

高等教育出版社



# 《市场调查》

Marketing Research

# 第四单元 抽样设计



第一节 抽样调查概述



第二节 抽样技术 几种抽样方法



第三节 误差来源与计算



第四节 样本容量的计算

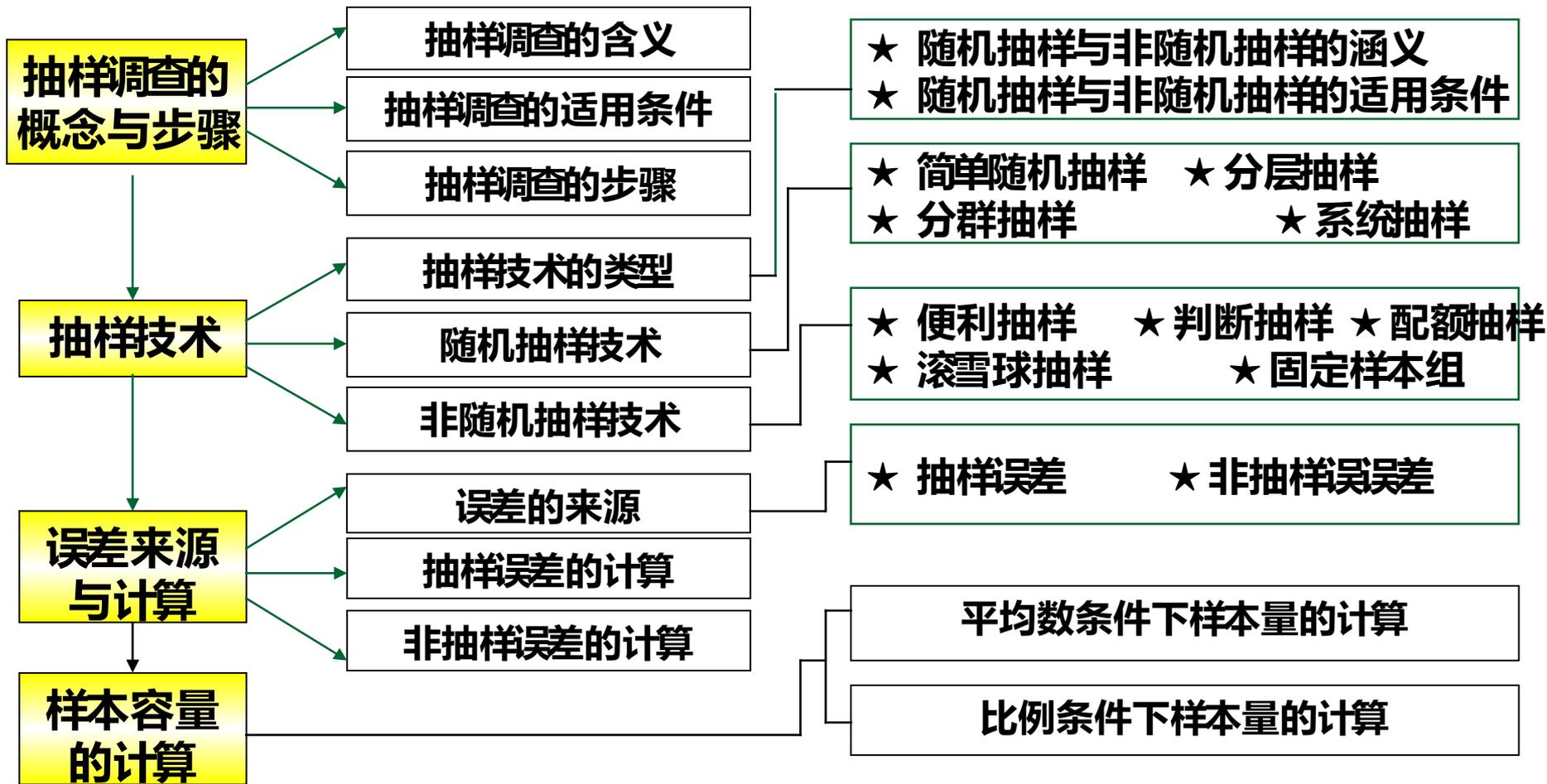
第五节 问题思考单元小结

总目录

# 【本单元内容】

- 📖 了解抽样调查的相关概念、过程和适用条件；
- 📖 理解抽样技术的含义；
- 📖 掌握抽样技术的种类；
- 📖 理解非随机抽样和随机抽样的含义和实际应用；
- 📖 理解抽样误差的来源；
- 📖 掌握平均数和比例条件下误差和样本量计算方法。

# 【学习导航】



GO

# 第一节 抽样调查概述

## 一、调查种类与抽样调查

1、**普查**：（ census ）是在某一时点上对构成总体的每一个体都进行调查，并从中得出有关总体特征的结论的一种调查方法。

**普查特点**：由于普查是一种全面调查，因而所取得的资料**全面、准确，标准化程度高**，可以为国家和企业制定方针、政策提供重要的依据。但是，由于普查需要动用大量的人力、财力、物力，调查**费用较高**，调查工作所需**时间较长**，组织管理工作**难度较大**，因而不宜经常采用。

**适用范围**：主要适用于小范围的市场调查和一些重要的现象的基本特征的调查。

普查在实际应用中，组织实施的方式有两种，一种是由调查的组织者制定**统一的调查表**，规定统一的调查时间，由被调查单位统一填写；另一种是设置专门的调查机构，培训调查人员，然后上门对调查对象进行登记调查。由于普查是对某一时点的现象的一种调查，工作量较大，所以后一种方式很少采用。



## 2、重点调查

**重点调查**，是从调查对象总体中选取少数重点单位进行调查，并以此推断总体的一种**非全面调查**。

**特点：**由于重点调查只需对少数重点单位进行调查，所以比较**节省人力、财力、物力**；调查对象少，更**有利于确定较多、较复杂的调查项目**；调查内容更深入、更细致；由于重点单位在总体中所占标志值的比重较大，重点调查能**更快速**地收集到所需要的总体的基本特征。但由于重点调查只是对一少部分对象进行调查，所以可能**缺乏全面性**，影响对总体推断的准确性。

