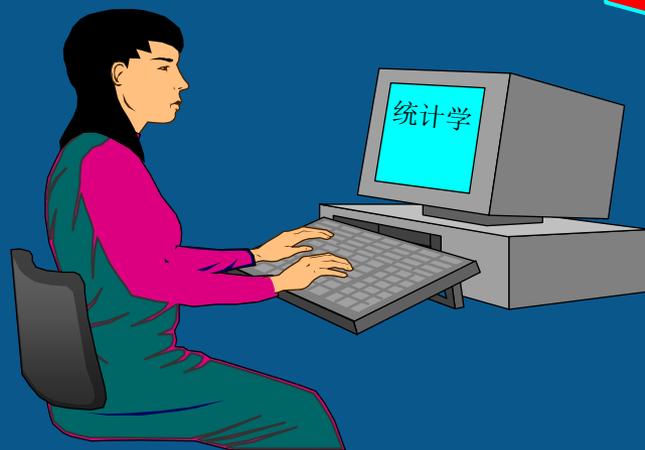


第 6 章 统计量及其抽样分布

PowerPoint



第 6 章 统计量及其抽样分布

6.1 统计量

6.2 由正态分布导出的几个重要分布

6.3 样本均值的分布与中心极限定理

学习目标

1. 了解统计量及其分布的几个概念
2. 了解由正态分布导出的几个重要分布
3. 理解样本均值的分布与中心极限定理
4. 掌握单样本比例和样本方差的抽样分布

6.1 统计量

6.1.1 统计量的概念

6.1.2 常用统计量

统计量 (statistic)

1. 设 X_1, X_2, \dots, X_n 是从总体 X 中抽取的容量为 n 的一个样本，如果由此样本构造一个函数 $T(X_1, X_2, \dots, X_n)$ ，不依赖于任何未知参数，则称函数 $T(X_1, X_2, \dots, X_n)$ 是一个统计量
 - 样本均值、样本比例、样本方差等都是统计量
2. 统计量是样本的一个函数
3. 统计量是统计推断的基础

常用统计量

(1) $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 是样本的均值，它反映出总体 X 数学期望的信息。样本均值是最常用的统计量。

(2) $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 是样本方差，它反映的是总体 X 方差的信息。样本方差 S^2 及样本标准差 S 也是最常用的统计量。

(3) $V = S/\bar{X}$ 是样本变异系数，它反映出总体变异系数 C 的信息。其中变异系数定义为 $C = \sqrt{D(X)}/E(X)$ ，它反映出随机变量在以它的均值为单位时取值的离散程度。此统计量消除了均值不同对不同总体的离散程度的影响，常用来刻画均值不同时不同总体的离散程度。它在投资项目的风险分析中、不同群体或行业的收入差距描述中有着广泛的应用。

(4) $m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ ，称 m_k 为样本 k 阶矩。它反映出总体 k 阶矩的信息。显然， $m_1 = \bar{X}$ 就是样本均值。

(5) $v_k = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^k$ ，称 v_k 为样本 k 阶中心矩。它反映出总体 k 阶中心矩的信息。显然， v_2 就是样本方差。

(6) $\alpha_3 = \sqrt{n-1} \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{(\sum_{i=1}^n (X_i - \bar{X})^2)^{3/2}}$ ，称 α_3 为样本偏度。它反映出总体偏度的信息。偏度反映了随机变量密度函数曲线在众数（密度函数在这一点达到最大值）两边的偏斜性。如果 $X \sim N(\mu, \sigma^2)$ ，则偏度 $\alpha_3 = 0$ 。

(7) $\alpha_4 = n-1 \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2} - 3$ ，称 α_4 为样本峰度。它反映出总体峰度的信息。峰度反映了密度函数曲线在众数附近的“峰”的尖锐程度。正态随机变量 $X \sim N(\mu, \sigma^2)$ 的峰度 $\alpha_4 = 0$ 。

6.2 由正态分布导出的几个重要分布

6.2.1 抽样分布

6.2.2 χ^2 分布

6.2.2 t 分布

6.2.3 F 分布

抽样分布

(sampling distribution)

