



# 社会调查方法

南京农业大学通识课

# 数据一定靠得住吗?

**讨论1：**有3份抽样数据：一份样本为**50**份，一份样本为**500**份，另一份样本为**50000**份，**你觉得哪一份数据更有说服力？**



# 第三章

## ➔ 抽样

---

### ➤ 抽样的介绍

### ➤ 概率抽样的原理与程序

### ➤ 概率抽样方法

### ➤ 非概率抽样方法

### ➤ 抽样实务及问题

## 一、与抽样相关的概念

- 总体——样本——抽样
- 抽样单位——抽样框
- 参数值——统计值
- 置信水平——置信区间

## 二、与抽样相关的概念（一）

总体（population）：

**总体是构成它的所有元素的集合，而元素则是构成总体的最基本单位。**

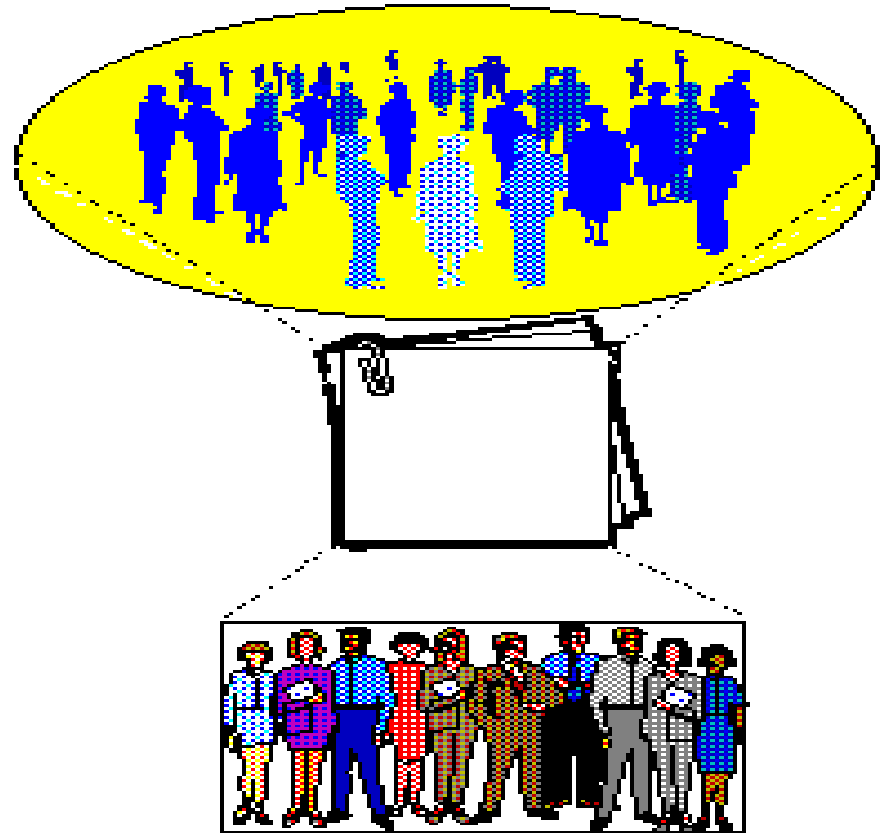
样本（sample）：

**样本是从总体中按一定方式抽取出一部分元素的集合。**

## 二、与抽样相关的概念（一）续

**抽样（sampling）：**

**抽样是从总体中按一定方式选择或抽取样本的过程。**



## 二、与抽样相关的概念（二）

抽样单位（sampling unit）：

