

# 统计学

主讲：经济管理教学部 杨世鉴

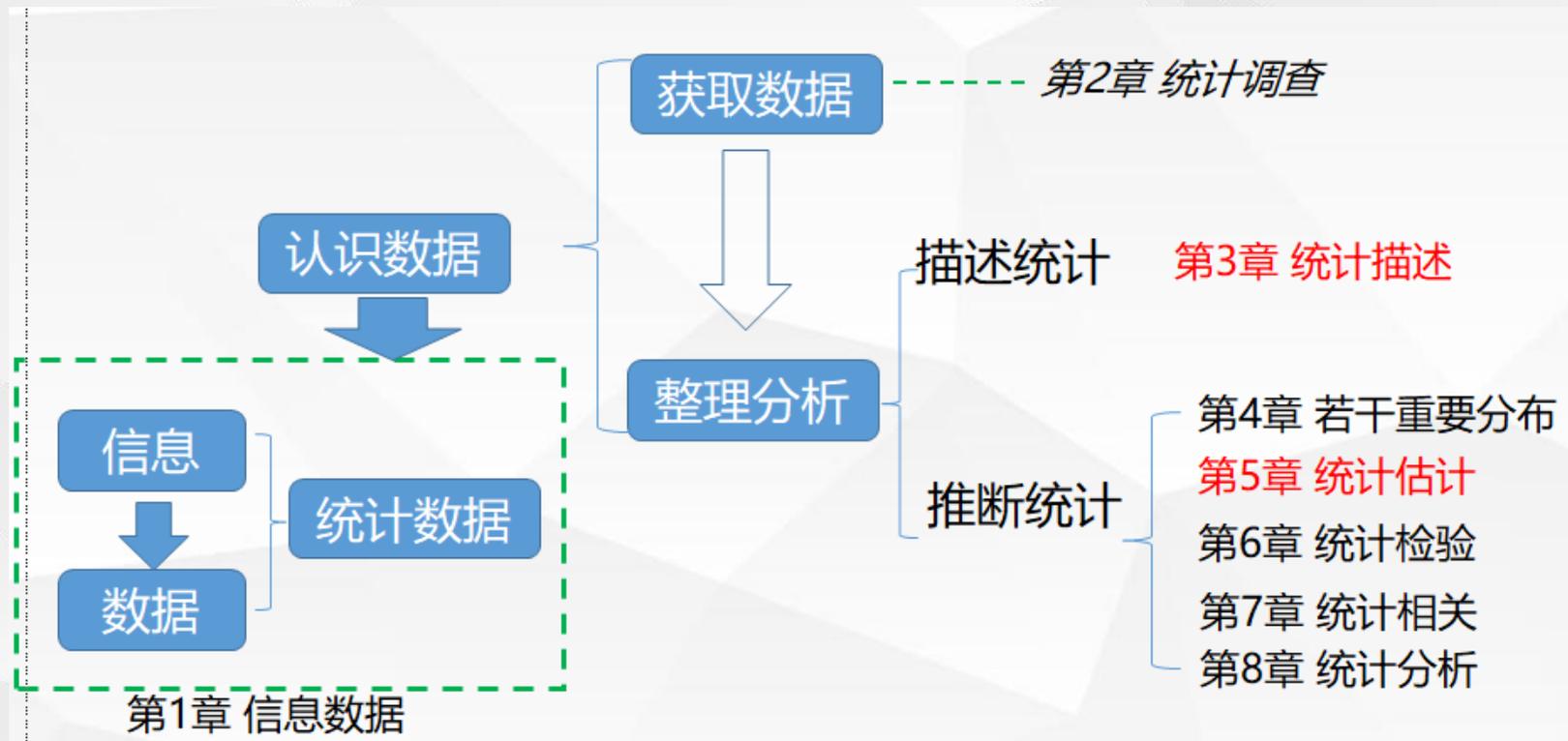


2

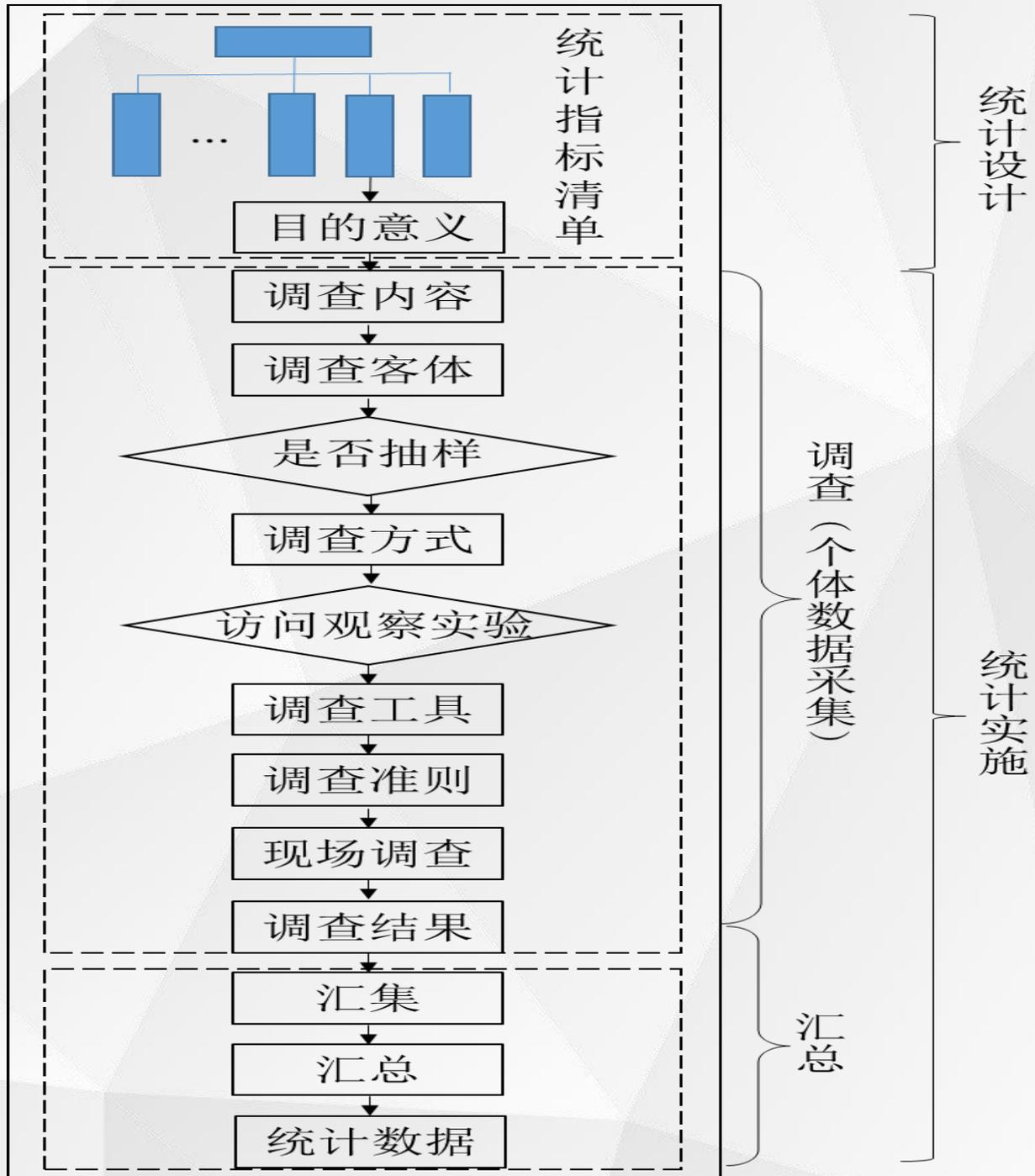
# 第2章 统计调查

# 本章导学

统计调查是根据调查的目的与要求，运用科学的调查方法，有计划、有组织地搜集数据信息资料的统计工作过程。