## 重点调查使用中应注意的问题

(1) 重点单位必须客观存在，重点单位的确定不能带有任何主观色彩。

(2) 重点调查的目的只是想了解总体的基本特征，而不是为了准确地推断总体。

(3) 重点调查较适合于调查对象集中、调查内容集中的情况。例如，商品资源调查、销售渠道调查等。



### 3、典型调查

典型调查，是指在对调查总体深入细致了解的基础上，选择具有代表性的单位进行调查，并对总体进行推断的一种非全面调查。

#### 选典的方法

- (1) 选择中等水平的单位作为调查对象
- (2) 选择后进单位或先进单位作为调查对象，即所谓的“解剖麻雀式选典”
- (3) 在对总体进行充分了解的基础上，进行分组，然后，在每组中再有意识地选择一部分典型单位作为调查对象，即所谓的“划类选典”。



# 第一节 抽样调查概述

**4、抽样调查：**是抽取具有代表性的样本，测算样本结果，并以样本结果推断总体特征的一种调查方法。



## 二、抽样调查的术语

- 1、总体：是构成它的所有元素的集合。是研究者感兴趣的群体，是研究者要将研究结果进行概念化的目标。
- 2、样本：是为研究提供信息的总体中的一部分。
- 3、抽样：选择愿意参与研究的个体的过程。
- 4、抽样单位、个体：一次直接抽样所使用的基本单位。
- 5、抽样框：抽样范围，一次直接抽样时总体中所有抽样单位的名单。
- 6、参数值：总体值，关于总体中某一变量的综合描述。
- 7、统计值：样本值，关于样本中某一变量的综合描述。

提供了一种实现“由部分认识总体”的途径和手段；节省人力、时间、经费

### 三、抽样调查的特点

- 经济性。样本量减少，节约人、财、物。
- 及时、高效。工作量小，快捷、迅速、效率高。
- 准确、灵活。规模小，参与人员精干，登记误差小。

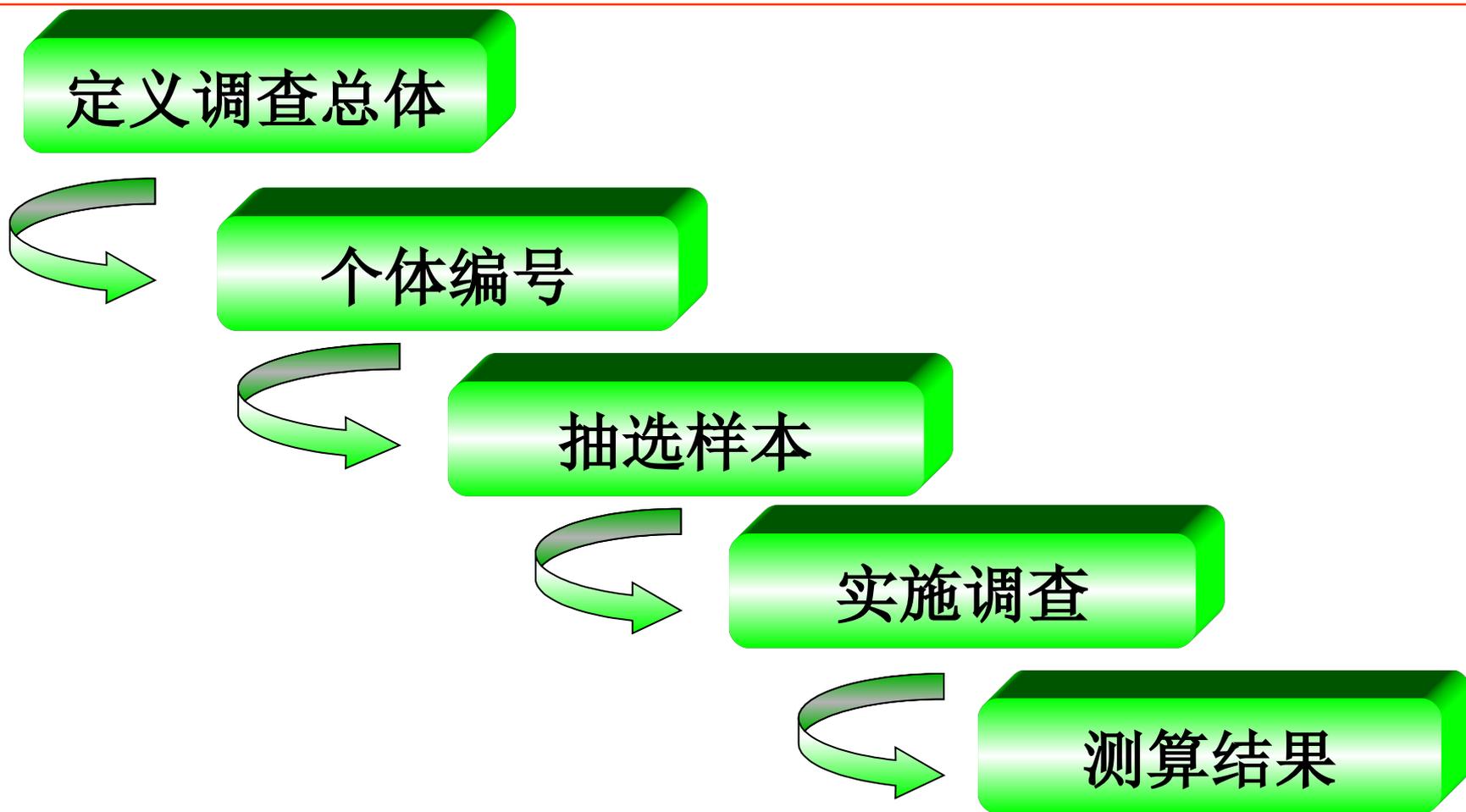
### 四、抽样调查的作用

- 1、对一些不可能或不必要普查的项目采用。
- 2、在经费、人力、物力、时间有限情况下采用。
- 3、对全面调查进行验证。
- 4、对同一现象在不同时间连续进行调查。
- 5、企业管理过程。