**一次直接的**抽样所使用的基本单位。

抽样框（sampling frame）：

**一次直接抽样时总体中所有抽样单位的名单。**

## 二、与抽样相关的概念（三）

参数值（parameter）：

**总体**中所有元素的某种特征的综合数量表现。

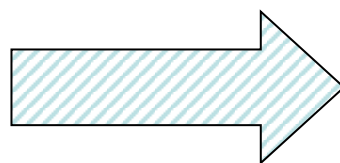
统计值（statistic）：

**样本**中所有元素的某种特征的综合数量表现。



## 样本统计值

平均值  $\bar{x}$   
标准差  $S$



## 总体参数值

平均值  $\mu$   
标准差  $\sigma$

不确定的、变化的  
对于特定样本是已知的

确定的、唯一的、  
未知的

## 二、与抽样相关的概念（四）

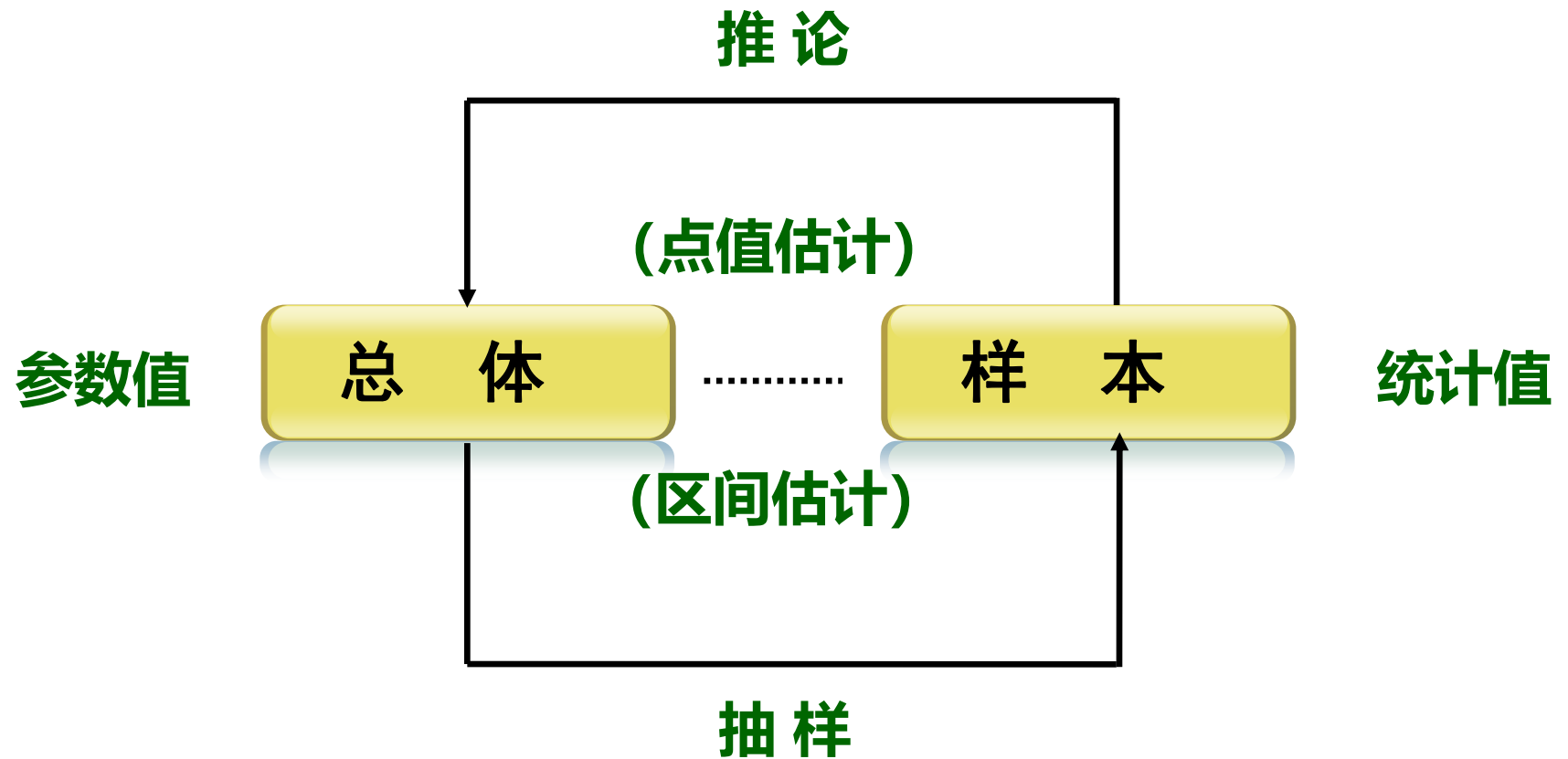
**置信水平（confidence level）：** **可靠性**

**总体参数值落在样本统计值某一区间内的概率。**

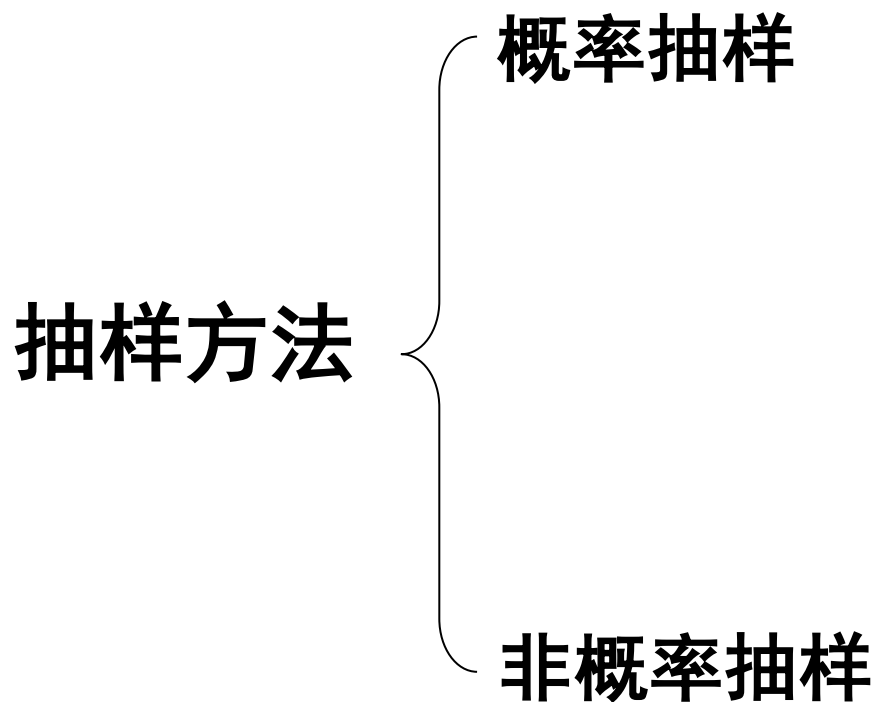
**置信区间（confidence interval）：** **精确性**

**指在一定的置信度下，样本统计值与总体参数值之间的误差范围。**

# 小结



### 三、抽样的类型



简单随机抽样

