



浙江财经大学
Zhejiang University of Finance & Economics

抽样分布



教师: 李金昌
浙江财经大学数据科学学院



抽样分布

一、抽样分布的基本问题



总体分布

总体中所有个体关于某个变量的取值所形成的分布。

均值

标准差



样本分布

样本中所有个体关于某个变量的取值所形成的分布。



抽样分布

一、抽样分布的基本问题

抽样分布：样本统计量的概率分布，由样本统计量的所有可能取值和相应的概率组成。（关于样本分布的分布）

从容量为N的总体中抽取容量为n的样本最多可抽取m个样本，m个样本统计值形成的频率分布，即为抽样分布。

\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_k
π_1	π_2	π_3	π_k
π_i 为某一样本统计值出现的频率， $\sum \pi_i = 1, k \leq m$				

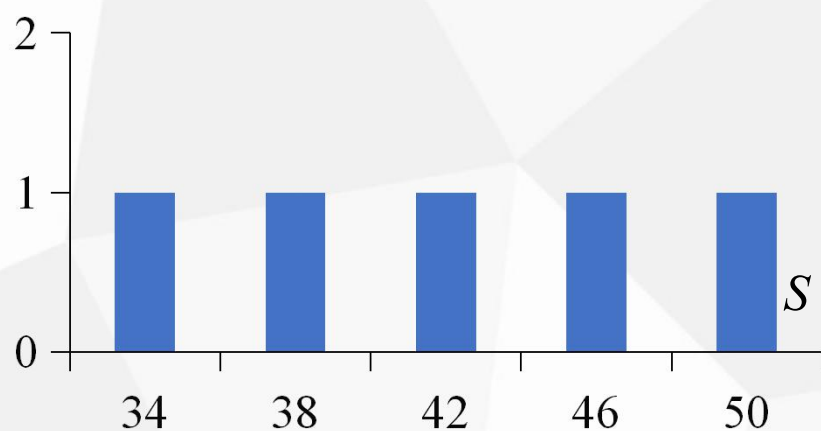


抽样分布

一、抽样分布的基本问题

举例

某车间班组5个工人的时工资为34, 38, 42, 46, 50元, 采用重复抽样方法从5个工人中随机抽取2人构成样本。



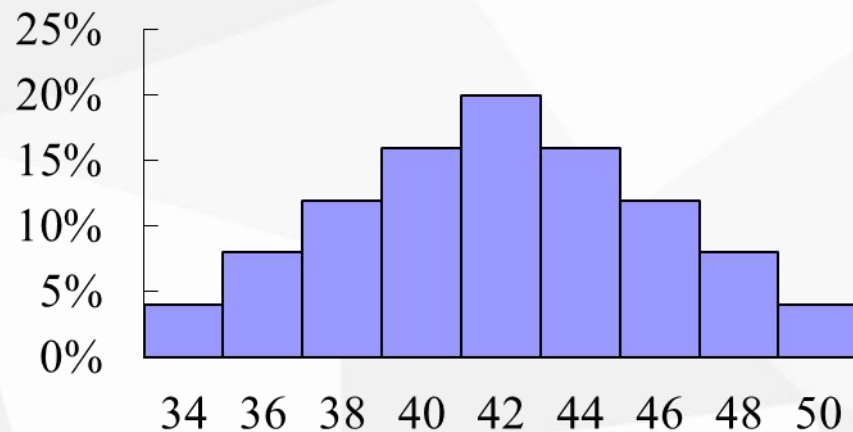
$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^N X_i}{N} = \frac{34 + 38 + 42 + 46 + 50}{5} = 42(\text{元}) \\ S^2 &= \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 \\ &= \frac{(-8)^2 + (-4)^2 + (0)^2 + 4^2 + 8^2}{5} = 32(\text{元}^2)\end{aligned}$$



抽样分布

一、抽样分布的基本问题

	34	38	42	46	50
34	34	36	38	40	42
38	36	38	40	42	44
42	38	40	42	44	46
46	40	42	44	46	48
50	42	44	46	48	50



$$\begin{aligned} E(\bar{x}) &= \frac{\sum \bar{x}_i f_i}{\sum f_i} = \sum \bar{x}_i \frac{f_i}{\sum f_i} \\ &= 34 \times 4\% + 36 \times 8\% + \cdots + 50 \times 4\% \\ &= 42(\text{元}) = \bar{X} \end{aligned}$$

$$\begin{aligned} Var(\bar{x}) &= \sum (\bar{x}_i - E(\bar{x}))^2 \pi_i \\ &= 16(\text{元}) \end{aligned}$$