## 四、抽样调查的适用条件

比较项目	抽查	普查
预算	小	大
可用时间	短	长
总体规模	大	小
特征差异	小	大
抽样误差的成本	低	高
非抽样误差的成本	高	低
对个体案例的关注	是	否

# 五、抽样调查的步骤



# 1、定义总体

- 总体界定的正确与否直接影响抽样调查结果应用的有效性。
- 总体进行定义：在实际的调查中，通常借助调查表中的过滤性问题来判断某个个体是否属于调查总体，定义总体的因素包括：

因素	举例
地理因素	抽样调查的地域范围——顾客活动的范围，可以是一个国家，一个城市，一个县。
人口统计因素	根据调查目的和产品的目标市场，从人口统计的角度确定对调查结果起重大影响的那些个体。如年龄阶段、收入层次、职业群体等。
产品 / 服务使用情况	对于一些市场调查来说，它们的调查总体通常可以根据消费者对企业产品或者服务的使用情况来定义，包括是否使用以及使用频率等。

## 2、个体编号

- 即对所确定的调查总体中的个体进行编号。
- 一般，只有当调查者选择随机抽样技术时，这一步骤才是必要的，其目的是为了简化抽样工作。
- 当调查总体规模较大时，编号的工作量会增大，调查者可能会采用分层抽样或分群抽样来简化编号工作。

# 3、抽选样本

这一过程实施涉及两个方面的内容：

- 明确所采取的抽样方法；
- 根据调查者对抽样精确度等要求，计算从调查总体中所要抽取的样本数量，运用事先确定的抽样方法抽选样本。

4、实施调查

5、测算结果

[返回单元目录](#)