系统抽样

分层抽样

整群抽样

多段抽样

PPS抽样

户内抽样

偶遇抽样

判断抽样

定额抽样

雪球抽样

# 第三章

## ➔ 抽样

---

- ↗ 抽样的介绍
- ↗ 概率抽样的原理与程序
- ↗ 概率抽样方法
- ↗ 非概率抽样方法
- ↗ 抽样实务及问题

# 一、概率抽样的基本原理

## （一）同质性与异质性

牛说：水很浅才到我膝盖

狐狸说：水很深，差点淹死



# 一、概率抽样的基本原理

## （二）等概率抽样



## 二、抽样分布

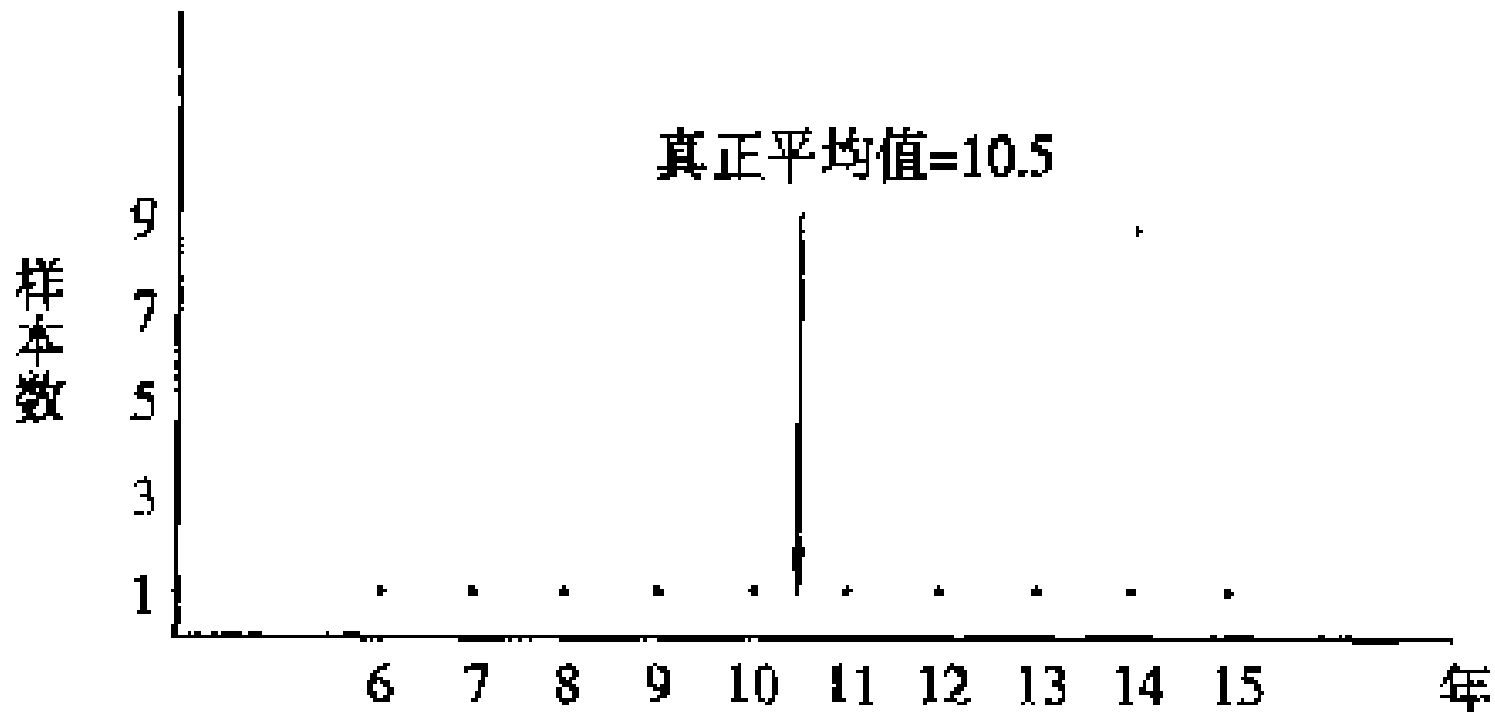


图 4-2 容量为 1 的样本的抽样分布(总数 = 10)



## 二、抽样分布

$$C_{10}^2 = \frac{10 \times 9}{2 \times 1} = 45$$

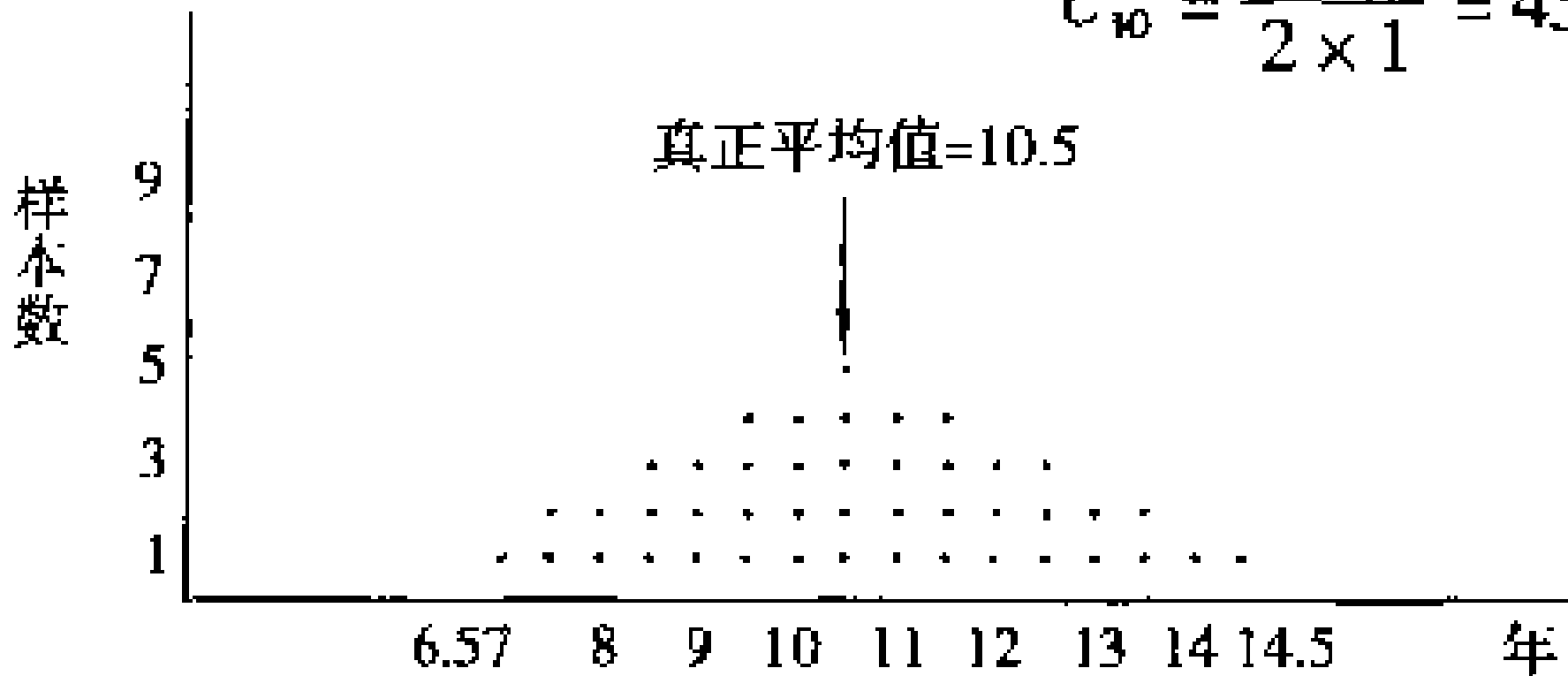


图 4-3 容量为 2 的样本的抽样分布(总数 = 45)

## 二、抽样分布

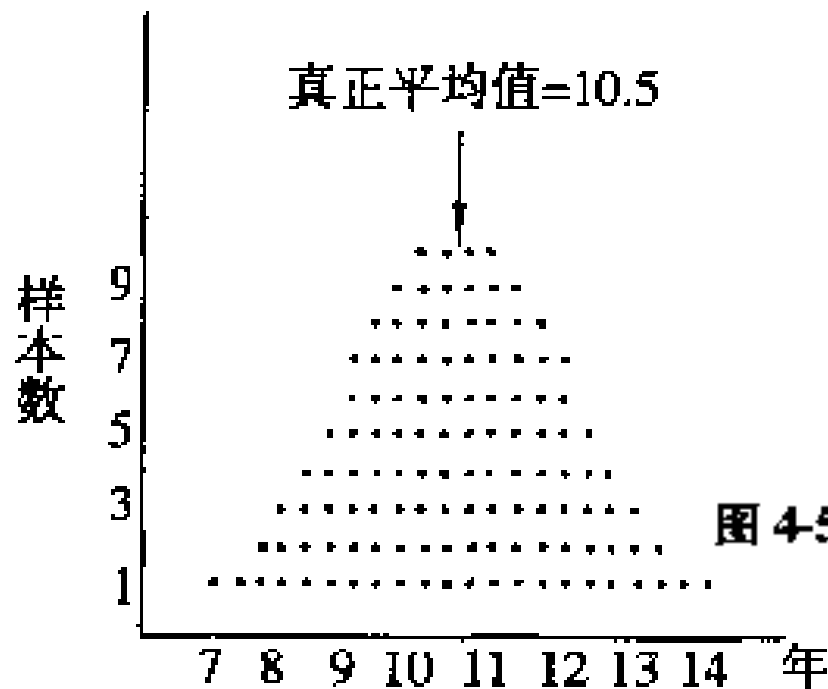


图 4-5 容量为 4 的样本的抽样分布 (总数 = 210)