1. 样本统计量的概率分布，是一种理论分布
 - 在重复选取容量为 n 的样本时，由该统计量的所有可能取值形成的相对频数分布
2. 随机变量是 样本统计量
 - 样本均值, 样本比例, 样本方差等
3. 结果来自容量相同的所有可能样本
4. 提供了样本统计量长远而稳定的信息，是进行推断的理论基础，也是抽样推断科学性的重要依据

χ^2 分布

χ^2 分布

(χ^2 distribution)

1. 由阿贝(Abbe)于1863年首先给出, 后来由海尔墨特(Hermert)和卡·皮尔逊(K·Pearson)分别于1875年和1900年推导出来

2. 设 $X \sim N(\mu, \sigma^2)$, 则 $z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

3. 令 $Y = z^2$, 则 Y 服从自由度为1的 χ^2 分布, 即

$$Y \sim \chi^2(1)$$

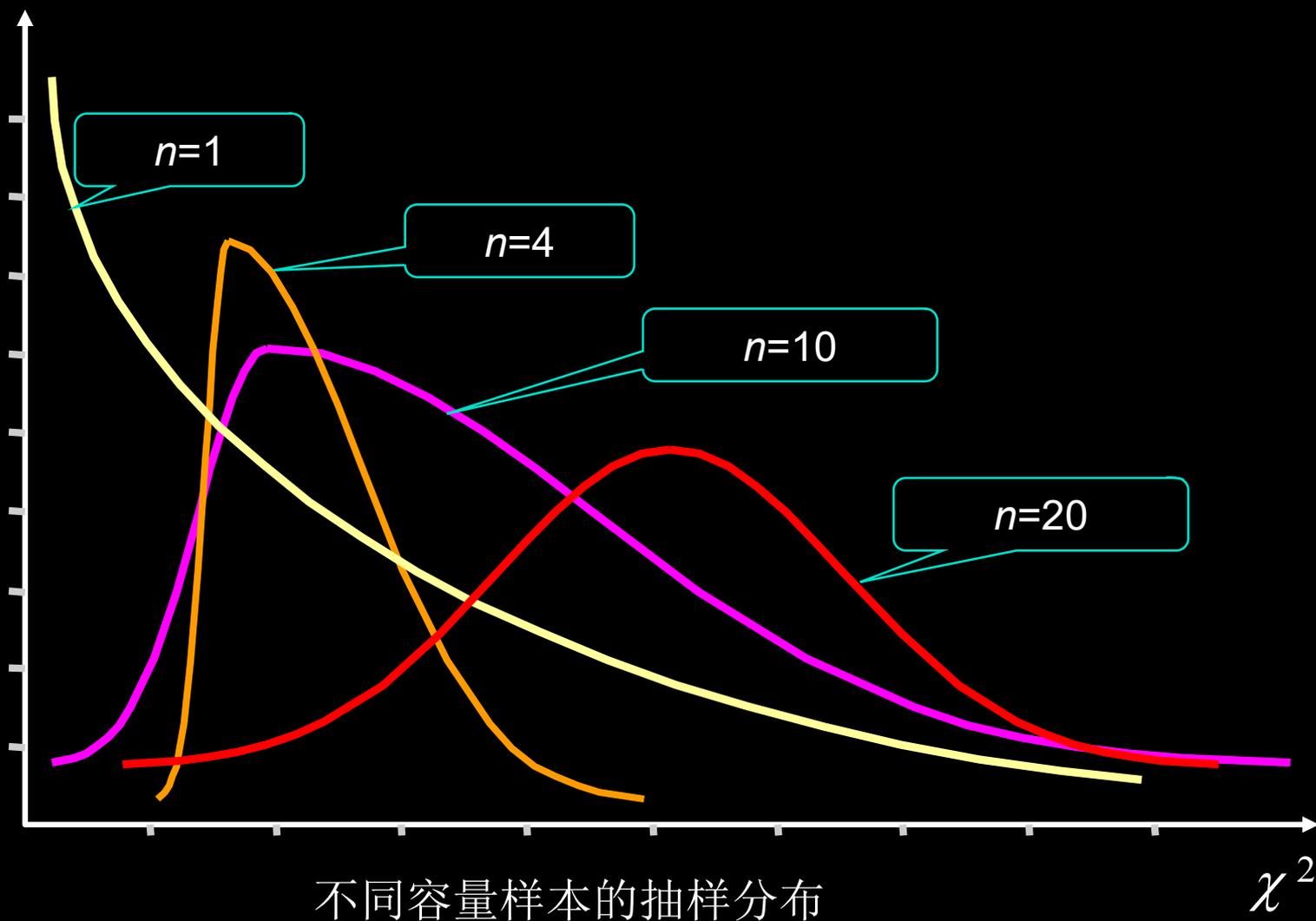
4. 当总体 $X \sim N(\mu, \sigma^2)$, 从中抽取容量为 n 的样本, 则

