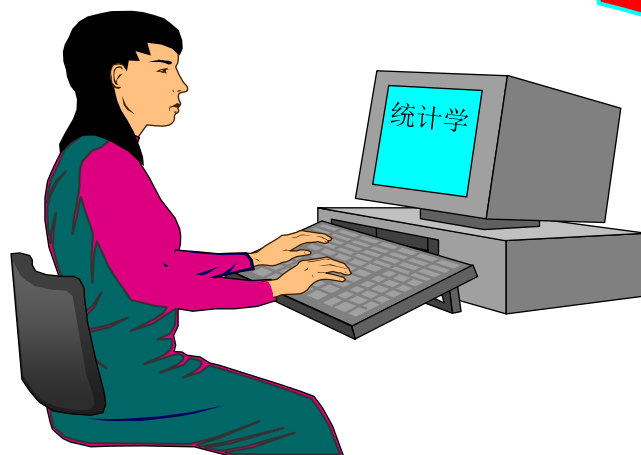


第 5 章 抽样分布

PowerPoint



第 4 章 抽样分布

1. 三种不同性质的分布
2. 一个总体参数推断时样本统计量分布
3. 两个总体参数推断时样本统计量分布

以统计量对总体参数作推论



研究对象总体



以样本计算统计量



抽出样本



推断统计要解决二个问题

1、抽样误差

2、推断正确的可靠性

要解决上述二个问题，样本统计量最好是正态分布。

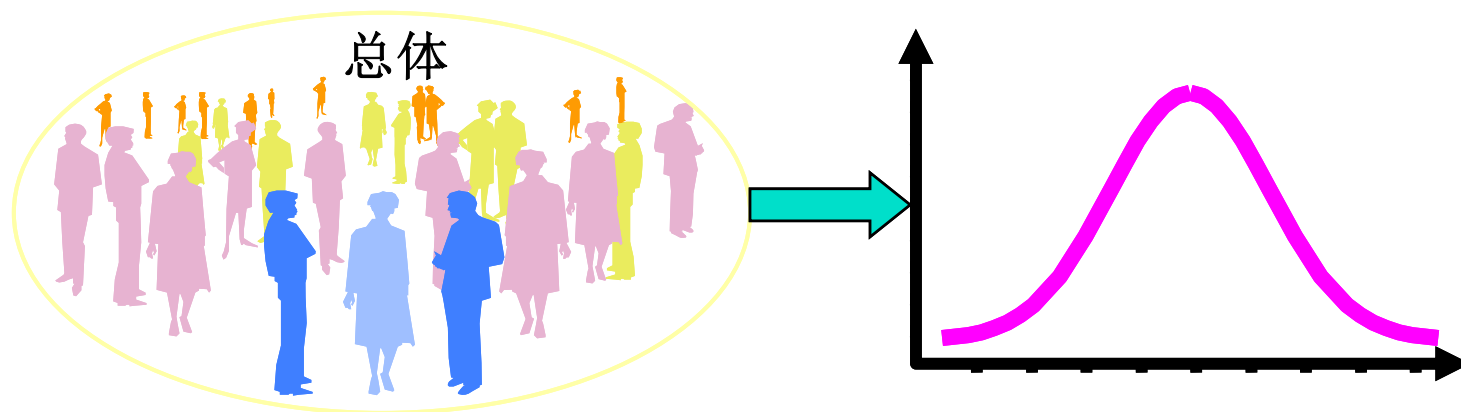
本章要阐述样本统计量在什么条件下服从正态分布。

三种不同性质的分布

- 一. 总体分布
- 二. 样本分布
- 三. 抽样分布

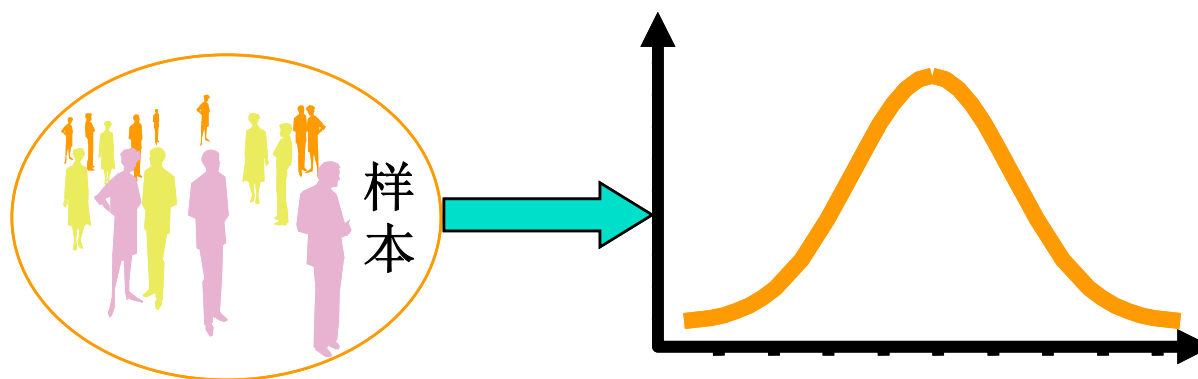
总体分布

1. 总体中各元素的观察值所形成的分布