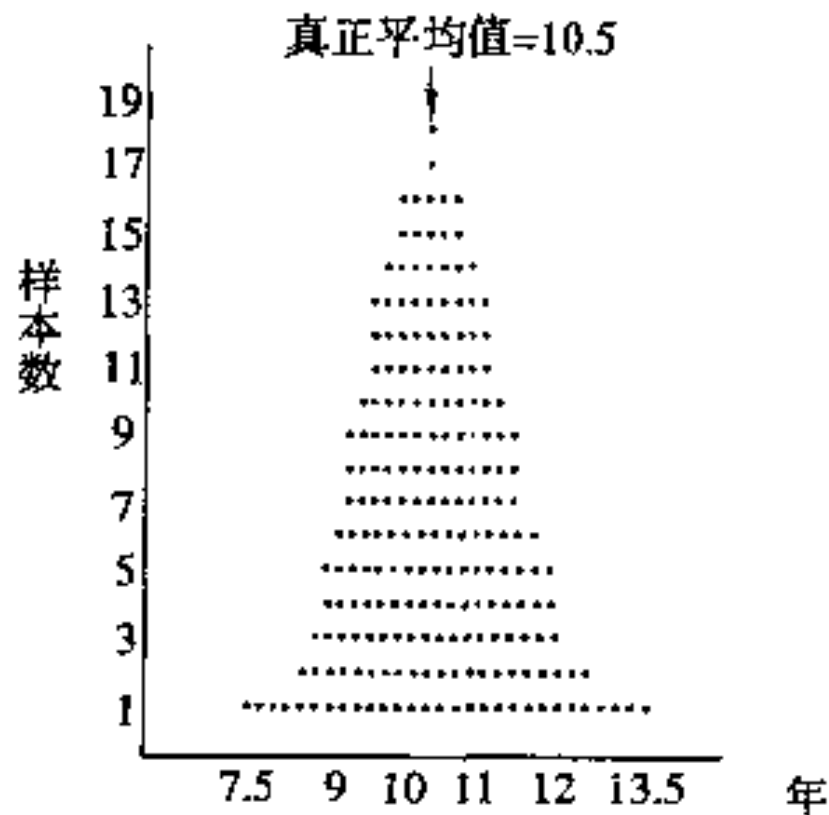


图 4-4 容量为 3 的样本的抽样分布 (总数 = 120)

## 二、抽样分布

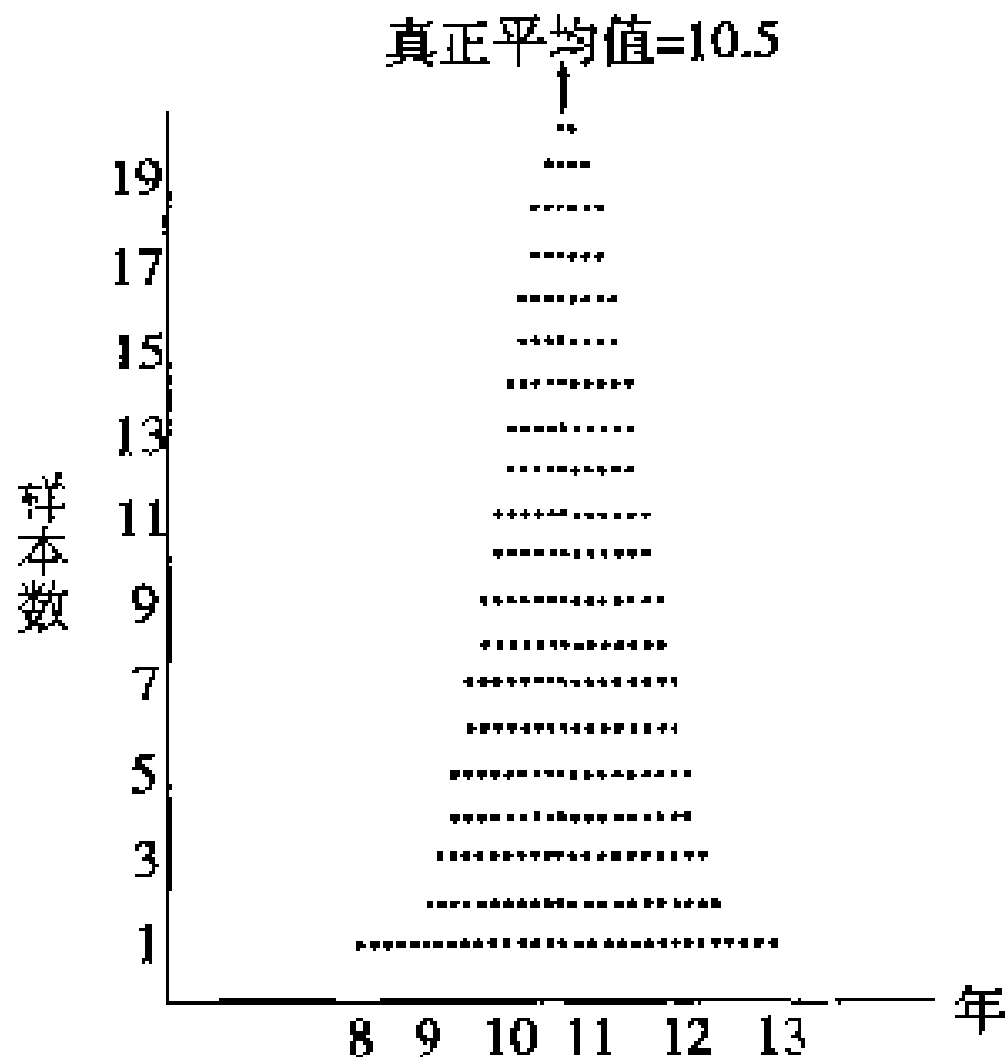


图 4-6 容量为 5 的样本的抽样分布(总数 = 252)