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} \sim \chi^2(n-1)$$

χ^2 分布 (性质和特点)

1. 分布的变量值始终为正
2. 分布的形状取决于其自由度 n 的大小，通常为不对称的正偏分布，但随着自由度的增大逐渐趋于对称
3. 期望为： $E(\chi^2)=n$ ，方差为： $D(\chi^2)=2n$ (n 为自由度)
4. 可加性：若 U 和 V 为两个独立的 χ^2 分布随机变量， $U\sim\chi^2(n_1)$ ， $V\sim\chi^2(n_2)$ ，则 $U+V$ 这一随机变量服从自由度为 n_1+n_2 的 χ^2 分布

χ^2 分布 (图示)

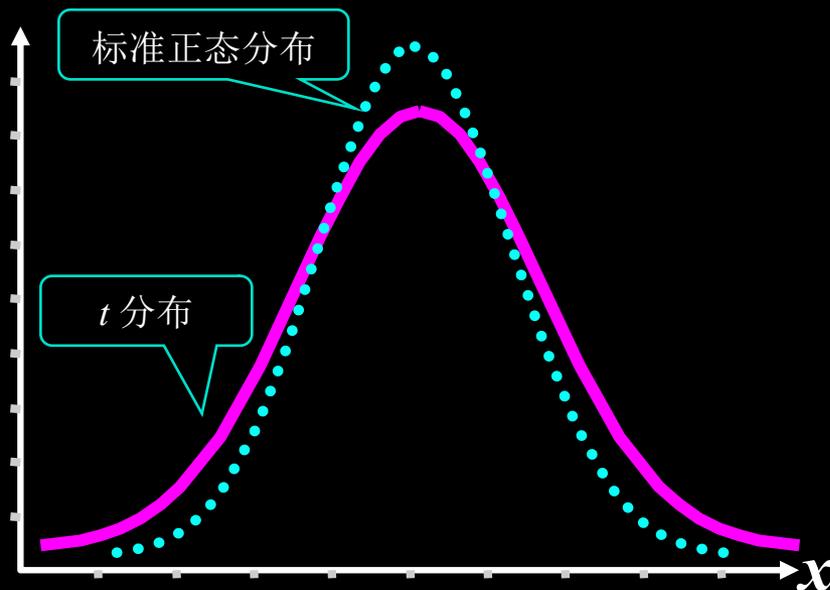


t 分布

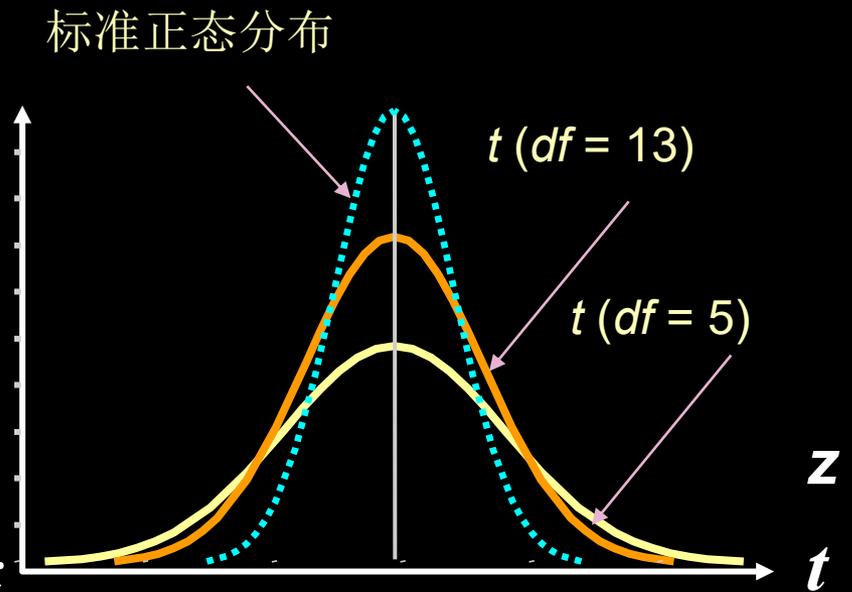
t 分布