## 第二节 抽样技术

两种：随机；  
非随机

### 一、抽样技术的类型

**1、随机抽样 ( probability sampling )**：也叫**概率抽样**，就是总体中的每一个体都具有相等的机会被抽取为样本单位的抽样技术。

- ◆依据概率论的基本原理，按照随机原则进行抽样
- ◆主要有：简单随机抽样、分层抽样、整群抽样、等距抽样

## 第二节 抽样技术

### 一、抽样技术的类型

**2、非随机抽样 ( non-probability sampling )**：调研总体中的每一个体被抽取的机会不平等，这种不平等的产生是由于调研人员根据主观设定的某些标准来选择样本。

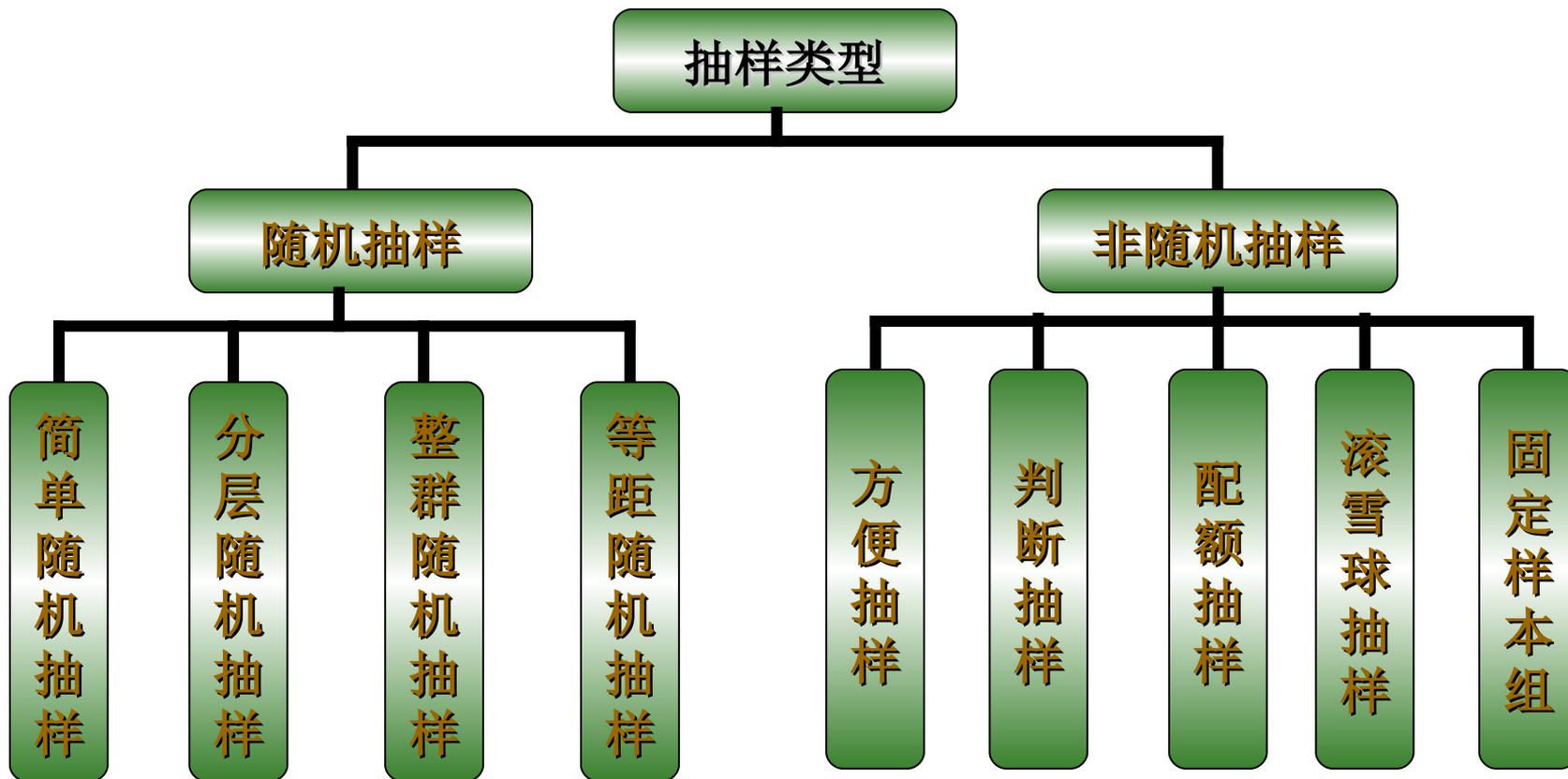
◆依据研究者的主观意愿、判断或是否方便等因素抽取样本。

◆主要有：方便抽样、判断抽样、配额抽样、雪球抽样和固定样本组抽样。

## 二、随机抽样与非随机抽样的适用条件

抉择考虑方面	非随机抽样	随机抽样
研究的性质	探索性	归纳性
误差	非抽样误差大	抽样误差大
变异程度	同质(低)	异质(高)
统计	不利	有利
操作	有利	不利

# 第二节 抽样技术



返回单

## 三、随机抽样

提问：你做过掷硬币、骰子和抽奖游戏吗？

### (一) 简单随机抽样

- **简单随机抽样**：又称纯随机抽样，是按等概率原则直接从含有  $N$  个单位的总体中随机抽取  $n$  个单位组成样本（ $N > n$ ）。抽样过程与调研人员的经验判断完全无关。
- 简单随机抽样是概率抽样的最基本形式，也是其它各种随机抽样技术的基础，通常适用于总体单位之间差异程度较小和数目较少的抽样。
- 常用的是**抽签法**和**随机数表法**。

# 1、抽签法

训练：各小组准备纸、笔、容器，按我说的去做。  
(5分钟)

- **抽签法的操作**：(1)把调查总体的每个个体依次编号，并将这些号码分别标记在相互独立的、完全相同、匀质物体上；(2)将这些物体放入某一容器（如纸盒）搅拌均匀；(3)由调研人员从中任意抽取，将与抽到的号码相对应的个体作为样本单位；(4)如此重复数次，直到抽足预定的样本数目为止；(5)由抽中的号码所代表的个体组成的就是一个简单随机样本。
- **抽签法的优点**在于操作简便，但当调查总体单位很多时，编号的工作量很大，搅拌均匀也比较难以做到。
- **抽签法往往只适用于**总体规模较小的市场调研。

## 2、随机数表法

- **随机数表法**：又称**乱数表法**，其操作过程如下：  
：(1)将总体中的每个个体随机编号，(2)从随机数表的任一位置开始，或向左，或向右，或向上，或向下，或一定间隔向一个固定方向顺序取数，选定的数字所对应的单元即入样，(3)重复的数字和没有对应单元的数字去掉，直至抽足所需样本量为止。要注意的是，所有号码的位数均应相同。
- 对于总体单位很多的情形，通常采用随机数表法来抽样。

## 【随机数表法示例】

**【例】某大学共有 30000（五位数）学生，要求采用随机数表法从中抽取 1000 人作为样本进行关于大学生消费态度的调查。首先，调研人员要将总体中的每个个体从 1 到 30000 进行编号；然后从随机数表的任意一行和任意一列的某一个五位数开始，按照从左到右的顺序，或者从下到上的顺序，以 30000 为标准，对随机数表中依次出现的每个五位数进行取舍：凡小于或等于 30000 的编码就选出来，凡大于 30000 的数以及已经选出的数码则抛弃。直到选够 1000 个编码为止；最后按照所抽取的数码，从总体名单中找到它们所对应的 1000 个个体。这 1000 个个体就构成一个随机样本。随机数表（片断）如表所示。**

# 随机数表（片段）

	( 1 )	( 2 )	( 3 )	( 4 )	( 5 )	( 6 )	( 7 )	( 8 )	( 9 )
1	10480	15011	01536	02011	81647	91646	69179	14194	62590
2	22368	46573	25595	85393	30955	89198	27982	53403	93965
3	24130	48390	22527	97265	76393	64809	15179	24830	49340
4	42167	93093	06243	61680	07856	16376	39440	53537	71341
5	37570	39975	81837	16656	06121	91782	60468	81305	49684

# 随机数表法随机抽样训练

时间：8分钟，确定出各小组抽样选取的样本学生名单。

利用课本 P115 的随机数表，在我们班级现有同学中随机选取 8 位同学作为样本进行调查。

请各小组进行样本确定。

返回

## (二) 分层抽样

**分层随机抽样**：亦称**分类抽样**或**类型抽样**，指把调查总体按照其属性不同分为若干层次或类型，然后在各层或类型按一定的比例关系随机抽取样本。

分层随机抽样需要注意以下几点：（1）分层标志必须可以把总体分隔为相互排斥的次层，而层间不能有交叉重叠；（2）分层的标志必须与关心的总体特征相关；（3）层间差异大，层内差异小；（4）各层个体数目和比例必须可以获得；（5）层数量不宜太多；（6）在各层抽取样本的方法依然是简单随机抽样。

**类型**：通常包括**比例分层抽样**和**非比例分层抽样**。

# 1、比例分层抽样

**例：**根据某产品的销售统计数据，家庭用户总体为4万户，其中，低收入家庭为1.2万户，中等收入家庭为2万户，高收入家庭为8千户。为进一步研究该产品用户使用状况，需对300户进行抽样调查，如采用比例分层抽样，低、中、高收入家庭用户的样本容量分别为多少？抽样结果如下。