抽样分布

一、抽样分布的基本问题

结论1：样本均值的平均值精确地等于总体均值；

结论2：样本均值抽样分布的散布范围比总体分布散布范围更小；

结论3：样本均值抽样分布的形状接近于正态分布。



抽样分布

二、抽样分布定理

1. 正态分布的再生定理

若总体服从正态分布 $X_i \sim N(\bar{X}, S^2)$ 则样本均值 \bar{x} 也服从正态分布。

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

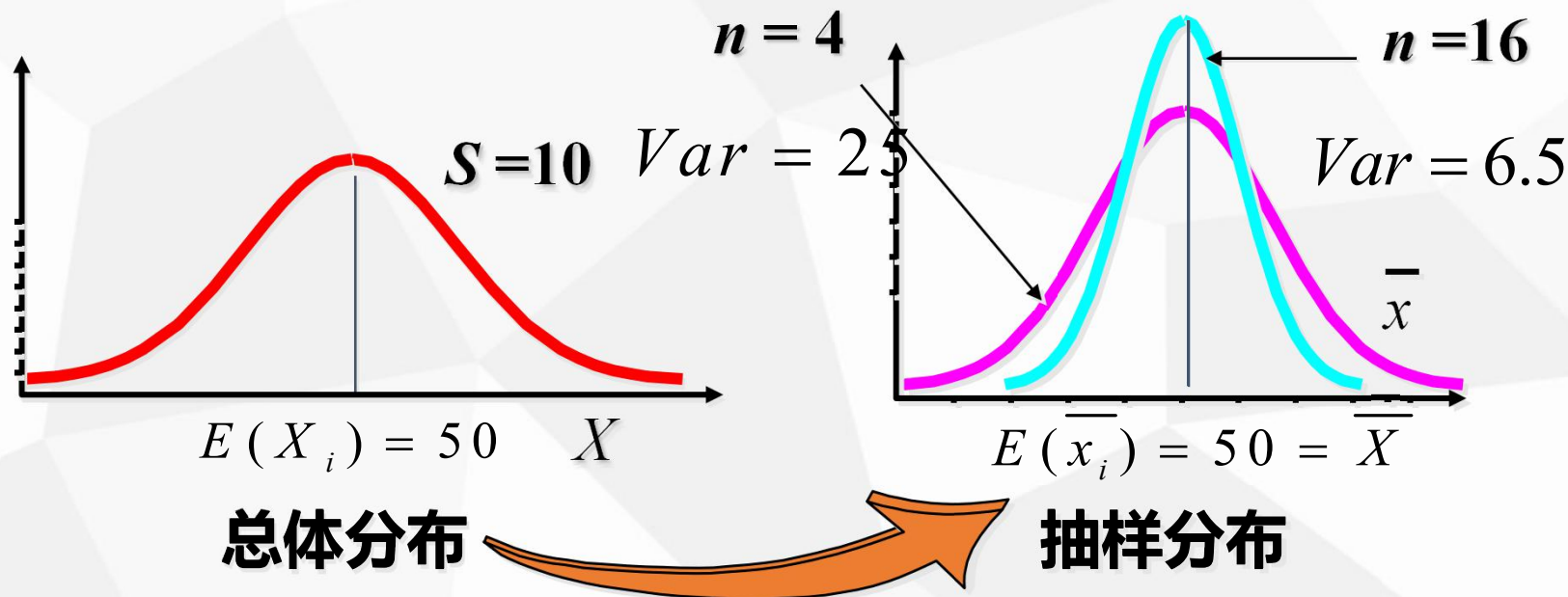
$$Var(\bar{x}) = \frac{S^2}{n} \text{ (重复抽样)}$$

$$Var(\bar{x}) = \frac{S^2}{n} \times (1 - f) \text{ (不重复抽样)}$$



抽样分布

二、抽样分布定理





抽样分布

二、抽样分布定理

2. 中心极限定理

任一总体（不要求正态），期望值 \overline{X} ，方差 S^2 ，当样本容量 n 足够大（当 $n \geq 30$ ，大样本）， \overline{x} 趋于服从正态分布。