1. 高塞特(W.S.Gosset)于1908年在一篇以“Student”(学生)为笔名的论文中首次提出
2. t 分布是类似正态分布的一种对称分布，它通常要比正态分布平坦和分散
3. 一个特定的分布依赖于称之为自由度的参数。随着自由度的增大，分布也逐渐趋于正态分布

t 分布图示



t 分布与标准正态分布的比较



不同自由度的 t 分布

F 分布

F 分布

(F distribution)

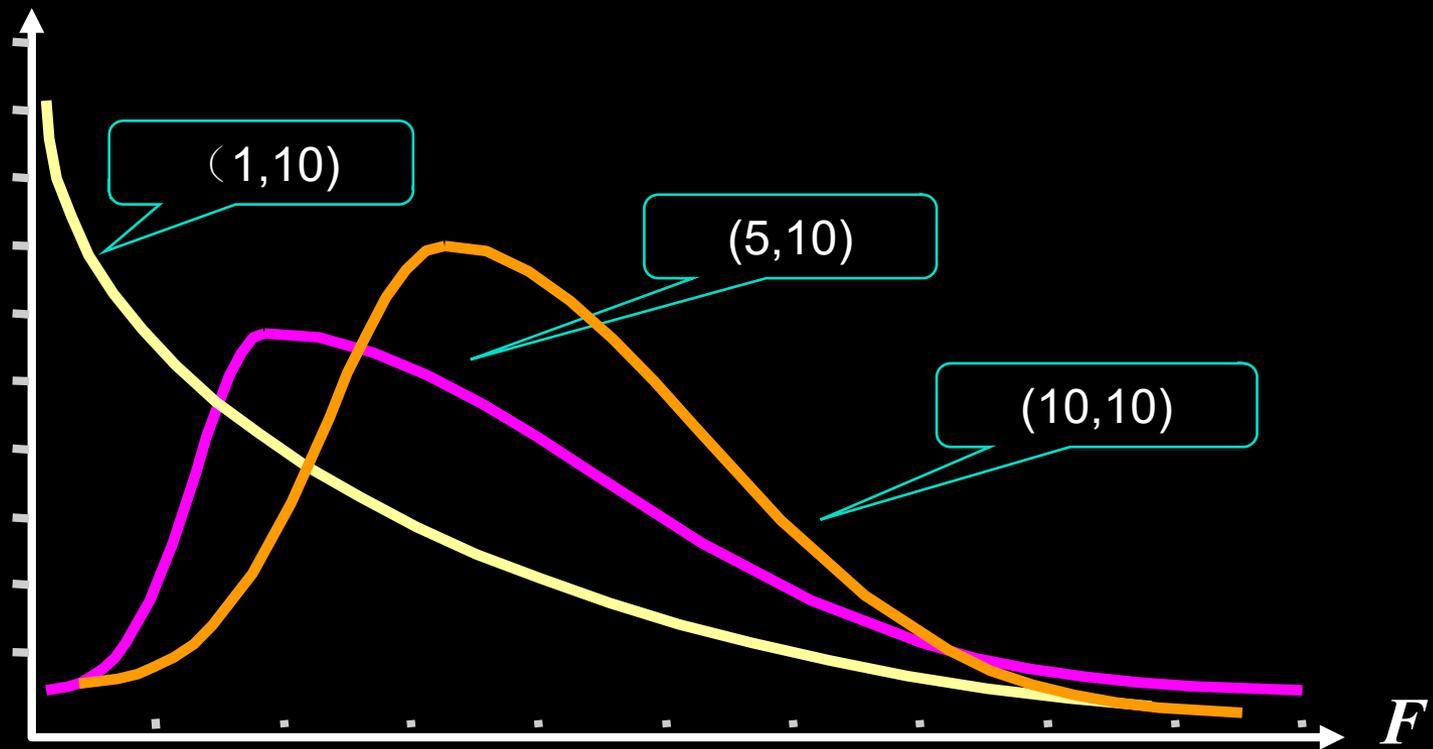
1. 由统计学家费希尔(R.A.Fisher)提出的, 以其姓氏的第一个字母来命名
2. 设若 U 为服从自由度为 n_1 的 χ^2 分布, 即 $U \sim \chi^2(n_1)$, V 为服从自由度为 n_2 的 χ^2 分布, 即 $V \sim \chi^2(n_2)$, 且 U 和 V 相互独立, 则称 F 为服从自由度 n_1 和 n_2 的 F 分布, 记为