# 统计调查





**国家开放大学**

THE OPEN UNIVERSITY OF CHINA

**01**

**调查设计（变量设计）**

**02**

**数据采集**

**03**

**抽样设计**

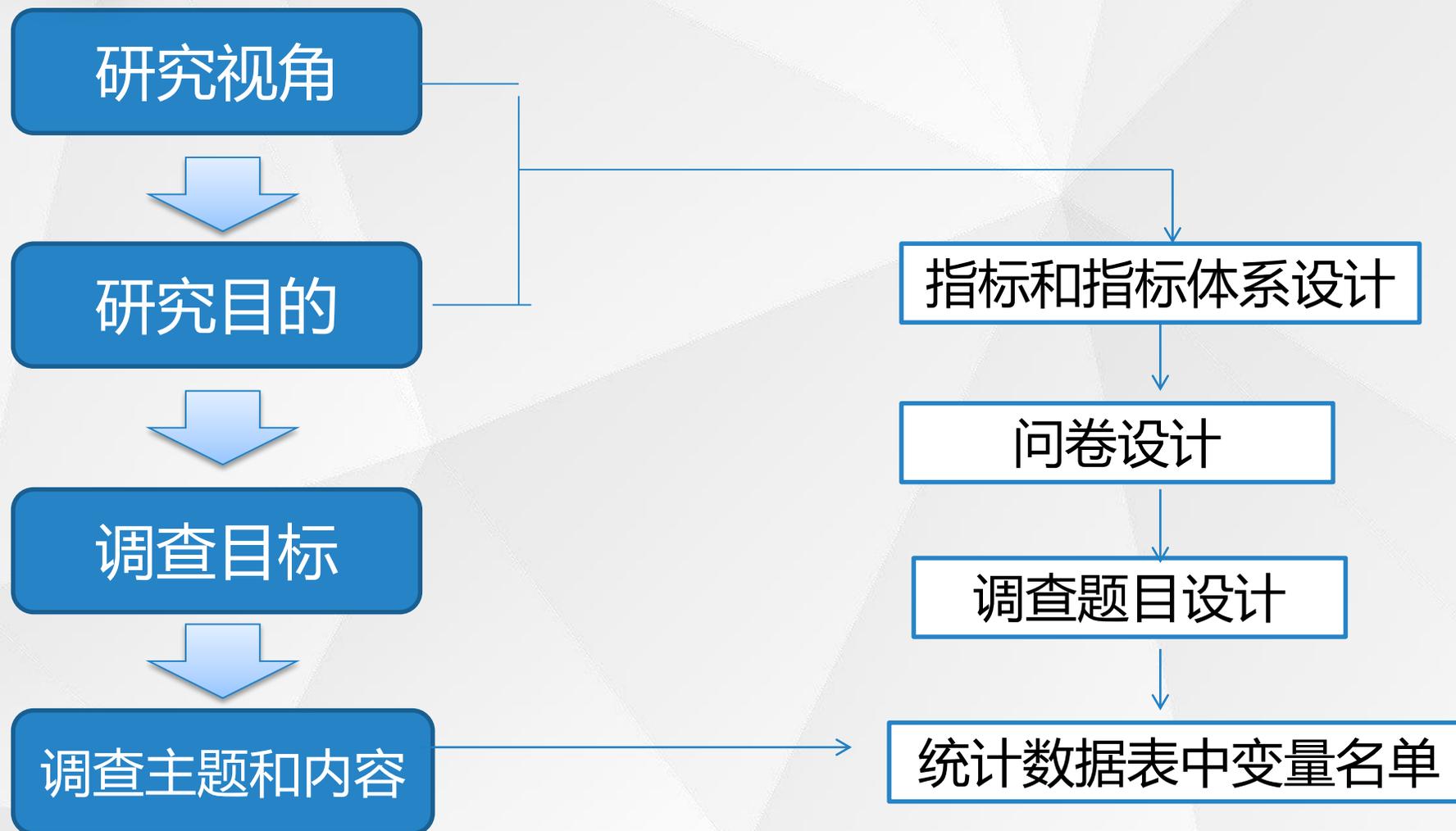
调查设计和现场实施是个体数据采集的两个阶段。