## 二、抽样分布

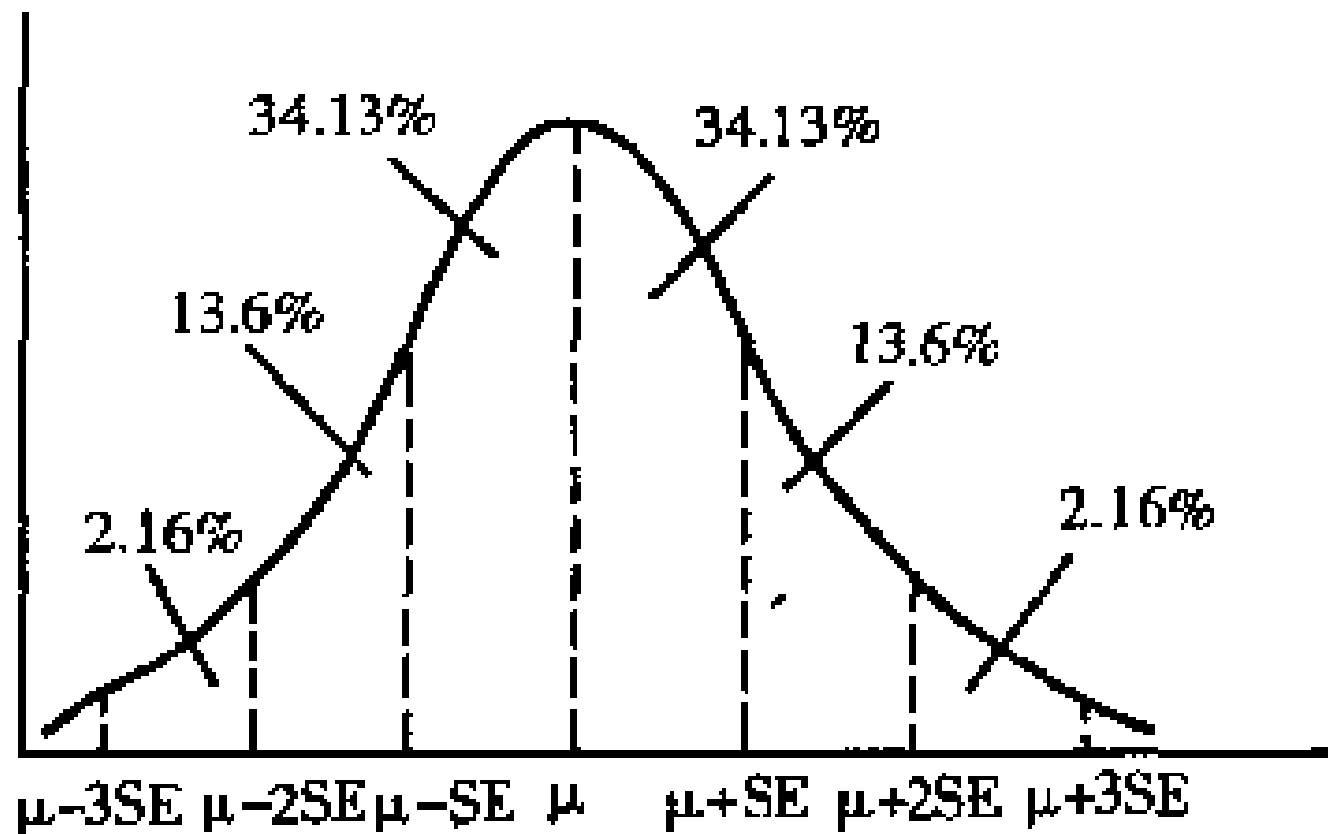


图 4-7 正态分布图

### 三、抽样的一般程序

界定总体

制定抽样框

决定抽样方案

实际抽取样本

评估样本质量

# 第三章

## ➔ 抽样

---

- ↗ 抽样的介绍
- ↗ 概率抽样的原理与程序
- ↗ 概率抽样方法
- ↗ 非概率抽样方法
- ↗ 抽样实务及问题

## 一、简单随机抽样

简单随机抽样（simple random sampling）：

按等概率的原则从含有 $N$ 个元素的总体中抽取 $n$ 个元素组成一个样本。

# 一、简单随机抽样

## ➤ 抽签法：

总体较小



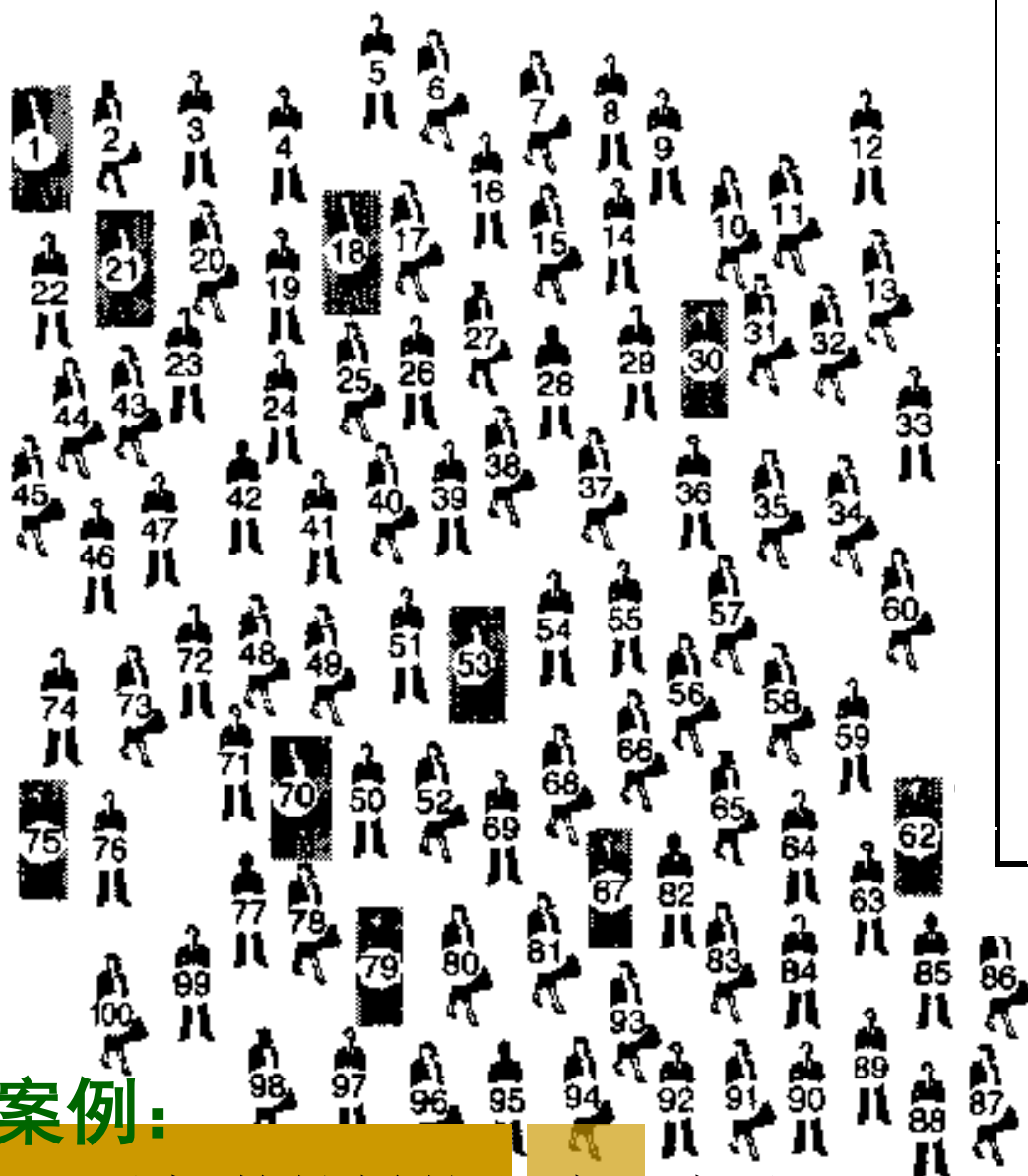
## ➤ 随机数法：

总体较大





# 一、简单随机抽样



附录 E  
随机数表

10480	15011	01536
22368	46573	25595
24130	48360	22527
42167	93092	06242
37570	399	
77921	069	
99562	729	
96301	919	
89579	143	
85475	368	
28918	685	
63553	409	
09429	939	