$$F = \frac{U/n_1}{V/n_2}$$

$$F \sim F(n_1, n_2)$$

F 分布 (图示)

→ 不同自由度的 F 分布



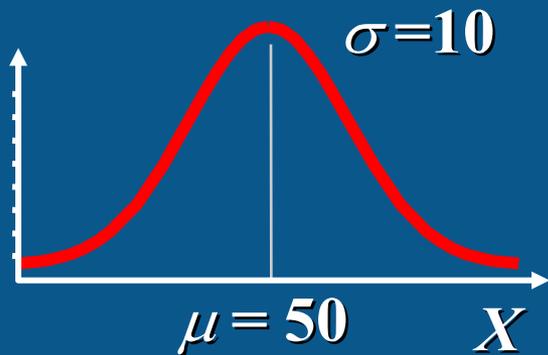
6.3 样本均值的分布与中心极限定理

样本均值的抽样分布

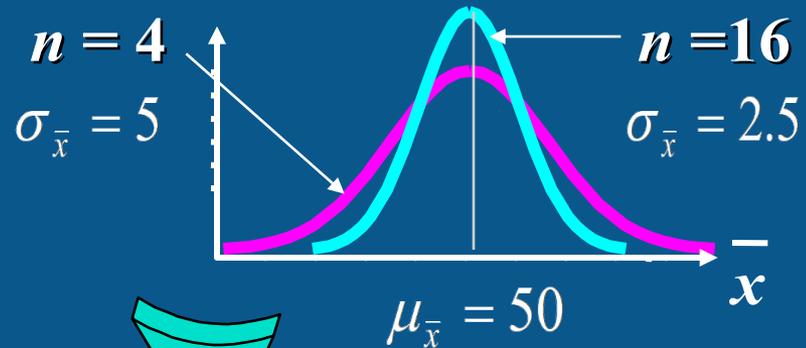
1. 在重复选取容量为 n 的样本时，由样本均值的所有可能取值形成的相对频数分布
2. 一种理论概率分布
3. 推断总体均值 μ 的理论基础

样本均值的抽样分布 与中心极限定理

当总体服从正态分布 $N(\mu, \sigma^2)$ 时，来自该总体的所有容量为 n 的样本的均值 \bar{x} 也服从正态分布， \bar{x} 的数学期望为 μ ，方差为 σ^2/n 。即 $\bar{x} \sim N(\mu, \sigma^2/n)$



