

第一讲-2 抽样分布与参数估计



本讲内容

■ 有关概念

■ 参数估计

◆ 单一总体均值估计

■ 确定样本容量



有关概念

❖ 参数与统计量

❖ 统计误差



参数与统计量

- ❖ 参数：反应总体分布特征的指标统称为总体参数，简称参数。常用的有

$$\mu, \sigma^2, \sigma \text{ 和 } \pi$$

- ❖ 统计量：反应样本分布特征的指标统称为样本统计量，简称统计量。常用

$$\bar{x}, s^2, s, p$$



总体参数

样本统计量

平均数

$$\mu = \frac{\sum X}{N}$$

$$\bar{x} = \frac{\sum x}{n}$$

比率

$$\pi = \frac{N_1}{N}$$

$$p = \frac{n_1}{n}$$

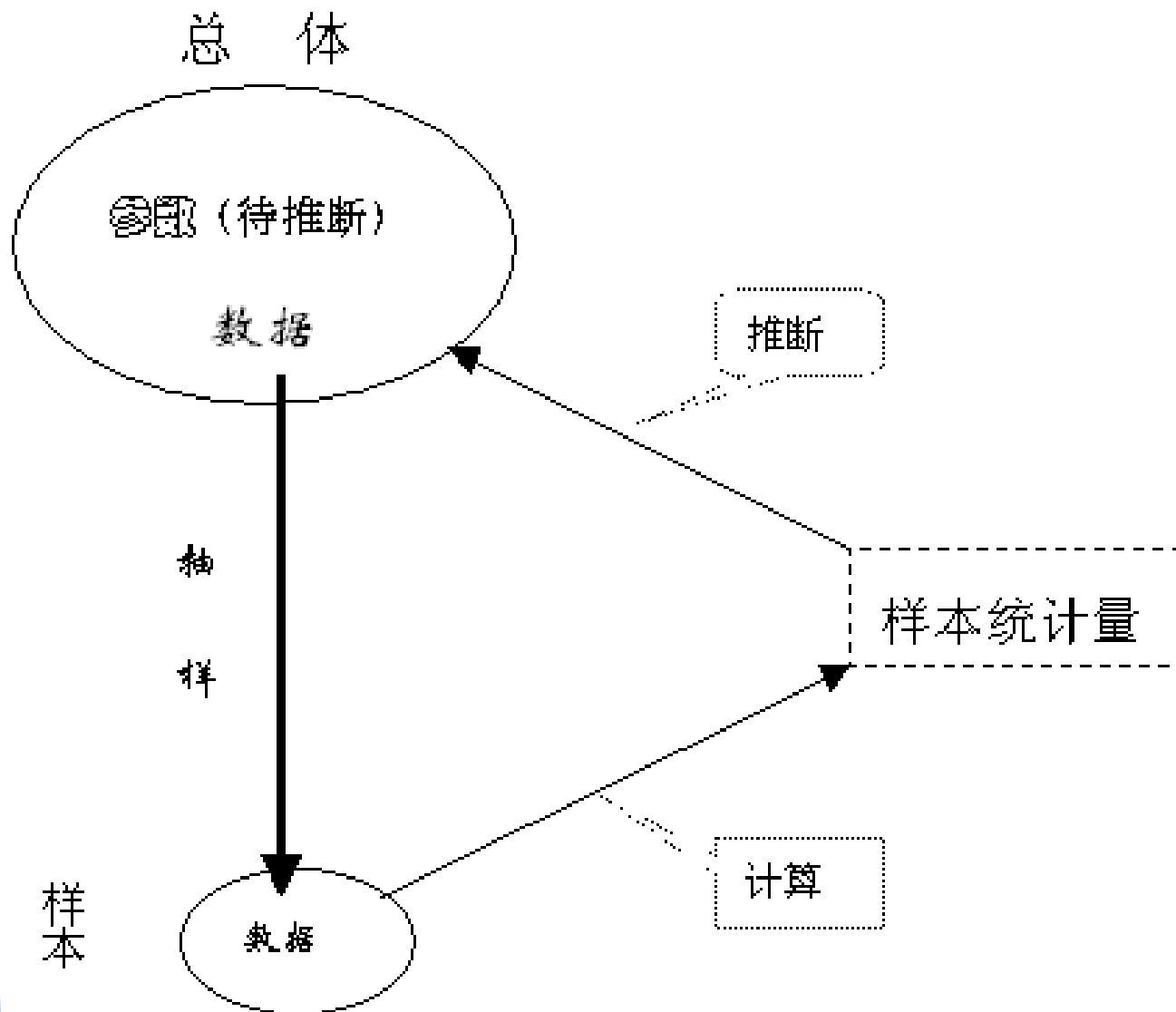
方差

$$\sigma^2(X) = \frac{\sum (X - \mu)^2}{N}$$

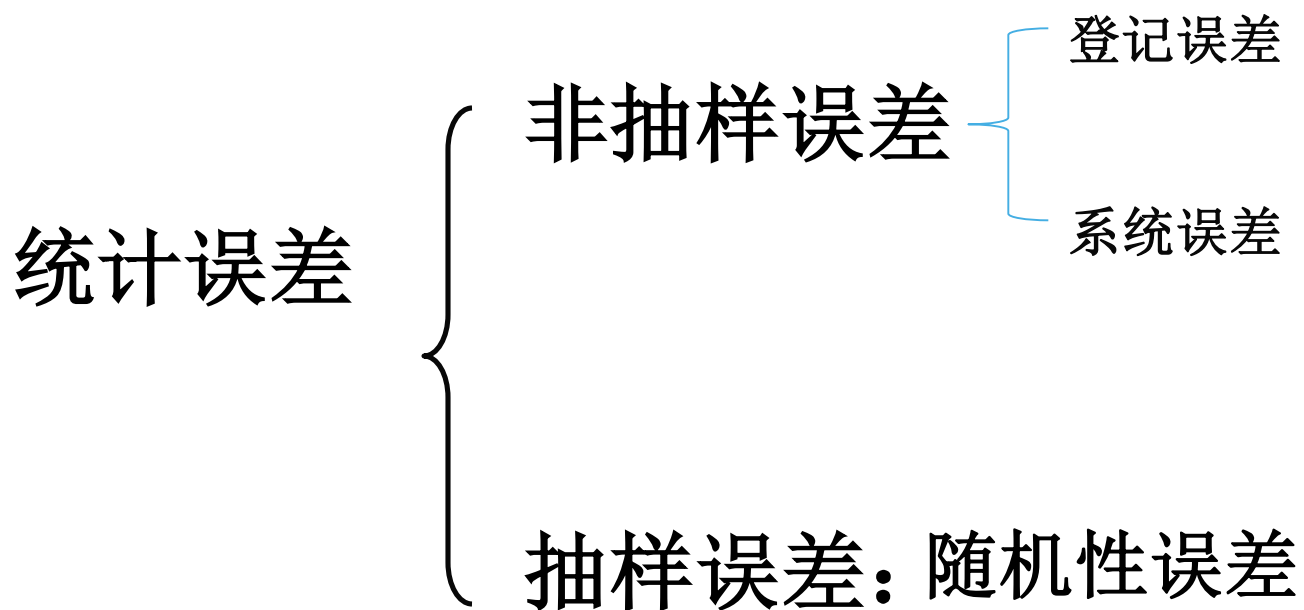
$$s^2(x) = \frac{\sum (x - \bar{x})^2}{n - 1}$$



抽样分布的流程



统计误差



参数估计

❖ 估计方法

- 点估计
- 区间估计

❖ 单一总体均值估计



参数估计的方法

- 点估计
- 区间估计



点估计

1. 用样本的估计量的某个取值直接作为总体参数的估计值

- 例如：用样本均值直接作为总体均值的估计；用两个样本均值之差直接作为总体均值之差的估计

2. 无法给出估计值接近总体参数程度的信息

- 由于样本是随机的，抽出一个具体的样本得到的估计值很可能不同于总体真值
- 一个点估计量的可靠性是由它的抽样标准误差来衡量的，这表明一个具体的点估计值无法给出估计的可靠性的度量



区间估计

- 区间估计的定义
- 区间估计的原理
- 区间估计的程序
- 单一总体平均数的区间估计



区间估计的定义

- ◆ 区间估计是在一定的置信系数的保证下，根据统计量得到的一个取值范围去估计总体的参数。

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha (0 < \alpha < 1)$$

$(\hat{\theta}_1, \hat{\theta}_2)$ 为 θ 的置信区间, $1 - \alpha$ 为置信度, $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别为置信下限和置信上限



区间估计的几个关键概念

- ❖ 置信系数 $(1-\alpha)$ 使人相信区间包含总体均值的概率,一般取 **0.95,0.90,0.99**.它的大小说明估计的把握性的大小.