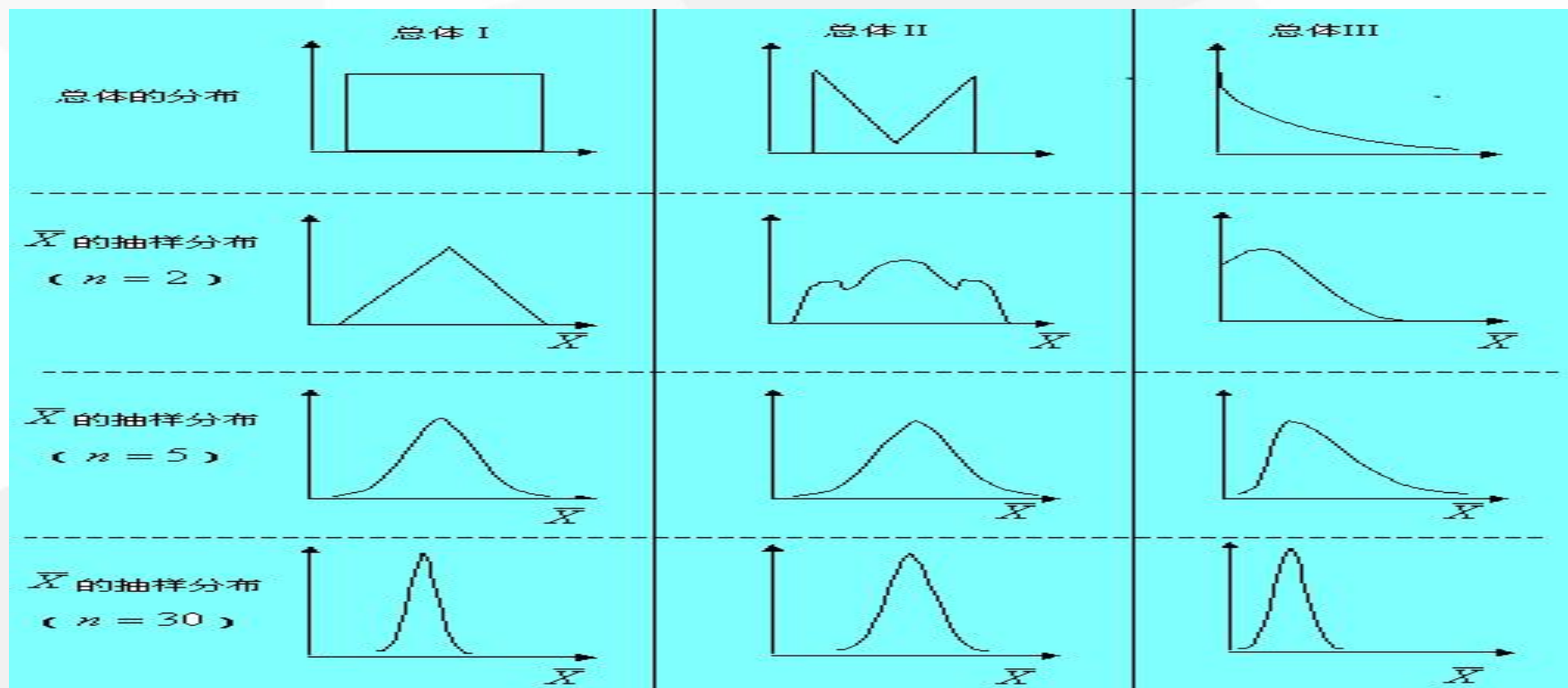
$$E(\overline{x}) = \overline{X} \quad \text{Var}(\overline{x}) = \frac{S^2}{n} \text{ (重复抽样)}$$

$$\text{Var}(\overline{x}) = \frac{S^2}{n} \times (1 - f) \text{ (不重复抽样)}$$



抽样分布

二、抽样分布定理





抽样分布

二、抽样分布定理

举例

某厂商产量数以亿计，声称其生产的产品具有均值为60个月，标准差为6个月的寿命分布。假设质检部门决定检验该厂的说法是否准确，为此随机抽取（不重复抽样）该厂生产的50件产品进行寿命试验。

请问：

- (1) 假定厂商声称是正确的，请描述50件产品的平均寿命的抽样分布；
- (2) 假定厂商声称是正确的，则50件产品组成的样本的平均寿命不超过57个月的概率是多少？



抽样分布

二、抽样分布定理

(1) 解:

**假定厂商声称是正确的，相当于已知均值为60和方差为 6^2 ；
总体容量未知，但数以亿计，样本容量为50，这样 $(N-n)/N$
接近1；**

可采用中心极限定理确定抽样分布特征：

平均寿命的均值为60月；

平均寿命的方差= $6^2/50=0.72$ ，标准差约为0.85月；

即，平均寿命服从均值60月，方差0.72的正态分布。



抽样分布

二、抽样分布定理

(2) 解:

假定厂商声称是正确的, 则50件产品组成的样本的平均寿命不超过57个月的概率是多少?

$$\begin{aligned} P(\bar{X} \leq 57) &= P\left(\frac{\bar{X} - 60}{0.85} \leq \frac{57 - 60}{0.85}\right) = P(Z \leq -3.529) \\ &= 1 - P(Z \leq 3.529) = 1 - \Phi(3.529) = 1 - 0.9998 \\ &= 0.0002 \end{aligned}$$