样本

30	67
70	21
82	01
79	75
18	53

案例:

随机数抽样的方法和步骤

Social survey method

## 二、系统抽样

**系统抽样（systematic sampling）：**

**又称等距抽样或间隔抽样，是把总体的元素进行编号后排序，再计算出间隔，然后按照一定的间隔抽取元素组成样本的方法。**

## 二、系统抽样

### ➤ 系统（等距）抽样步骤：

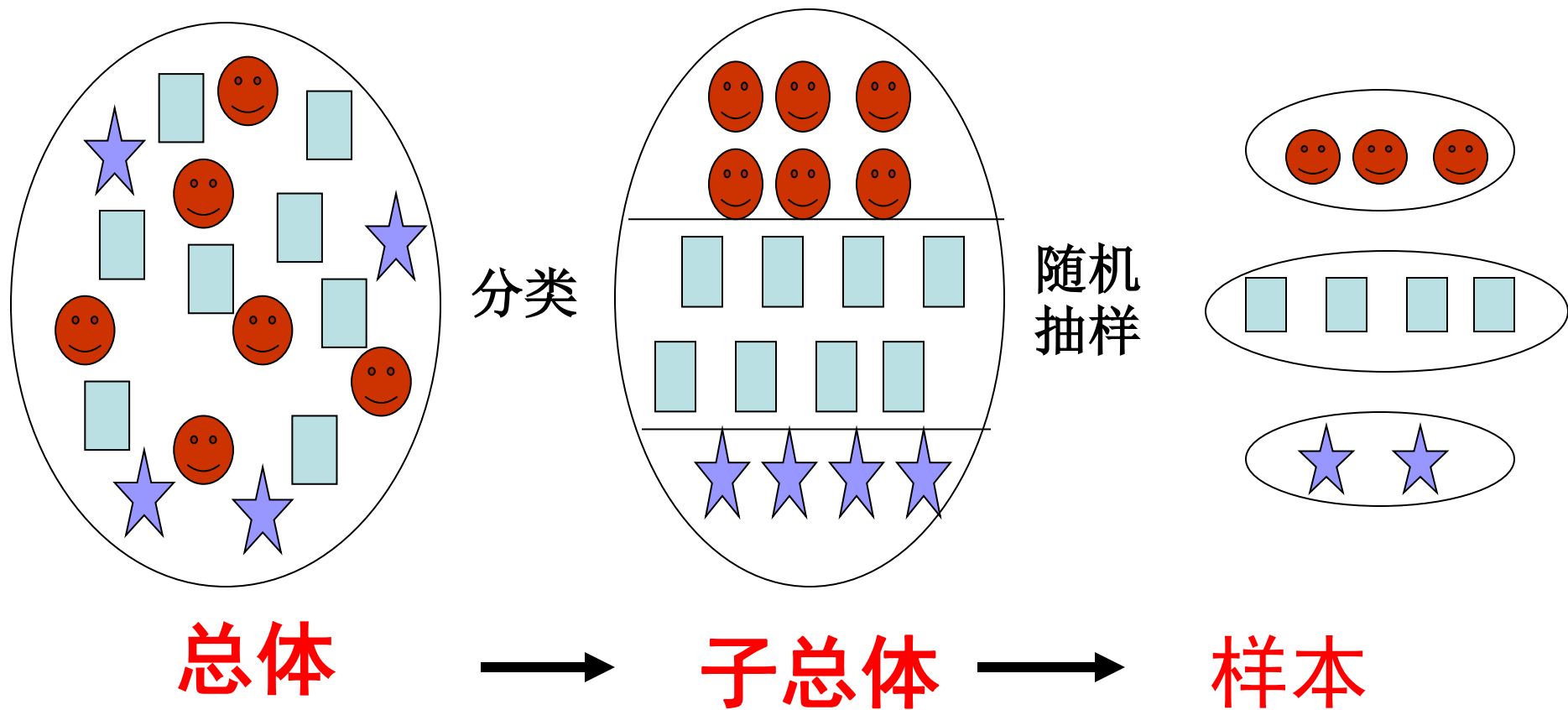
- (1) 将总体的所有个体按顺序编号；
- (2) 计算间距： $K=N/n$ ；
- (3) 用简单随机的方法在第一个 $k$ 中抽取一个记下编号 $A$ ；
- (4)  $A, A+K, A+2K, \dots, A+(n-1)K$ ；
- (5) 将 $n$ 个元素合起来构成一个样本。