家庭收入分层	户数（万）	所占比例（%）	各层样本容量
低	1.2	30	90
中	2	50	150
高	0.8	20	60
总计	4	100	300

## 2、非比例分层抽样

考虑各层的代表性权重和变异程度（标准差）

- **非比例分层抽样**，也称**最佳抽样法**，是指根据其他因素，如各层的标准差来分配各层所应抽取的样本数目。这种方法的优势在于可以降低各层的差异，提高样本的可信度。其计算公式为：

$$n_i = n \times \frac{N_i S_i}{\sum N_i S_i}$$

回忆：离均差、方差、均方差（标准差）？

- 式中： $n_i$ ：第  $i$  层应抽取的样本数目； $n$ ：应抽取的样本总数； $N_i$ ：第  $i$  层的包含的单位总数； $S_i$ ：第  $i$  层的标准差

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

# 非比例分层抽样示例

- ★ 上题中，如果知道低、中、高层次收入的标准差分别为 100，200 和 300 元，请计算非比例分层抽样条件下的样本量？（抽样数  $n=300$ ）

45, 165, 90

家庭收入分层	户数（万）	标准差（元）	$N_i$	$S_i$	各层样本容量
低	1.2	100	120		
中	2	220	440		
高	0.8	300	240		
总计	4	—	800		300

### (三) 整群抽样

- **整群抽样**：也称分群抽样，是按一定的标准把调查总体区分为若干不同的子群体，然后按照简单随机抽样法选取部分子群体进行调查的抽样方法。
- 分群抽样又分为**单阶段分群抽样**和**多阶段分群抽样**。

# 整群抽样与分层抽样的区别

比较项目	分层抽样	整群抽样
分层(群)标志	与研究的总体特征相关	与研究的总体特征无关
子集间差异	各层之间差异大	各群间同质或差异小
子集内个体差异	层内个体同质或差异小	各群内个体差异大
抽样方式	各层内分别随机抽取适当个体组成样本	随机抽取一个或几个群作为样本

返回

## （四）等距抽样

- **等距抽样**：又称**系统抽样**，指先将总体中的个体按一定标志进行排列编号，并根据总体单位数和样本单位数计算出抽样距离，然后按照相同的距离或间隔选取样本的一种抽样方法。
- 它和简单随机抽样一样，需要有完整的抽样框，样本的抽取也是直接从总体中抽取个体，而无其他中间环节。

# 系统抽样示例

**【例】**某连锁超市要从某地区的 100 个超市中抽取 10 个进行调查，采用系统抽样法应如何抽取。

- 第一步，采用随机的方式将总体中的所有个体编号： $1, 2, 3, \dots, 100$ ；
- 第二步，计算抽样间隔，总体数（ $N$ ）/ 样本数（ $n$ ）=  $100/10 = 10$ ，抽样距离为 10；
- 第三步，用简单随机抽样在抽样距离内（1 至 10）确定起始编号，假设抽中 3；
- 第四步，根据起始号和抽样间隔确定 10 个应抽取的个体，最终抽取的编号分别为： $3, 3 + 10, 3 + 20, 3 + 30, \dots, 3 + 90$ ，即  $3, 13, 23, 33, \dots, 93$ ，共 10 个超市组成样本。

[返回](#)

## 四、非随机抽样

### (一) 任意抽样

- **方便抽样**：也称**便利抽样**或**任意抽样**，指调研人员从工作方便出发，在调研对象范围内随意抽选一定数量的样本进行调查，通常没有严格的标准。如在超级市场和车站等进行的街头拦截调查，都属于任意抽样的范畴。
- **操作简单、节约时间和费用**是任意抽样的优点。但可信度较低。
- **任意抽样一般只适用于非正式的探索性调研**，否则会造成较大的误差。

[返回](#)

## (二) 判断抽样

- **判断抽样**：是指由调查人员依据自己的经验抽取样本，或由某些有见解的专家选定一定数量的样本进行调研的抽样方法。判断抽样是主观的。
- 判断抽样具体操作中有三种可供选择的方案：一是多数型；二是平均型；三是主观型。
- 判断抽样具有抽样简便的优点，当调研项目的精确度要求不高的情况下会被经常采用。

返回

### (三) 配额抽样

- **配额抽样**：是按一定标准分配样本数额，并在规定数额内由调查人员任意抽选样本。即首先要按控制特征将总体分为若干层次，然后由调查人员主观选定样本。
- **分层抽样与配额抽样的区别**

比较项目	分层抽样	配额抽样
分层标志	与研究的总体特征相关	分层标志无特殊要求
抽样方式	分别从各层内随机抽取适当个体组成样本	分别从各层内主观抽取适当个体组成样本
分层目的	提高样本代表性	比较不同群体，解释关系或检验理论

# 配额抽样与判断抽样的区别

比较项目	判断抽样	配额抽样
抽样要求	注重样本“质”的分配	注重样本“量”的分配
抽样方式	主观判断后抽取适当个体组成样本	从各层内主观抽取适当个体组成样本
特点	简便	复杂精密

# 1、独立控制配额抽样

■ **独立控制配额抽样**：调研人员分别独立地按照各类控制特性分配样本数额。

【例】某超市需要对其顾客进行调查，选定的三个分类标准为性别、年龄、收入，确定的样本数目为360个，独立控制配额抽样情况如表所示。

性别		年龄		收入	
男	60	16至29岁	20	高	24
		30至44岁	30	中	60
女	60	45至59岁	40	低	36
		60岁以上	30		
合计	120		120		120

## 2、交叉控制配额抽样

- **交叉控制配额抽样**：也称非独立控制配额抽样，是调研人员充分考虑各层之间的交叉关系，同时对具有两种或两种以上特性的样本做出具体的规定。
- **交叉控制配额抽样的步骤**如下：
  - 首先，确定总体分类标准；
  - 其次，按确定的分类标准把总体分为若干子群体；
  - 第三，计算各子群体在总体中所占的比例；
  - 第四，根据各子群体的比例分配各子群体所应抽取的样本数目；
  - 第五，根据分配的样本数，在各群抽取相应数量的样本。