- ❖ 置信区间:在一定概率的保证下,包含总体均值的区间,区间的宽窄说明估计精度的大小.区间越宽,估计的精度就小;否则就大.
- ❖ 临界值:置信区间的上限和下限
- ❖ 注意置信系数和区间宽窄的关系



置信水平（置信系数）

1. 将构造置信区间的步骤重复很多次，置信区间包含总体参数真值的次数所占的比例，也称置信度（置信水平）
2. 表示为 $(1 - \alpha) \%$
3. 常用的置信水平值有 99%, 95%, 90%



总体均值区间的一般表达式

1. 总体均值的置信区间是由样本均值加减估计误差得到的
2. 估计误差由两部分组成：一是点估计量的**标准误差**，它取决于样本统计量的抽样分布。二是估计时所要求**置信水平**，统计量分布两侧面积的分位数值，它取决于事先所要求的可靠程度
3. 总体均值在置信水平下的置信区间可一般性地表达为

样本均值 \pm 分位数值 \times 样本均值的标准误差



当 σ 已知时计算 μ 的置信区间的步骤

- 选定置信系数 $1-\alpha$
- 抽取一个样本容量为 n 的样本
- 计算 \bar{X}
- 确定 \bar{X} 统计量的概率分布
- 求置信区间的临界值
- 单一总体平均数的置信区间的临界值为

$$\bar{X} \pm z\sigma_{\bar{X}}$$



总体均值的区间估计

【例】一家保险公司收集到由36个投保人组成的随机样本，得到每个投保人的年龄(单位：周岁)数据如下表。试建立投保人年龄90%的置信区间

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



总体均值的区间估计

解：已知 $n=36$, $1-\alpha = 90\%$, $z_{\alpha/2}=1.65$ 。根据样本数据计算得： $\bar{x} = 39.5$ $s = 7.77$

总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\begin{aligned}\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} &= 39.5 \pm 1.65 \times \frac{7.77}{\sqrt{36}} \\ &= 39.5 \pm 2.14 \\ &= (37.36, 41.64)\end{aligned}$$

投保人平均年龄的置信区间为37.36岁~41.64岁



样本容量的确定

●估计总体均值时

根据均值区间估计公式可得样本容量 n 为

$$n = \frac{Z^2 \cdot \sigma^2}{e^2}$$



例：根据某城市随机抽样调查结果，被调查的家庭在过去的一年中耐用消费品的购买额均值为**450**元。另根据经验估计的标志差为**120**。如果置信系数为**0.95**且误差在**± 4.5**户以内，问样本量应确定为多少？



Thank You !