2. 分布通常是未知的
3. 可以假定它服从某种分布



样本分布

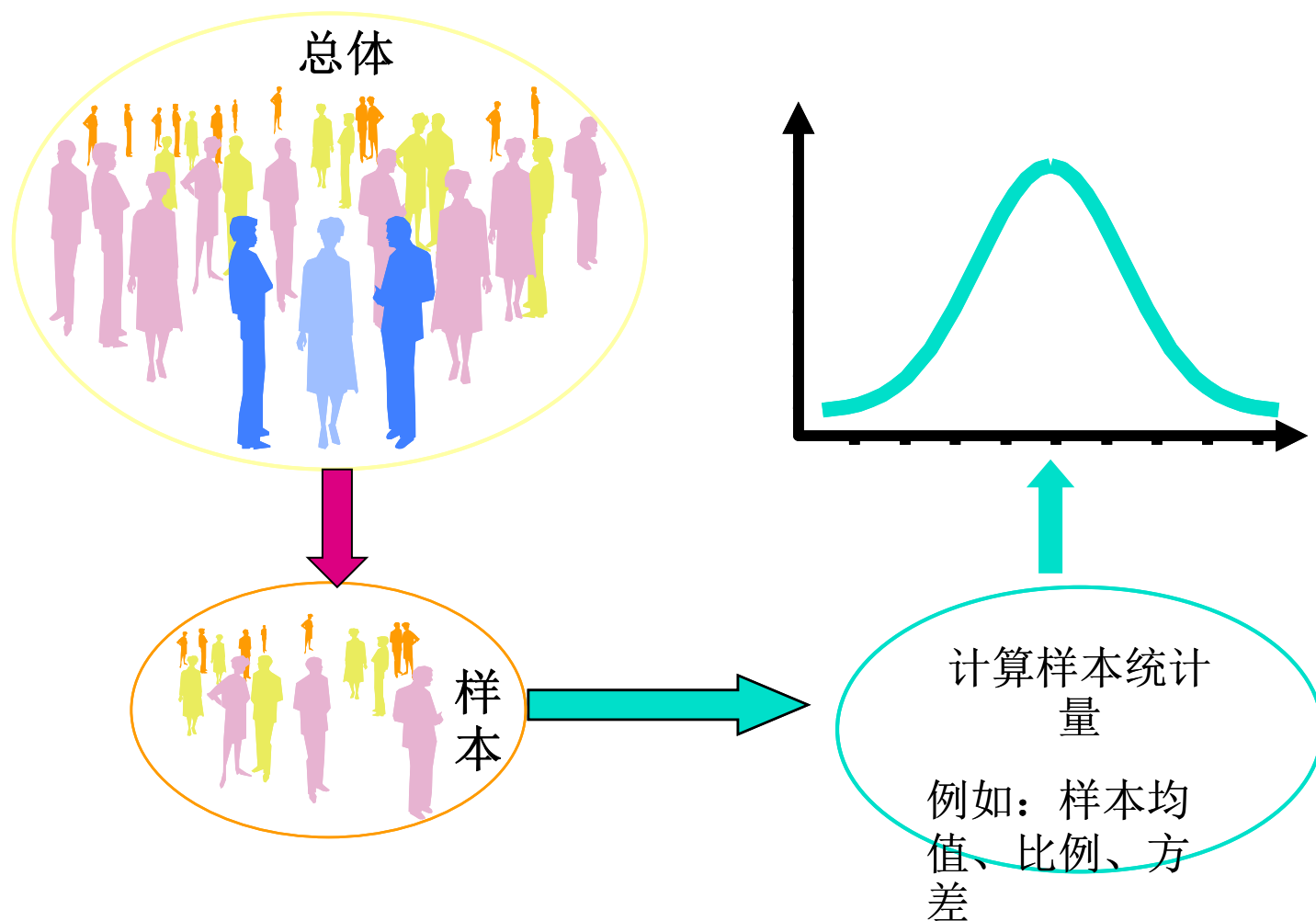
1. 一个样本中各观察值的分布
2. 当样本容量 n 逐渐增大时，样本分布逐渐接近总体的分布



抽样分布

1. 样本统计量的概率分布
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

抽样分布的形成过程



样本统计量的抽样分布 (一个总体参数推断时)

- 一. 样本均值的抽样分布
- 二. 样本比例的抽样分布

样本均值的抽样分布

样本均值的抽样分布

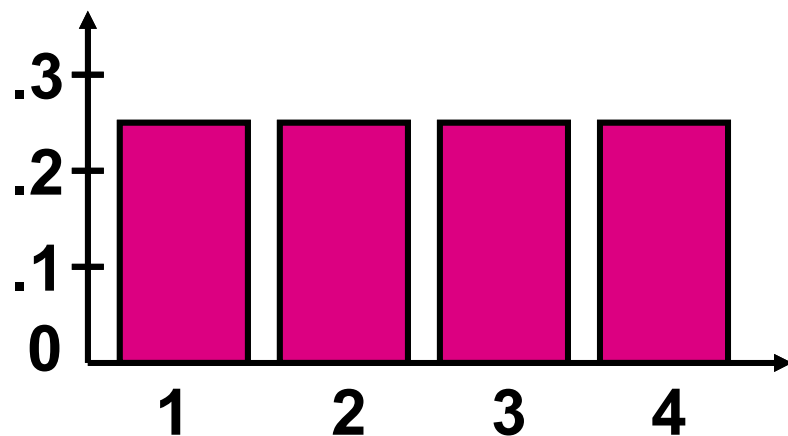
1. 容量相同的所有可能样本的样本均值的概率分布
2. 进行推断总体总体均值 μ 的理论基础

样本均值的抽样分布

(例题分析)