### 三、分层抽样

分层抽样（stratified sampling）：

**又称类型抽样，是先将总体中的所有元素按某种特征或标志划分成若干类型或层次，然后再在各个类型或层次中采用简单随机抽样或系统抽样的办法抽取一个子样本，最后，将这些子样本合起来构成总体的样本。**

### 三、分层抽样



### 三、分层抽样

#### ➤ 分层抽样的优点：

- (1) 不增加样本规模的前提下降低抽样误差，把异质性较强的总体分成一个个同质性较强的子总体。
- (2) 便于了解、比较总体内不同层次的情况。

### 三、分层抽样

**思考：** 如何进行分层？

➤ 分层的标准

➤ 分层的比例

### 三、分层抽样

#### ➤ 分层标准

**a: 以分析的主要变量或相关变量**

**b: 保证各层内部同质性强，各层之间异质性强**

**c: 以那些已有明显层次区分的变量作为分层变量**

#### ➤ 分层比例

**a: 按比例分层抽样**

**b: 不按比例分层抽样**



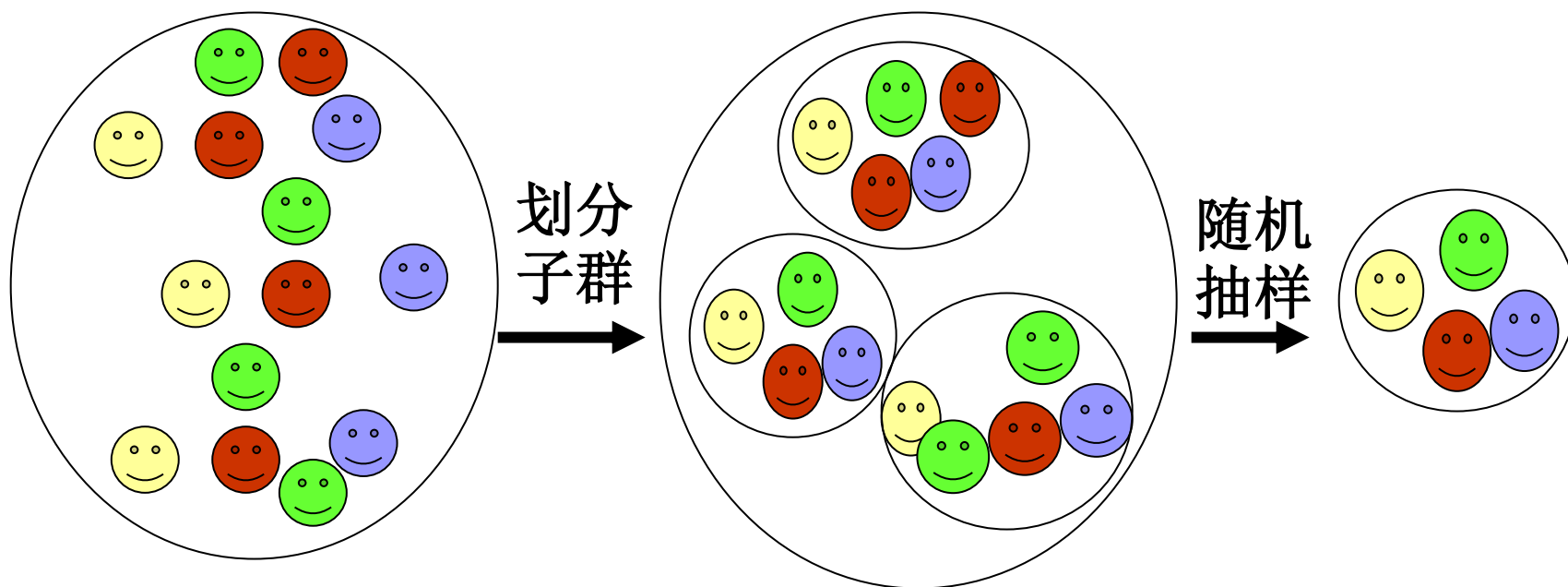
## 四、整群抽样

**整群抽样（cluster sampling）：**

**抽样单位不是单个的元素，而是成群的元素。**

**它是从总体中随机抽取一些小的群体，然后由所抽出的若干个小群体内的所有元素构成调查的样本。**

## 四、整群抽样



## 四、整群抽样

### ➤ 优点

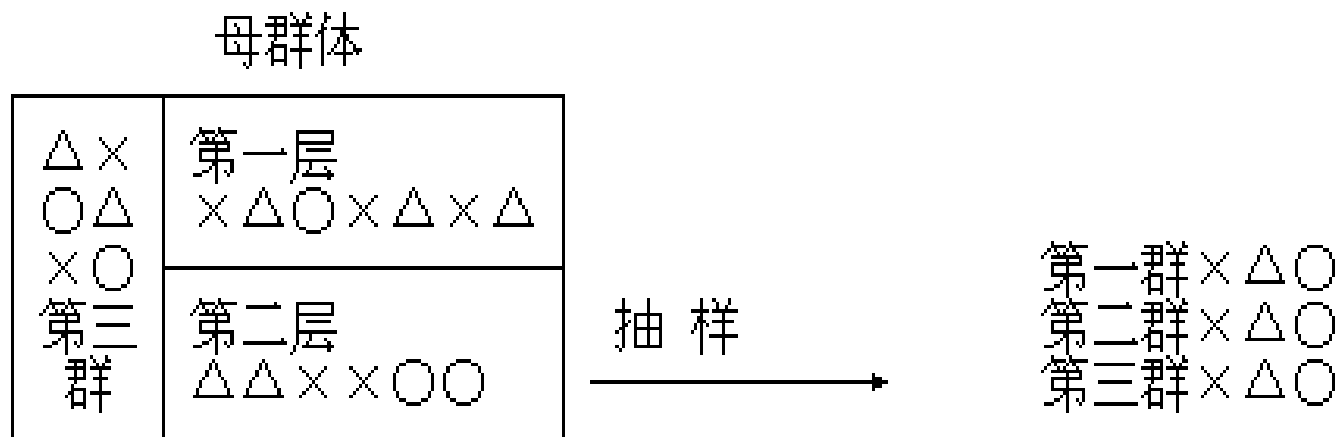
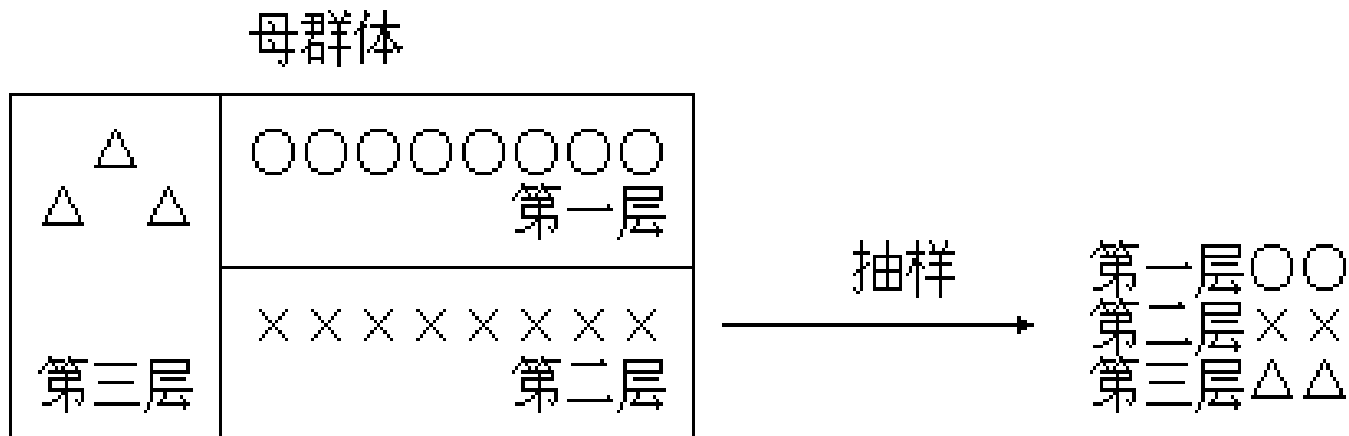
**无需详细的所有元素的名单，简单，费用低**