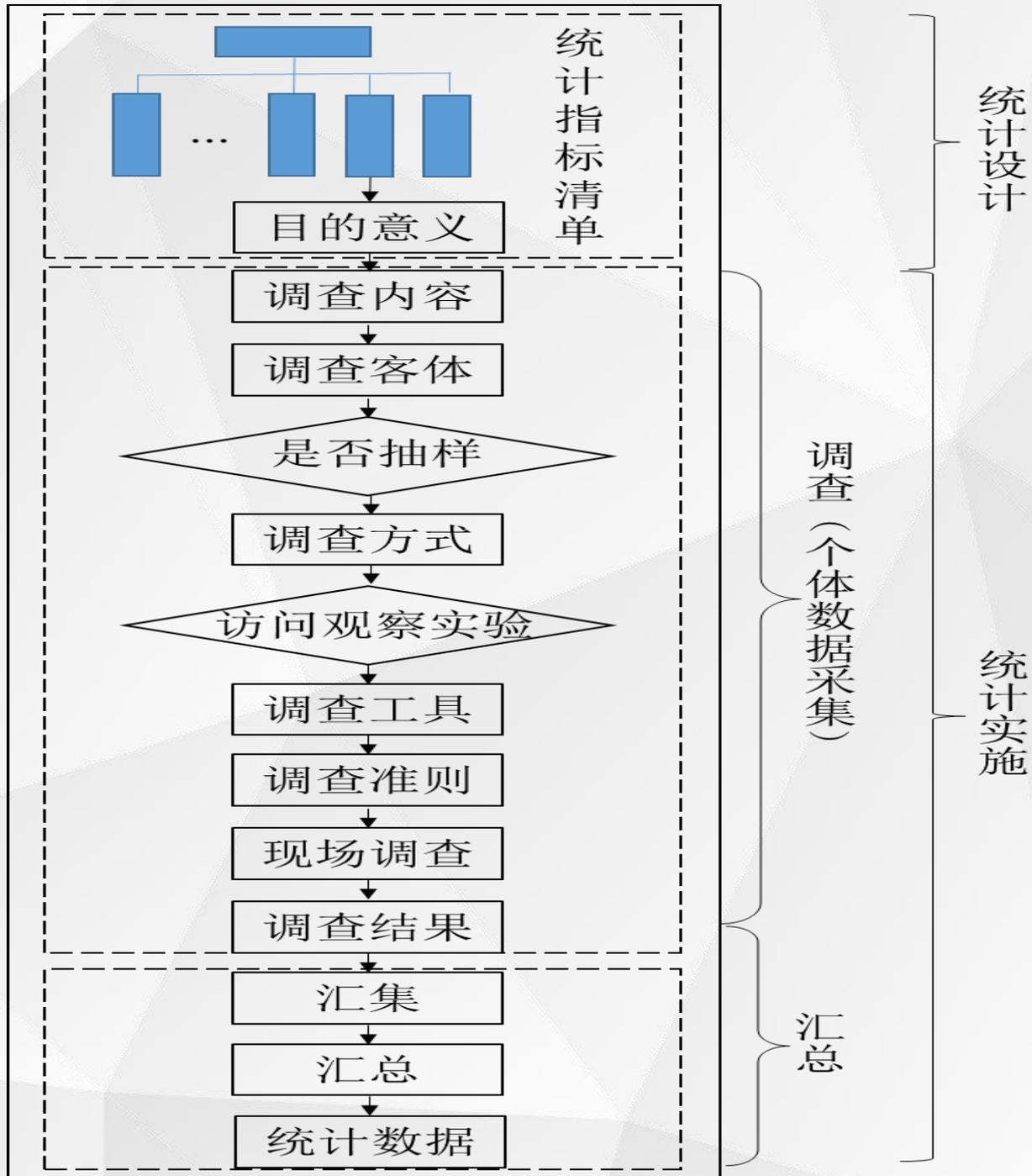
统计调查设计包括调查项目清单、调查表或问卷设计、调查客体确定、是否抽样、具体抽样方式、调查方法、调查工具安排、调查准则制订、调查人员组织、时间、步骤程序、进度安排等内容。

变量设计是调查设计的最重要阶段。这个阶段解决的是整个调查的战略性问题，确定调查所要针对的究竟是哪些变量？目标决定手段，只有确定目标后，才能制订实现此目标的方案和手段。

# 01

## 调查设计 (变量设计)





# 统计调查

统计数据的获取过程包含**调查**和**汇总**两个阶段

1. 调查就是调查主体针对调查客体，就调查内容所圈定的属性或变量，采取特定的调查方法，使用相应的调查工具，遵从规定的调查准则和程序，采集调查客体中的个体属性值作为调查结果即个体数据的过程或活动。（本章主要讲授的内容）