【例】设一个总体，含有4个元素(个体)，即总体单位数 $N=4$ 。4个个体分别为 $x_1=1$ 、 $x_2=2$ 、 $x_3=3$ 、 $x_4=4$ 。总体的均值、方差及分布如下

总体分布



均值和方差

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = 2.5$$

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = 1.25$$

样本均值的抽样分布

(例题分析)

➡ 现从总体中抽取 $n=2$ 的简单随机样本，在重复抽样条件下，共有 $4^2=16$ 个样本。所有样本的结果为

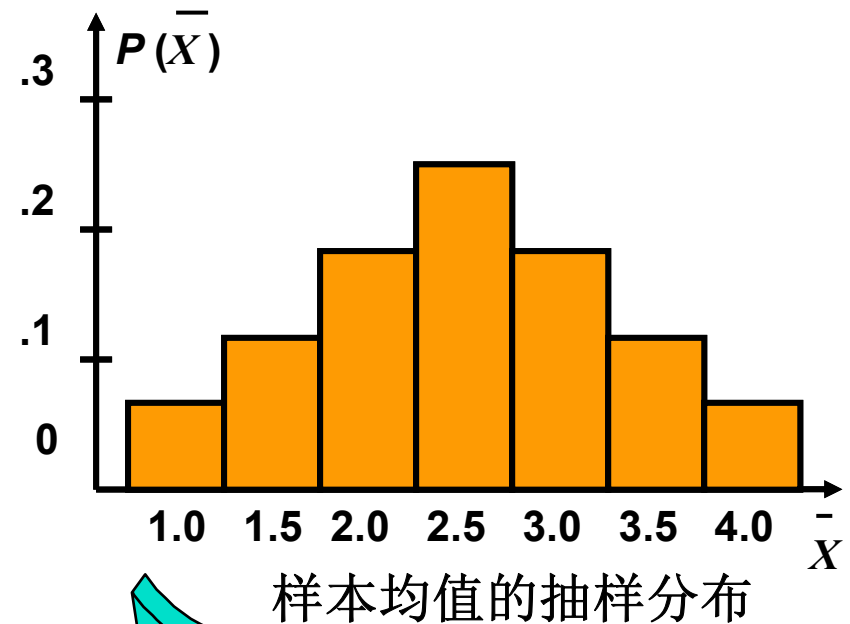
所有可能的 $n = 2$ 的样本（共16个）				
第一个 观察值	第二个观察值			
	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

样本均值的抽样分布

(例题分析)

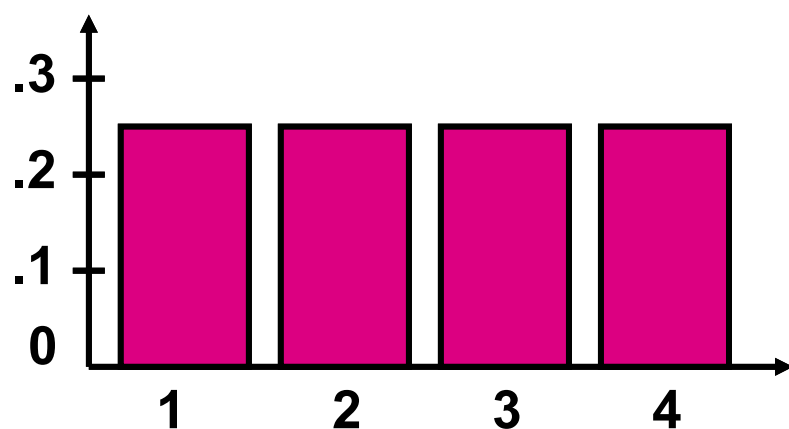
➡ 计算出各样本的均值，如下表。并给出样本均值的抽样分布

16个样本的均值 (\bar{x})				
第一个观察值	第二个观察值			
	1	2	3	4
1	1.0	1.5	2.0	2.5
2	1.5	2.0	2.5	3.0
3	2.0	2.5	3.0	3.5
4	2.5	3.0	3.5	4.0



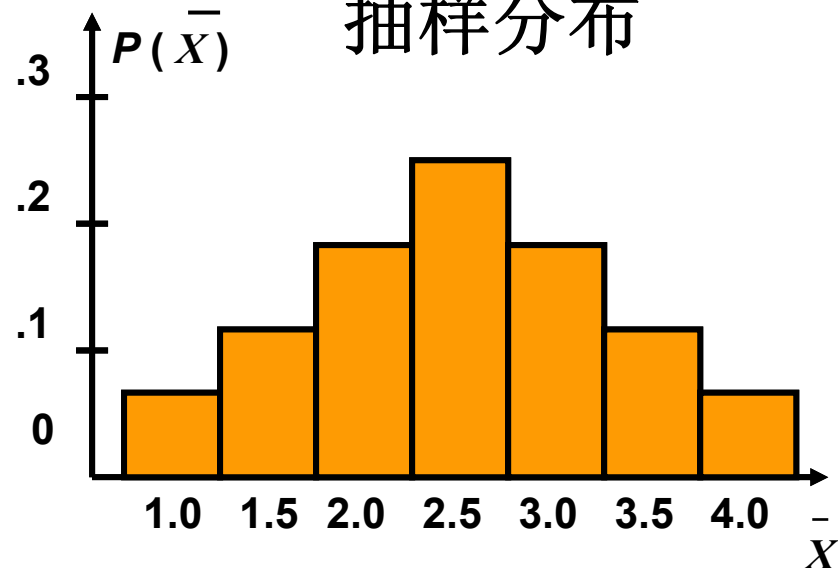
样本均值的分布与总体分布的比较 (例题分析)