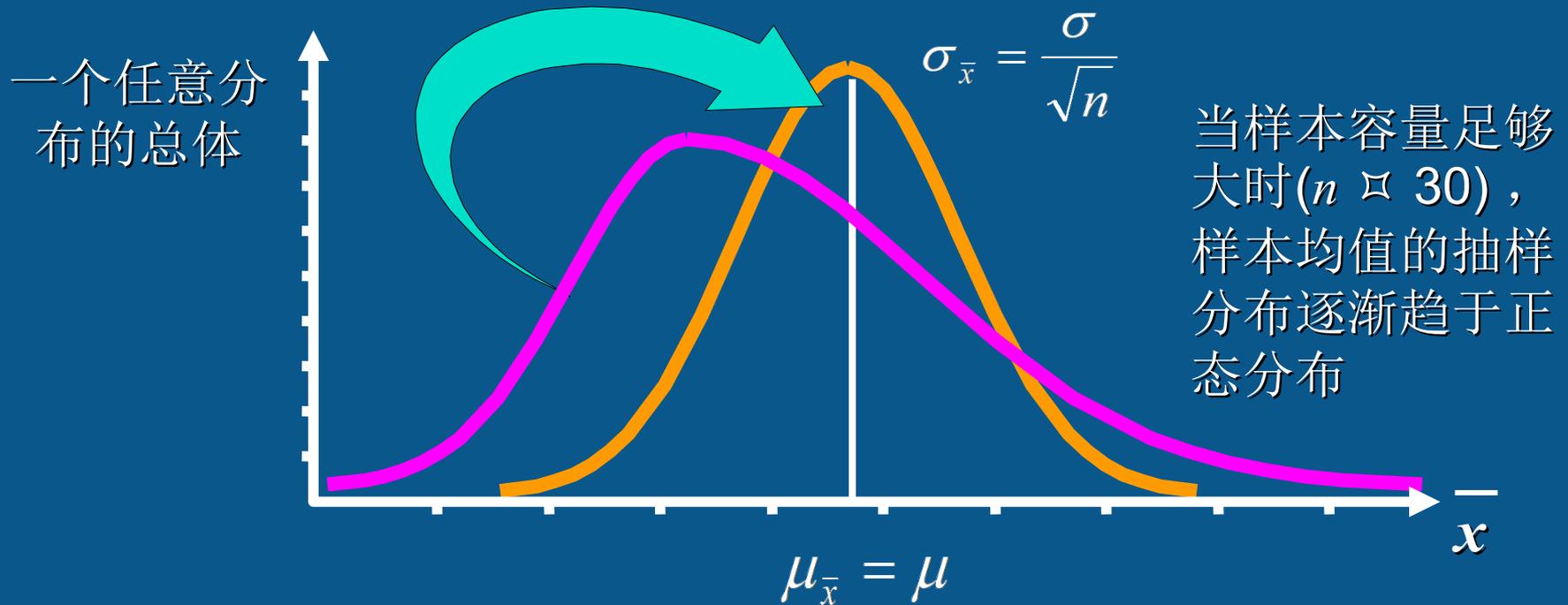
总体分布



抽样分布

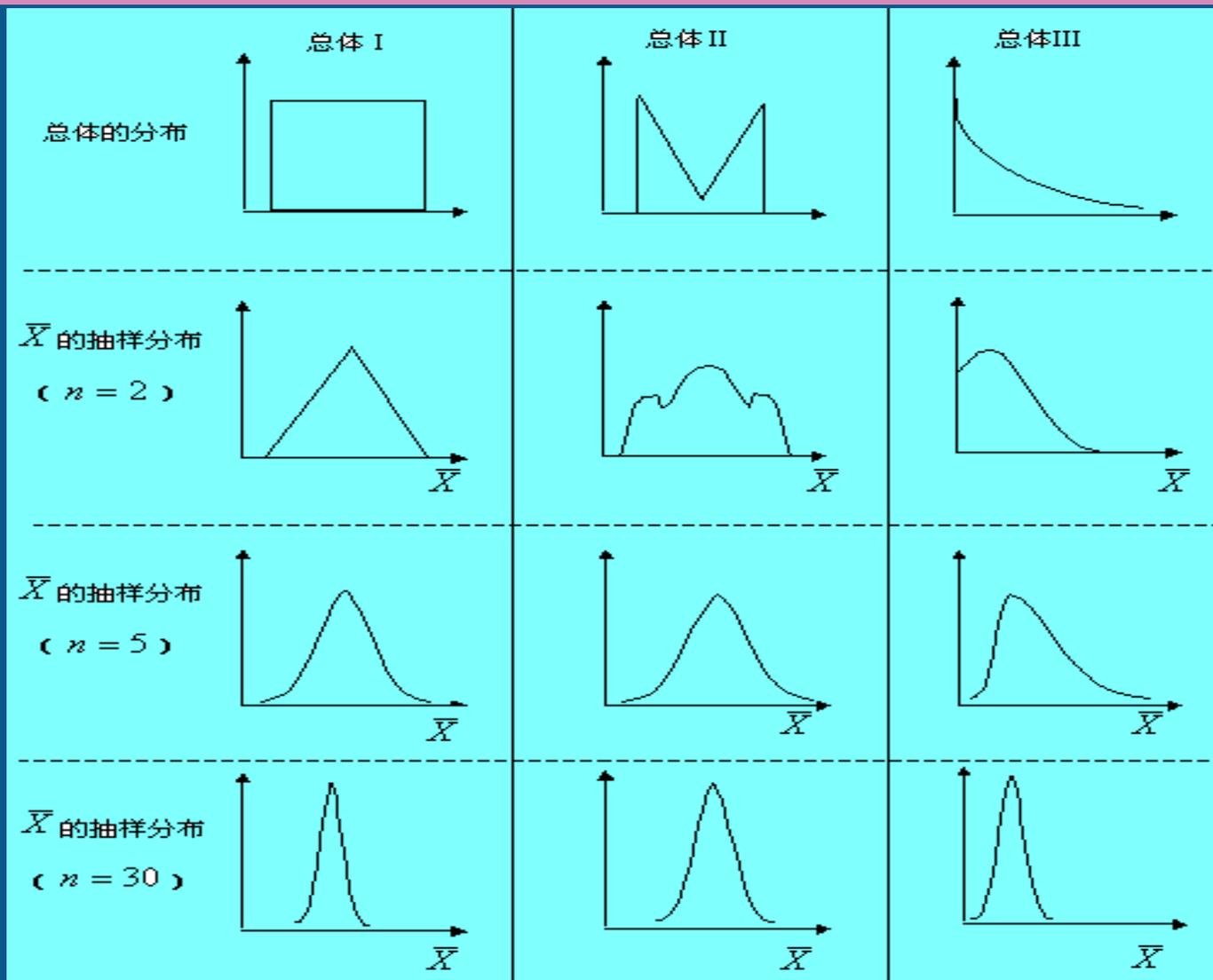
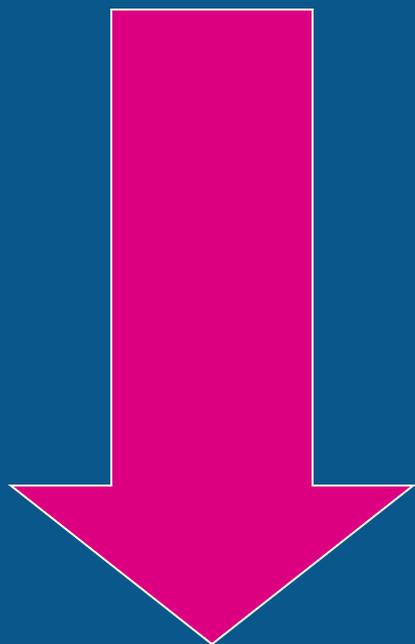
中心极限定理 (central limit theorem)

从均值为 μ ，方差为 σ^2 的一个任意总体中抽取容量为 n 的样本，当 n 充分大时，样本均值的抽样分布近似服从均值为 μ 、方差为 σ^2/n 的正态分布



中心极限定理 (central limit theorem)

\bar{x} 的分布趋于正态分布的过程



本章小结

1. 统计量及其分布
2. 由正态分布导出的几个重要分布
3. 样本均值的分布与中心极限定理

结 束



THANKS