2. 汇总的本质是将调查到的所有个体及其属性值汇集而后**进行整理**计算得到总体的分布或分布特征称为汇总。（第三章统计描述的内容）

统计调查具有八个要素：调查主体、调查客体、调查内容（项目）、调查方法、调查工具、调查准则、调查程序、调查结果。

## 二、统计调查要素

### (一) 调查主体

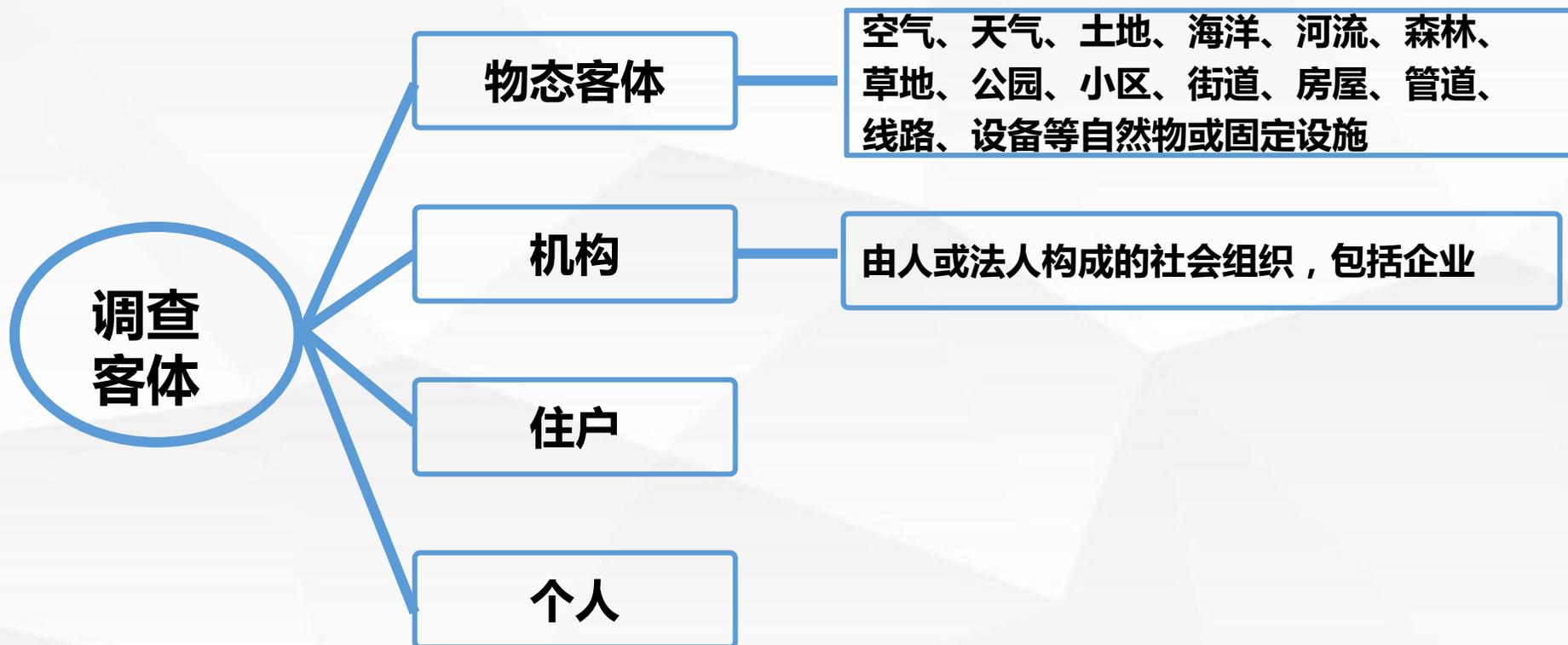
调查主体是发起或从事个体数据采集的人或由人构成的组织。

调查主体有三个层次：一是调查活动策划者或发起者；二是组织者；三是执行者。

策划者或发起者通常是最终用户；组织者是调查活动的管理者，确定调查目标、调查内容，配置调查资源，安排时称进度等；执行者是调查的实施者，确定调查项目和调查准则，选择调查工具，采集信息，最终获得结果。三个层次的主体彼此的角色划分不是绝对的，某些时候甚至是三合一的。