总体分布



$$\mu = 2.5$$
$$\sigma^2 = 1.25$$

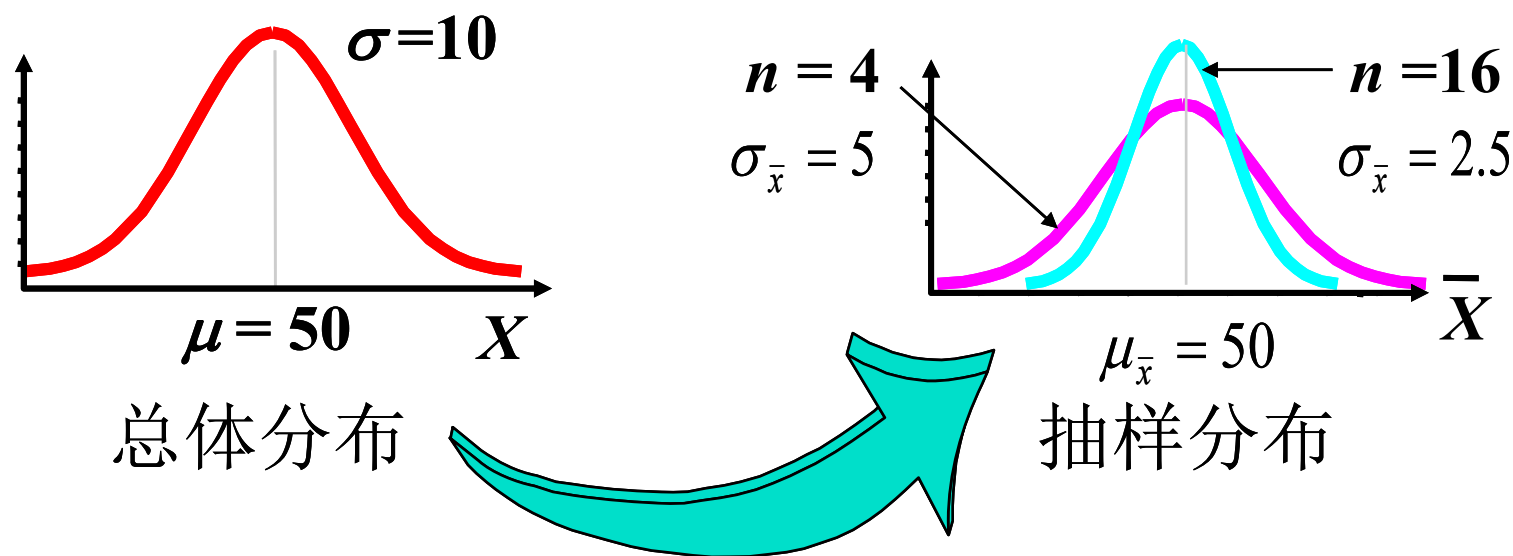
抽样分布



$$\mu_{\bar{X}} = 2.5$$
$$\sigma_{\bar{X}}^2 = 0.625$$

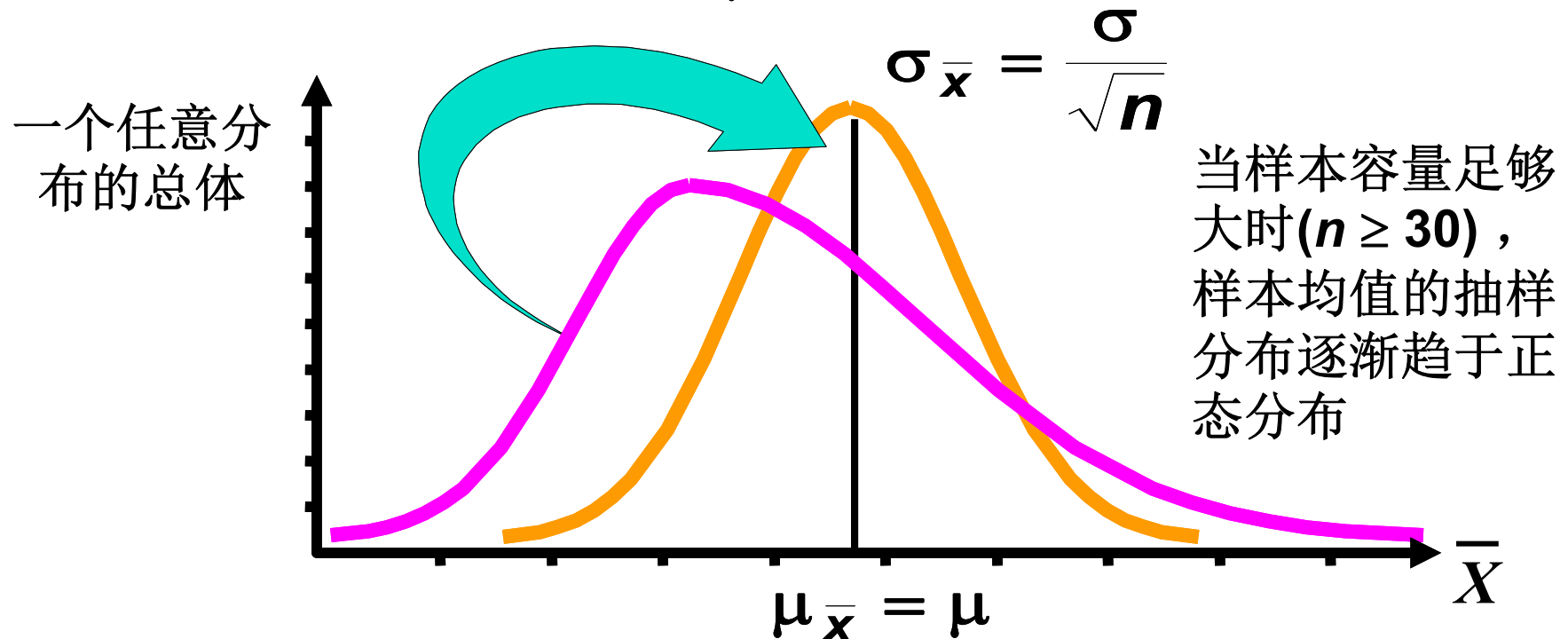
样本均值的抽样分布 与中心极限定理

当总体服从正态分布 $X \sim N(\mu, \sigma^2)$ 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 $\bar{X} \sim N(\mu, \sigma^2/n)$