### ➤ 缺点

**样本的分布面不广，代表性相对较差**

**比较：**整群抽样与分层抽样？

# 比较



子群体间差异大，内部差异小——**分层抽样**

子群体间差异小，内部差异大——**整群抽样**

## 五、多段抽样

**多段抽样（multistage sampling）：**

**又称多级抽样或分段抽样，它是按抽样元素的隶属关系或层次关系，把抽样过程分为几个阶段进行。**

## 五、多段抽样

### ➤ 优点

**无需详细的所有元素的名单，各阶段的抽样单位数一般较少，容易操作。**

### ➤ 缺点

**每一个阶段都存在误差，抽样误差可能较大。**

**应对方法：尽量增加开头阶段的样本数适当减少最后阶段的样本数。**

# 抽样方法

概率抽样

简单随机抽样

系统抽样

分层抽样

整群抽样

多段抽样

PPS抽样

户内抽样

非概率抽样

偶遇抽样

判断抽样

定额抽样

雪球抽样

## 六、PPS抽样

### PPS抽样

(sampling with probability proportional to size) :

**“概率与规模成比例”。**

**阶段1: 每个群按照其规模被给予大小不等的抽取概率。**

**阶段2: 从每个抽中的群中都抽取同样多的元素。**



## 七、户内抽样

### （一）kish选择法

“Kish选择法”，保证每户家庭中所有的成年人都具有同等被抽中的概率。

- 问卷编上8种不同的号（对应不同比例）；
- 对每户成年人进行排序和编号  
（年长男-年轻男-年长女-年轻女）；
- 根据问卷编号和相应的选择卡确定调查对象。

## 七、户内抽样

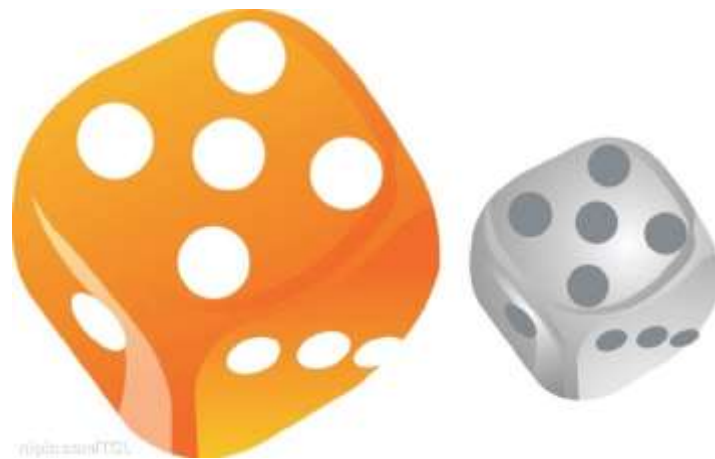
### ➤ kish抽样表

表 3—5 抽样表

抽样表的 序次	住户中成年成员数目					
	1	2	3	4	5	6 或更多
	抽取成年住户成员的号码					
A1	1	1	1	1	1	1
B1	1	1	1	1	2	2
B2	1	1	1	2	2	2
C	1	1	2	2	3	3
D	1	2	2	3	4	4
E1	1	2	3	3	3	5
E2	1	2	3	4	5	5
F	1	2	3	4	5	6

## 七、户内抽样

### (二) 掷骰子法



## 七、户内抽样

### (三) 生日法



## ➤ 概率抽样可能的困难

- 抽样框不易得或不明确
- 概率抽样成本较高
- 抽样方法的适用性
- 抽样中的拒访率、样本轮换等问题



非概率抽样

# 第三章

## ➔ 抽样

---

- 抽样的介绍
- 概率抽样的原理与程序
- 概率抽样方法
- 非概率抽样方法
- 抽样实务及问题

# 抽样方法

概率抽样

简单随机抽样

系统抽样

分层抽样

整群抽样

多段抽样

PPS抽样

户内抽样

非概率抽样

偶遇抽样

判断抽样

定额抽样

雪球抽样

## 一、偶遇抽样

**偶遇抽样（accidental sampling）：**

**又称方便抽样，是指研究者根据现实情况，以自己方便的形式抽取偶然遇到的人作为调查对象，或者仅仅选择那些离得最近的、最容易找到的人作为调查对象。**



## 一、偶遇抽样

### ➤ 优点

**便利、方便。**

### ➤ 缺点

**抽样误差大，不能推论总体。**

## 二、判断抽样

**判断抽样（judgmental sampling）：**

**又称立意抽样，是指调查者根据研究的目标和自己主观的分析来选择和确定调查对象的方法。**

## 二、判断抽样

**思考：** 什么情形下使用判断抽样？

**适用于无法确定总体边界或总体规模、或无法进行大规模抽样等情形。**

## 二、判断抽样

### ➤ 优点

**当研究者对研究的总体情况比较熟悉，研究者的分析判断能力较强、研究方法与技术十分熟练、研究的经验比较丰富时，采用这种方法往往十分方便。**

### ➤ 缺点

**样本的代表性难以判断。**

### 三、定额抽样

定额抽样（quota sampling）：

**使得样本中的成员在上述各种因素、各种特征方面的构成和在样本的比例尽量接近总体情形。**

**比较：** 定额抽样与分层抽样？

**(1) 目的不同：总体的模拟物**

——提高各层的异质性，同层的同质性

**(2) 性质不同：主观分析，非概率抽样**

——客观分析，概率抽样

### 三、定额抽样

#### ➤ 优点

**比偶遇抽样和判断抽样的代表性略好。**

#### ➤ 缺点

**只能对有限几个特征进行控制；**

**调查员抽取对象的自由度较大，易造成偏差。**

## 四、雪球抽样

**雪球抽样（snowball sampling）：**

**当无法了解总体情况时，可以从总体中少数成员人手，对他们进行调查，向他们询问还知道哪些符合条件的人，然后去找那些人并再询问他们知道的人。**



# 第四章 ➤ 抽样

---

- 抽样的介绍
- 概率抽样的原理与程序
- 概率抽样方法
- 非概率抽样方法
- 抽样实务及问题

## 一、样本规模

样本规模（sample size）：

**样本容量，指样本中所含个案的多少。**

## 二、影响确定样本规模的因素

- 总体的规模
- 估计的把握性与精准性
- 总体的异质性程度
- 调查者的时间、人力和物力

### 三、实际采用的样本规模

- 小型调查：100--299；
- 中型调查：300--1000；
- 大型调查：>1000。

## 四、抽样误差

### ➤ 抽样误差（sampling error）

**用样本统计值去估计总体参数值时所出现的误差，是由于抽样本身的随机性所引起的误差。**

### ➤ 非抽样误差

**度量误差，调查、记录、填答、汇总时出现的误差。**

## 四、影响抽样误差的因素

- 总体的分布方差
- 抽样规模

**注意：1、样本规模的确定需综合衡量；  
2、样本规模的扩大也会增加非抽样误差。**

# 社会调查方法

**- Thanks! -**