# 交叉控制配额抽样示例

**【例 1】** 某大学有 4000 名学生，其中男生占 60%，女生占 40%，管理类和理工类学生各占 50%，大一，大二，大三，大四的学生比例分别为 40%，30%，20%，10%，现要抽取 500 人进行调查，按交叉控制配额抽样，其情况如下表所示。

	男								女							
	管理类				理工类				管理类				理工类			
年级	一	二	三	四	一	二	三	四	一	二	三	四	一	二	三	四
人数	60	45	30	15	60	45	30	15	40	30	20	10	40	30	20	10

# 交叉控制配额抽样示例

**【例 2】** 以年龄、性别与收入水平作为交叉控制因素进行配额抽样的样本配额分布

		收入水平						合计
		高		中		低		
性别		男	女	男	女	男	女	
年龄	50 以上	6	19	4	13	2	6	50
	25-50	13	37	9	26	3	12	100
	25 以下	6	19	5	13	2	5	50
	合计	25	75	18	52	7	23	200
合计	100		70		30			

## （四）滚雪球抽样

提问：关于滚雪球游戏，你玩儿过吗？

- **滚雪球抽样的操作方法**：首先选出一组最初的调查对象，通常是随机选出的，访谈之后，要求这些被访者推荐一些属于目标总体的其他人，根据这些推荐选出后续的被访者。
- 滚雪球抽样主要是用于估计稀有的人群特征，例如私家车主、保姆等。其主要**优点**表现在，通过对调研总体设定期望的特征，从而增强了样本个体的相似性。因此，采用这种方法所产生的抽样误差比较小，成本比较低。对于一些难以获得的样本采取这种抽样方法能取得较好的效果。

返回

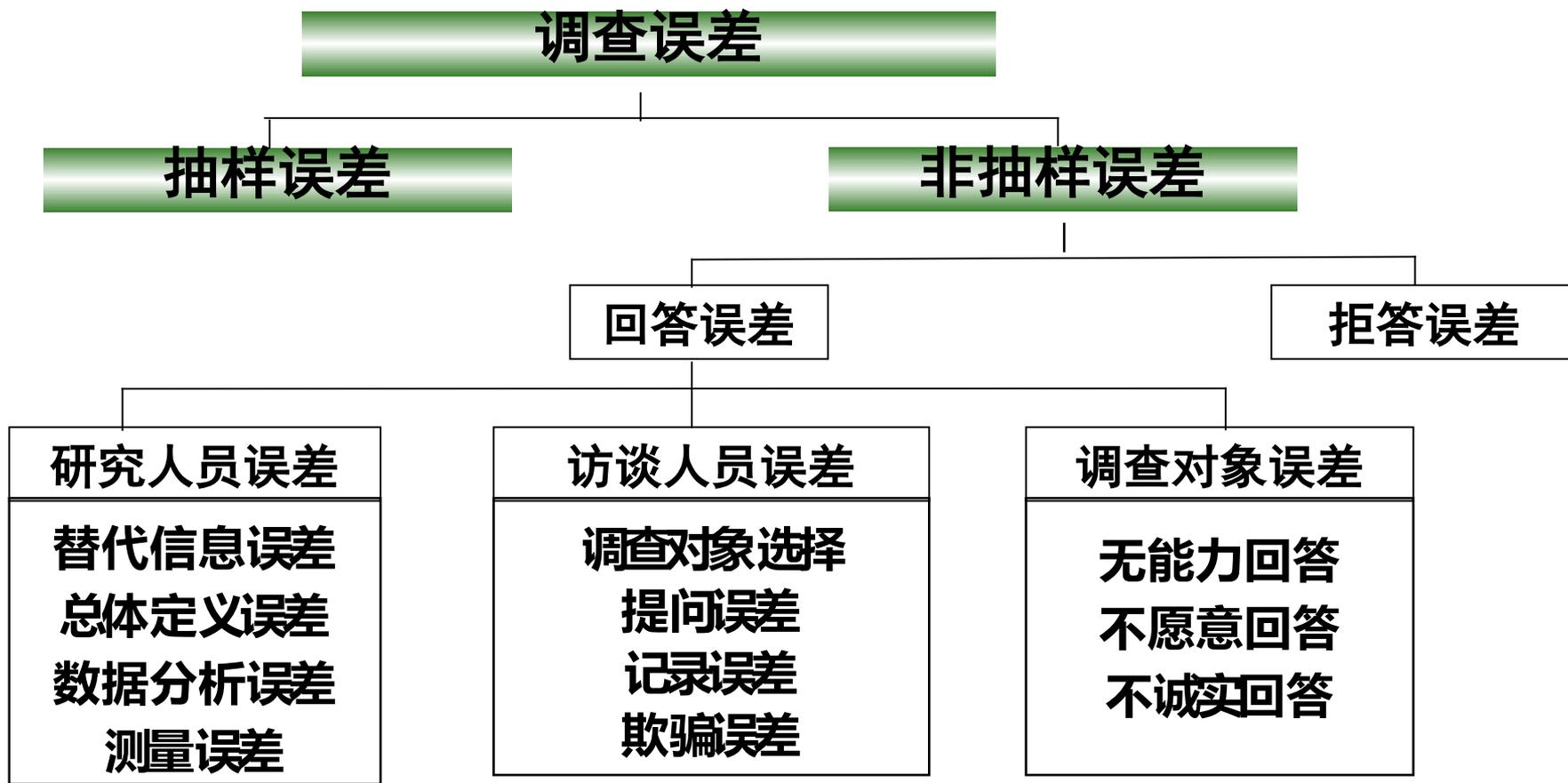
## （五）固定样本组抽样

- **固定样本组**：选取的样本固定下来，长期进行调查。
- 这种方法的优点在于其调查对象稳定，可以及时、全面的取得各种所需要的信息，费用低，回收率较高，这也是它在企业实践中被广泛应用的一个重要原因。但固定样本组也同时可能受到一些困扰，如被调查者是职业受访者，参加调研只是为了获取一定的利益，从而可能导致所收集信息的可靠性降低。