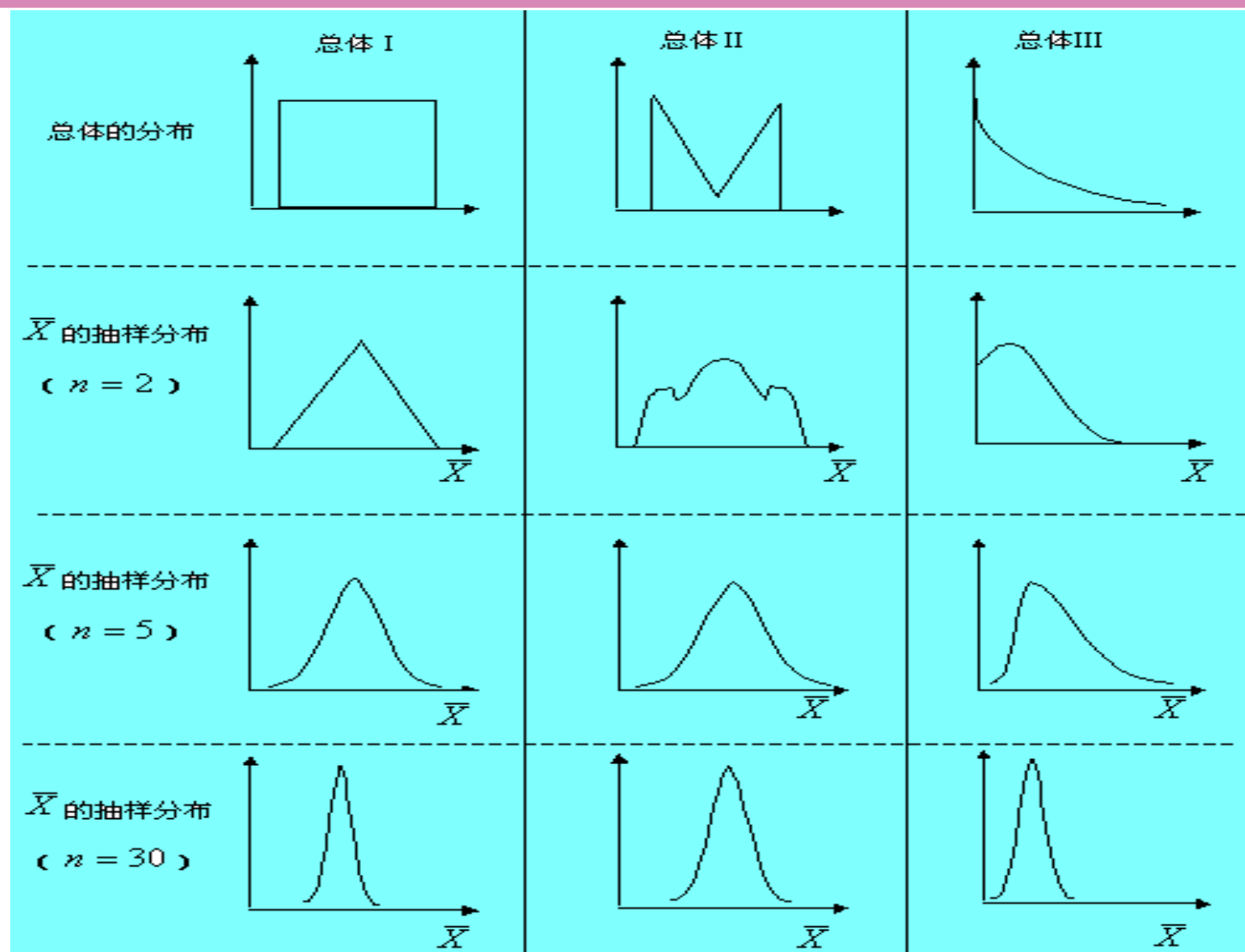
中心极限定理

中心极限定理： 设从均值为 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

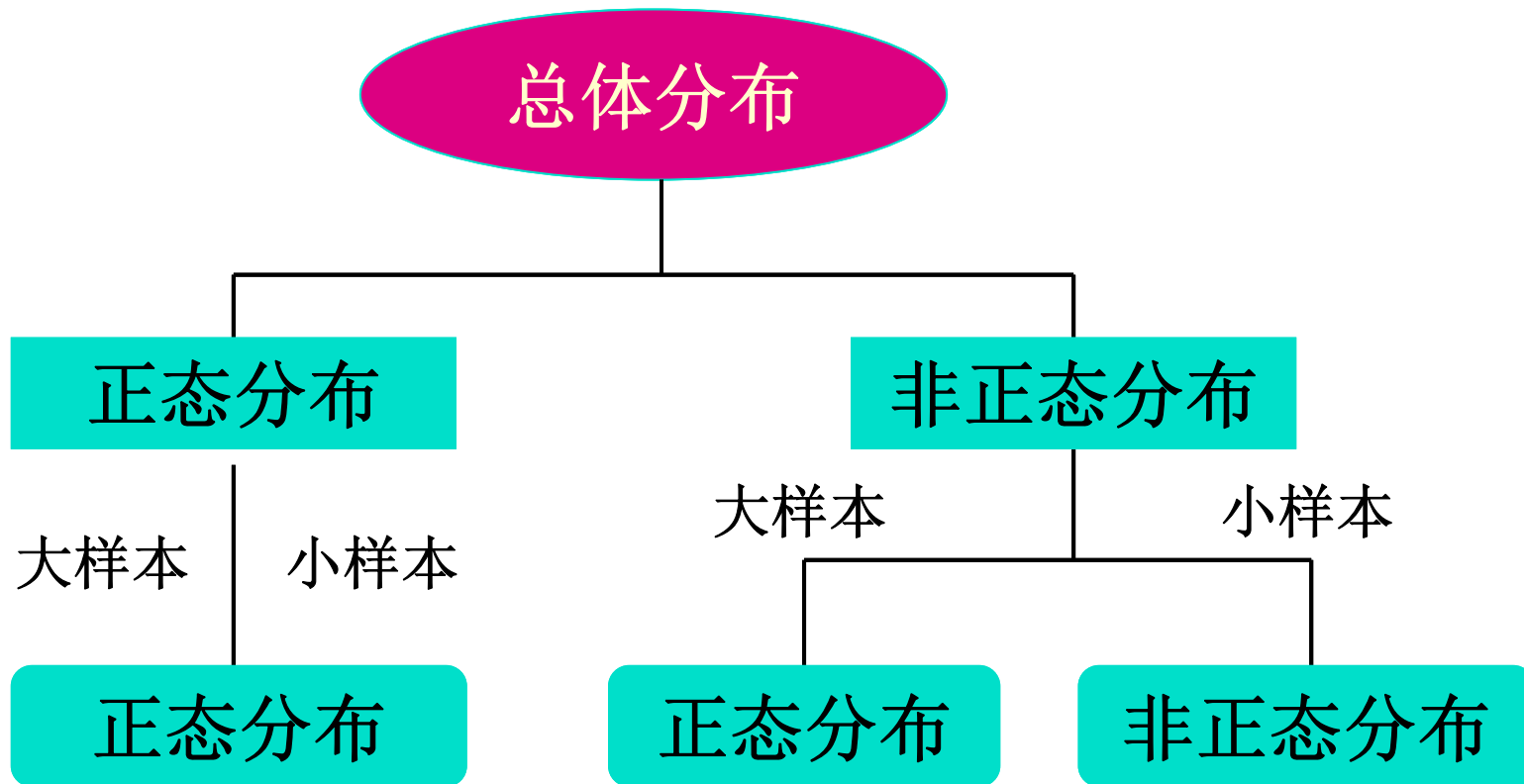


中心极限定理

\bar{X}
的分布趋于正态分布的过程



抽样分布与总体分布的关系



样本均值的抽样分布

(数学期望与方差)

1. 样本均值的数学期望

$$E(\bar{X}) = \mu$$

2. 样本均值的方差

- 重复抽样 $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$
- 不重复抽样 $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$

修正系数

抽样比小于**5%**则修正系数可以忽略不计

样本均值的抽样分布

(数学期望与方差)