返回

# 第三节 误差来源与计算

## 一、误差的来源



# 第三节 误差来源与计算

## 一、误差的类型

### (一) 非抽样误差

#### 1、非抽样误差

非抽样误差是指除抽样误差以外所有的误差的总和。

#### 2、非抽样误差的来源

- 引起非抽样误差的原因很多。比如抽样框不齐全、访问员工作经验有限，被访者不配合访问而加以虚假的回答，数据录入、问卷设计本身存在缺陷等等。
- 贯穿了市场调查的每一个环节。任何一个环节出错都有可能导致非抽样误差增加而使数据失真。
- 可以控制

# 第三节 误差来源与计算

## (一) 非抽样误差

3、鲍德威（1915）误差四个来源：

- 获得的信息不正确或不真实
- 定义和标准不严格、不确切、不适当
- 样本不能代表总体
- 部分数据对于总体的估计将要产生的误差

# 第三节 误差来源与计算

## (一) 非抽样误差

### 4、戴明非抽样误差：

- 回答的变异性
- 不同类型和不同水平的访问员
- 访问员引起的主观偏差
- 委托方对于数据的期望（人为影响）
- 问卷设计的缺陷
- 抽样前后总体发生的变化
- 无回答的偏差
- 过时记录的偏差
- 数据缺乏代表性
- 解释数据的误差

# 第三节 误差来源与计算

## 一、误差的类型

### (二) 抽样误差:

1、抽样方法本身所引起的误差。当由总体中随机地抽取样本时，那个样本被抽到是随机的，由所抽到的样本得到的样本指标  $x$  与总体指标  $\mu$  之间偏差，称为实际抽样误差。

不可能列出所有的实际抽样误差，而用平均抽样误差来表征各样本实际抽样误差的平均水平。

**抽样误差**是指样本指标值与被推断的总体指标值之差。

2、主要包括：样本平均数与总体平均数之差  $\mu_x$ ；样本成数与总体成数之差  $\mu_p$ 。

# 第三节 误差来源与计算

## 二、抽样误差的来源及特性

1、登记性误差或责任误差）；

2、代表性误差（A、系统性误差； B、偶然性误差）

抽样误差特指偶然性误差，包括实际抽样误差和抽样平均误差。

3、影响抽样误差的因素：

➤ 抽样单位数的多少；

➤ 总体中被研究标志的变动程度的大小。

4、所谓的抽样误差是由抽样的随机性引起的，一般来说是无法避免，但可以控制的。

# 三、误差的计算

## 1、平均数的抽样误差

重复抽样下

$$\mu_x = \sqrt{\frac{\sigma^2}{n}}$$

不重复抽样下

$$\mu_x = \sqrt{\frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)}$$

在实际调查中，总体标准差  $\sigma$  往往不知道而无法计算抽样误差  $\mu$ 。一般常用样本标准  $S$  来代替  $\sigma$  进行计算。样本标准差  $S$  可以根据抽样结果来计算。

# 误差的计算例题

假定某街道有 8000 户居民，用单纯随机抽样法抽取 200 户，调查居民对食用油的月均需求量。调查结果表明，每户居民食用油月均需求量为 800 克，标准差为 100 克。求抽样平均误差是多少？

解：已知：  $n=200$  ，  $S=100$

$$\text{抽样平均误差： } \mu_x = \frac{\sigma}{\sqrt{n}} = \frac{S}{\sqrt{n}} = \frac{100}{\sqrt{200}} = 7.1(\text{克})$$

答：此次抽样调查的抽样平均误差为 7.1 克。

## 2、比例条件下的抽样误差

重复抽样下

$$\mu_p = \sqrt{\frac{P(1-P)}{n}}$$

不重复抽样下

$$\mu_p = \sqrt{\frac{P(1-P)}{n} \left( \frac{N-n}{N-1} \right)}$$

## 抽样误差 计算练习

对市场营销专业大二 160 名学生期末《市场调查》考试成绩进行简单随机抽样调查。随机抽取的 25 名成绩如下：

编号	001	005	009	021	025	032	052	061	066
成绩	98	92	93	87	89	91	75	90	77
编号	072	099	101	108	112	123	125	132	138
成绩	94	72	88	89	96	81	85	81	85
编号	139	141	142	145	148	151	155		
成绩	91	89	79	82	96	91	84		

请计算：样本平均成绩、平均成绩抽样平均误差及以 95% 的置信度估计整体学生成绩范围。

返回单元

## 抽样误差 计算练习

请计算：平均成绩、平均误差及以 **95%** 的置信度估计整体学生成绩范围。

**提示：**总体标准差  $\sigma$  可以用样本标准差  $S$  来代替进行计算公式为

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

样本标准差  $S$  可以根据抽样结果来计算，计算公式为：  
 $n=25$ ， $X \sim (\bar{X}, \sigma^2)$ ，置信度  $1-\alpha=95\%$ ， $\alpha/2=0.025$ ，查标准正态概率双侧临界值表  $U$ ，计算平均误差和极限误差  $\Delta x = U \cdot \mu_x$ ，便可求出置信区间，即估计学生成绩范围。

返回单元

# 第四节 样本容量与计算

## 一、样本容量 $n$

样本容量即样本数目是指抽样单位数，在随机抽样时必须确定必要的样本数目。

样本数目过多，浪费人力、物力、财力和时间，抽样数目过少，会影响调查结果的精确度，造成较大误差，所以确定必要的样本数目极为重要。

在抽样调查中，样本数目要取多少为宜？这是一个比较复杂问题，它与人们给定的条件有关。这里所指的条件包括允许误差、置信度等。

## 二、影响样本规模确定的因素

1、总体的变异程度（总体方差  $\sigma^2$ ）

异质性大，所需要的样本规模大。

2、抽样的精确性（允许误差  $\Delta$  要求）

置信区间越小，抽样的精确度越高，样本值与总体值之间的误差越小，所要求的样本规模就越大。

3、调查结果的可靠程度（概率度  $t$  和概率保证度  $Ft$  也称置信度  $C$  要求）（如：置信度  $1-\alpha=95\%$ ，双侧临界值  $U_{\alpha/2}$  即  $U_{0.025}=1.96$ 。可查表）