$$\mu_{\bar{X}} = \frac{\sum_{i=1}^n \bar{X}_i}{M} = \frac{1.0 + 1.5 + \dots + 4.0}{16} = 2.5 = \mu$$
$$\sigma_{\bar{X}}^2 = \frac{\sum_{i=1}^n (\bar{X}_i - \mu_{\bar{X}})^2}{M}$$
$$= \frac{(1.0 - 2.5)^2 + \dots + (4.0 - 2.5)^2}{16} = 0.625 = \frac{\sigma^2}{n} \quad M \text{为样本数目}$$

- 比较及结论：
1. 样本均值的均值(数学期望) 等于总体均值
 2. 样本均值的方差等于总体方差的 $1/n$

均值的抽样标准误

1. 所有可能的样本均值的标准差，测度所有样本均值的离散程度
2. 小于总体标准差
3. 计算公式为

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

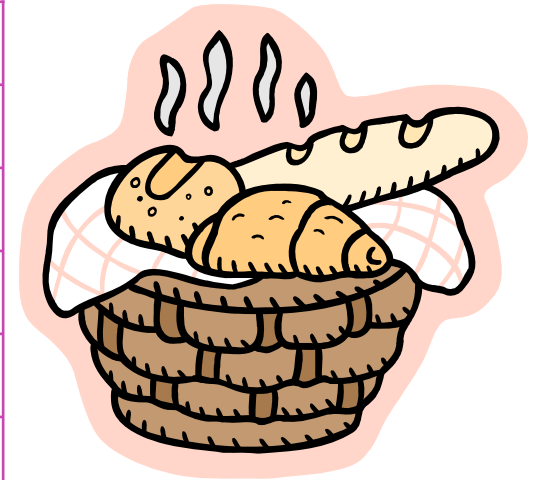
思考：

不重复抽样条件下，均值抽样标准误（
抽样平均误差）的计算公式

均值的抽样标准误 (例题分析)

1. 一家食品生产企业以生产袋装食品为主，为对产量质量进行监测，企业质检部门经常要进行抽检，以分析每袋重量是否符合要求。现从某天生产的一批食品中随机抽取了**25**袋，测得每袋重量如下表所示。已知产品重量的分布服从正态分布，且总体标准差为**10g**。**计算均值的抽样标准误。**

25袋食品的重量				
112.5	101.0	103.0	102.0	100.5
102.6	107.5	95.0	108.8	115.6
100.0	123.5	102.0	101.6	102.2
116.6	95.4	97.8	108.6	105.0
136.8	102.8	101.5	98.4	93.3



均值的抽样标准误 (例题分析)

解：已知 $X \sim N(\mu, 10^2)$ ，根据样本数据计算得：

$$\bar{x} = 105.36$$

$$\text{抽样标准误} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

均值的抽样标准误 (例题分析)

一家保险公司收集到由**36**投保个人组成的随机样本，得到每个投保人的年龄(周岁)数据如下表。

试计算均值的抽样标准误。

36个投保人年龄的数据					
23	35	39	27	36	44
36	42	46	43	31	33
42	53	45	54	47	24
34	28	39	36	44	40
39	49	38	34	48	50
34	39	45	48	45	32



均值的抽样标准误 (例题分析)

解：根据样本数据计算得：

$$\bar{x} = 39.5 \quad s = 7.77$$

$$\text{均值的抽样标准误} = \frac{s}{\sqrt{n}} = \frac{7.77}{\sqrt{36}} = 1.295$$

样本比例的抽样分布

比例

1. 总体(或样本)中具有某种属性的单位与全部单位总数之比

- 不同性别的人与全部人数之比
- 合格品(或不合格品) 与全部产品总数之比

2. 总体比例可表示为

$$\pi = \frac{N_0}{N} \quad \text{或} \quad 1 - \pi = \frac{N_1}{N}$$

3. 样本比例可表示为

$$P = \frac{n_0}{n} \quad \text{或} \quad 1 - P = \frac{n_1}{n}$$

样本比例的抽样分布

1. 容量相同的所有可能样本的样本比例的概率分布
2. 当样本容量很大时，样本比例的抽样分布可用正态分布近似
3. 推断总体总体比例 π 的理论基础

样本比例的抽样分布

(数学期望与方差)

1. 样本比例的数学期望

$$E(P) = \pi$$

2. 样本比例的方差

- 重复抽样 $\sigma_P^2 = \frac{\pi(1-\pi)}{n}$
- 不重复抽样 $\sigma_P^2 = \frac{\pi(1-\pi)}{n} \left(\frac{N-n}{N-1} \right)$

思考：

比例的抽样标准误？

比率的抽样标准误 (例题分析)

【例】某城市想要估计下岗职工中女性所占的比率，随机地抽取了**100**名下岗职工，其中**65**人为女性职工。试计算样本下岗职工中女性比率的**抽样标准误**

解：已知 $n=100$ ， $p=65\%$

$$\begin{aligned}\text{抽样标准误} &= \sqrt{\frac{p(1-p)}{n}} \\ &= \sqrt{\frac{65\%(1-65\%)}{100}} \\ &= 4.77\%\end{aligned}$$