置信度越高，推论的把握性越大，所要求的样本规模越大；反之，样本容量小。

4、经费、人力和时间—抽样方法和抽样组织形式

# 三、样本容量计算

以简单随机抽样为例

1、平均数的样本容量

重复抽样下

$$n = \frac{Z_p^2 \sigma^2}{\Delta^2}$$

不重复抽样下

$$n = \frac{Z_p^2 \sigma^2 N}{N\Delta^2 + Z_p^2 \sigma^2}$$

2、比例的样本容量

重复抽样下

$$n = \frac{Z_p^2 P(1-P)}{\Delta^2}$$

不重复抽样下

$$n = \frac{Z_p^2 NP(1-P)}{N\Delta^2 + Z_p^2 P(1-P)}$$

## 样本容量 课堂练习（参考 P124）

[例 1] 某市调查职工每月食品消费支出情况。已知职工平均每人月食品消费支出标准差为 40 元，若要求允许误差为 3 元，置信度达 95.45%，求样本数目需要多少人？若其他条件不变，置信度提高到 99.73%，样本数目又需要多少人？

## [例 1] 参考答案

已知：  $\Delta X=3$ ，  $\sigma=40$ ， 置信度  $F(t) = 95.45\% = 0.9545$ ，  
查正态分布概率表  $t=2$ ， 置信度  $F(t) = 99.73\% = 0.9973$ ，  
查正态分布概率表  $t=3$

$$\text{解： 当 } F(t) = 95.45\% \text{ 时， } n = \left(\frac{t\sigma}{\Delta X}\right)^2 = \left(\frac{2 \times 40}{3}\right)^2 = 711(\text{人})$$

这说明当允许误差为 3 元， 置信度为 95.45% 时， 要抽 711 人来调查。 当  $F(t) = 99.73\%$  时，

$$n = \left(\frac{t\sigma}{\Delta X}\right)^2 = \left(\frac{3 \times 40}{3}\right)^2 = 1600(\text{人})$$

这说明当允许误差为 3 元， 置信度上升到 99.73%， 需抽 1600 人调查。

从以上计算可以看出， 对抽样调查结果置信度要求愈高， 样本的数目就要越多。 反之亦然。

[返回单](#)

## 样本容量 课堂练习（参考 P124）

[例 2] 某市有 5 万户居民，抽选 500 户作样本，调查居民人均月生活费支出为 350 元，抽样误差为 15 元，请推断置信度为 95% 条件下，该市居民人均月生活费支出的置信区间。

## [例 2] 参考答案

根据题意，分析可知： $\bar{x} = 350$ 元  $\mu_x = 15$ 元

$F(t) = 95\%$ ，查正态分布表， $t = 1.96$

解：

$$\bar{x} - t\mu_x \leq \bar{X} \leq \bar{x} + t\mu_x$$
$$350 - 1.96 \times 15 \leq \bar{X} \leq 350 + 1.96 \times 15$$
$$350 - 29.4 \leq \bar{X} \leq 350 + 29.4$$
$$320.6 \leq \bar{X} \leq 379.4$$

在置信度为 95% 条件下，该市居民人均月生活费支出的置信区间或区间范围是 320.6 ~ 379.4 元。

请思考：

其他条件不变，置信度提高到 99.49%，该市居民人均月生活费支出置信区间又是多大？

当置信度为 99.49% 时，该市居民人均月生活费支出的置信区间为 308 ~ 392 元之间。

返回单

# 第五节 抽样设计有关问题

## 一、抽样精度问题

- 精度是误差的相反数，提高精度即减少误差。
- 影响误差的最重要因素是样本量，而样本量与调查费用直接相关，样本量愈大，费用就愈高。
- 一般控制置信度 **95%**，相对误差 **2%~3%**。

## 二、抽样框误差

要从一个总体中抽选样本，需要一个包括全部总体单元的一个框架，用来代表总体，从中抽选样本的一个框架就称作抽样框。构成抽样框的单元称为抽样单元。

抽样框误差主要来自：

- 1、丢失目标总体单位
- 2、包含非目标单位

误差控制：

- 重新定义总体；
- 筛选个体；
- 加权调整。

### 三、样本量的问题

**样本容量**即样本数目是指抽样单位数，在随机抽样时必须确定必要的样本数目。

样本数目过多，浪费人力、物力、财力和时间；

抽样数目过少，会影响调查结果的精确度，造成较大误差。

在实际抽样调查中设计要求：满足一定误差要求下使费用最少，或在一定费用限制下精度最高。

## 四、抽样方式选择的问题

结合各种方式的优缺点，综合考虑以下因素

:

1、抽样误差要求

2、调查对象特点

3、时间费用限制

并注意：多方法组合使用。

## 五、 样本轮换问题

固定样本调查出现的问题：

- 1、 样本老化
- 2、 被调查者合作态度
- 3、 被调查者行为影响

是否替换？ 替换多少？

新增样本应考虑费用的限制和调查的需要。

## 六、无回答问题

样本区分：

- 无回答层
- 回答层

问题：无回答层对样本的影响？

如：考试作弊人数调查、企业兼职人数调查。  
无回答的大多是作弊者、兼职者，影响较大。

解决措施：多次调查、改变调查方法、对调查结果进行修正。

返回单元

# 本单元小结

## ■ 小结

## ■ 思考题

1. 什么是抽样调查？它在什么条件下使用？
2. 抽样调查的步骤有哪些？
3. 抽样调查技术有哪些？
4. 分层抽样与分群抽样有何区别？
5. 配额抽样有哪些优缺点？
6. 抽样误差与非抽样误差产生的原因是什么，如何避免？
7. 确定样本容量的影响因素有哪些？

[返回单元](#)