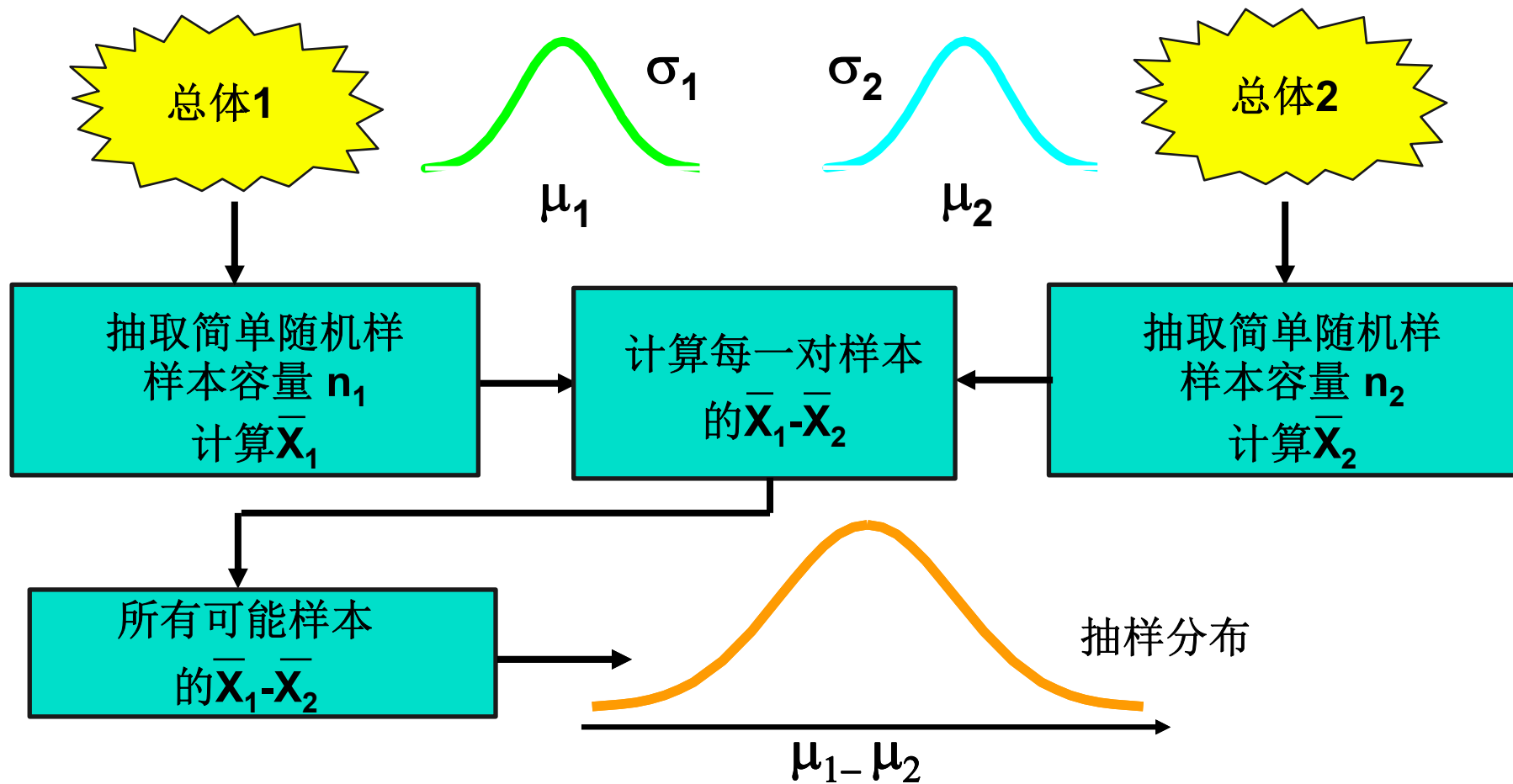
样本统计量的抽样分布

(两个总体参数推断时)

- 一. 两个样本均值之差的抽样分布
- 二. 两个样本比例之差的抽样分布

两个样本均值之差的抽样分布

两个样本均值之差的抽样分布



两个样本均值之差的抽样分布

1. 两个总体都为正态分布，即 $X_1 \sim N(\mu_1, \sigma_1^2)$ ，

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

2. 两个样本均值之差 $\bar{X}_1 - \bar{X}_2$ 的抽样分布服从正态分布，其分布的数学期望为两个总体均值之差

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$$

方差为各自的方差之和

$$\sigma_{\bar{X}_1 - \bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

两个样本比例之差的抽样分布

两个样本比例之差的抽样分布

1. 两个总体都服从二项分布
2. 分别从两个总体中抽取容量为 n_1 和 n_2 的独立样本，当两个样本都为大样本时，两个样本比例之差的抽样分布可用正态分布来近似
3. 分布的数学期望为

$$E(P_1 - P_2) = \pi_1 - \pi_2$$

方差为各自的方差之和

$$\sigma_{P_1 - P_2}^2 = \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}$$

抽样推断可靠性

把正态分布变为标准的正态分布

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

$$z_i = \frac{x_i - \bar{x}}{s}$$

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

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

抽样推断可靠性

把正态分布变为标准的正态分布

$$p \sim N(\quad)$$

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} \sim N(0,1)$$

中心极限定理的应用 (1)

假设某次高考数学为成绩正态分布，平均为**65**分，标准差为**12**分，要求计算 (1) 随机抽取一人，该人成绩在**77**分以上的概率； (2) 随机抽取**9**人，这**9**人平均成绩在**77**分以上的概率。

中心极限定理的应用 (2)

某学校教学楼的电梯可以容纳20人，规定限重1500公斤，假设该学校使用该电梯的师生体重服从正态分布，平均为68公斤，标准差为12公斤，计算电梯中乘满20人而重量超过1500公斤的概率。

中心极限定理的应用 (3)

某城市居民拥有信用卡的比例为10%，随机抽取500人，拥有信用卡在60人以上的概率有多大？

本章小结

- 1. 总体分布、样本分布、抽样分布**
- 2. 单总体参数推断时样本统计量的分布**
- 3. 双总体参数推断时样本统计量的分布**

一个具有 **$n=900$** 个观测值的随机样本选自
 $\mu=100$ 和 **$\sigma=10$** 的总体。

- 1、你预计样本均值的最大值和最小值是什么？
- 2、你认为样本均值至多偏离 **μ** 有多远？

美国汽车联合会**1999**年**5**月通过对会员调查得知一个**4**口之家出游中平均每日餐饮和住宿费用大约是**213**美元。假设这个花费的标准差是**15**美元，并且美国汽车联合会调查的每日消费是总体均值。又假设抽选**49**个**4**口之家，并对其在**1999**年**6**月期间的旅行费用进行记录。

- 1、描述样本均值的抽样分布；
- 2、对于样本家庭来说，平均每日消费大于**213**美元的概率是什么？

大于**217**美元的概率呢？在**209**美元和**217**美元之间的概率呢？

技术人员对奶粉装袋过程进行质量检验。每袋平均重量标准为**406**克，标准差为**10.1**克。监控这一过程的技术人员每天随机抽取**36**袋，并对每袋重量进行测量。现考虑这**36**袋奶粉组成样本的平均重量。

- 1、描述样本均值的抽样分布；
- 2、样本均值小于**400.8**克的概率；
- 3、假设某一天技术人员观察到样本均值为**400.8**克，这是否意味装袋过程出现问题？为什么？