## (二) 调查客体

调查客体即被调查的所有个体。调查客体是受查者，可能是人、可能是住户、可能是机构也可能是物，某种自然现象或人为活动。



## (二) 调查客体

注意：只有调查客体才是前面所提及的信息两个构成要素里的实体，调查主体不属于信息的构成要素，因而不能归于实体范围，除非其出现于另外的调查且居于客体之位。

调查客体有时也指受查的个体而不限于受查个体的集合，但**调查客体多数场合与总体是相同的概念**，统计学一般记总体规模为 $N$ 。

### （三）调查内容（项目）

调查内容即变量清单，调查项目即调查内容的展开，是更详细的变量清单

### （四）调查方法

调查方法指选择个体及采集个体属性值的途径。

### （五）调查工具

调查工具，调查主体中的执行者采集信息所使用的各种工具。

假如调查客体属于人或人所构成与控制的群体与机构，调查问卷、调查表和调查记录表单等表格常用来记载调查结果，而调查客体是自然物或人为活动时，除了这些表格，还要使用各种试验工具，例如温度计，血压计，血糖仪，X光机、CT机，监视器，计数器，激光测量仪等各种度量衡器具，甚至包括非常复杂的设备，如卫星、海洋调查船舶、地质钻探设备、常规飞机无人机，汽艇等等。不难看出，调查工具的选择与调查客体是相关连的。

### （六）调查准则

调查准则，是为了确保调查目标顺利实现的一整套规范，包括调查采用或遵循的程序，流程，方法，标准，规则，注意事项等，各个不同领域的调查往往都有自己的专业规范，例如“地面气象观测规范”，“海洋调查规范”，“地质调查规范”，“民意、市场及社会调查标准”、“全球定位系统城市测量技术规程”，“工程测量规范”。调查准则既是调查活动的指南，又是调查过程透明、调查结果可比的保证，是科学性和规范性的基础和体现。

### (七) 调查程序

调查程序则为调查的步骤规定，也可看作调查准则的延伸与泛化。

### (八) 调查结果

调查结果，是所采集信息的最终表现，主要内容是对应于特定实体之特定属性的具体表现，称之为属性值。

调查结果是实体、属性及属性值三者的统一体，是三者及其对应关系的记录。

属性值的形式因属性不同与实体不同而多种多样，包括数字、文字、图形、图像、声音等人类感观和思维能够辨识的形式。调查结果又称为数据（data），也就是说数据是实体、属性及属性值三位一体的公称。

统计调查结果：统计数据

变量 样品	$Y_1$	$\cdots$	$Y_j$	$\cdots$	$Y_p$
$Y_{(1)}$	$y_{11}$	$\cdots$	$y_{1j}$	$\cdots$	$y_{1p}$
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
$Y_{(i)}$	$y_{i1}$	$\cdots$	$y_{ij}$	$\cdots$	$y_{ip}$
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
$Y_{(n)}$	$y_{n1}$	$\cdots$	$y_{nj}$	$\cdots$	$y_{np}$

## 数据表示例——地震观测数据表

序号	纬度	经度	深度km	里氏震级	观测台站编号
1	-20.42	181.62	562	4.8	41
2	-20.62	181.03	650	4.2	15
3	-26	184.1	42	5.4	43
4	-17.97	181.66	626	4.1	19
5	-20.42	181.96	649	4	11
⋮	⋮	⋮	⋮	⋮	⋮
996	-25.93	179.54	470	4.4	22
997	-12.28	167.06	248	4.7	35
998	-20.13	184.2	244	4.5	34
999	-17.4	187.8	40	4.5	14
1000	-21.59	170.56	165	6	119

属性

实体

属性值

说明：1964年以来斐济附近 1000次地震发生的位置。

数据阵包含了调查所要获得的所有信息，调查客体（即总体）中包括的全部实体（即个体），所关注的全部属性（即变量或调查项目），和所有实体所有属性所对应的全部属性值（变量值），应有尽有。

虽然表面上看，调查的诸要素中，数据表里只显示了调查客体、调查项目和调查结果，调查的主体、方法、工具和准则四个要素并未显性地表现在数据表中，但这不等于四个要素不存在，不过却是人们使用数据库时容易忽视这种存在的部分原因。饮水不忘挖井人，数据表的形成离不开调查主体费心费时费力费钱，选择合适的科学方法，严格依照调查准则，正确使用调查工具的工作。

- 无论是普通的统计工作还是统计分析都离不开**统计数据**。
- 统计数据是一切一切的基础。
- 统计数据的特性在于总体规模通常很大，大得只能放到仓库里，称为**数据库**。
- 统计数据库具有清晰的棋盘式结构，称为**结构化数据**。

未标题2.sav [数据集1] - IBM SPSS Statistics 数据编辑器

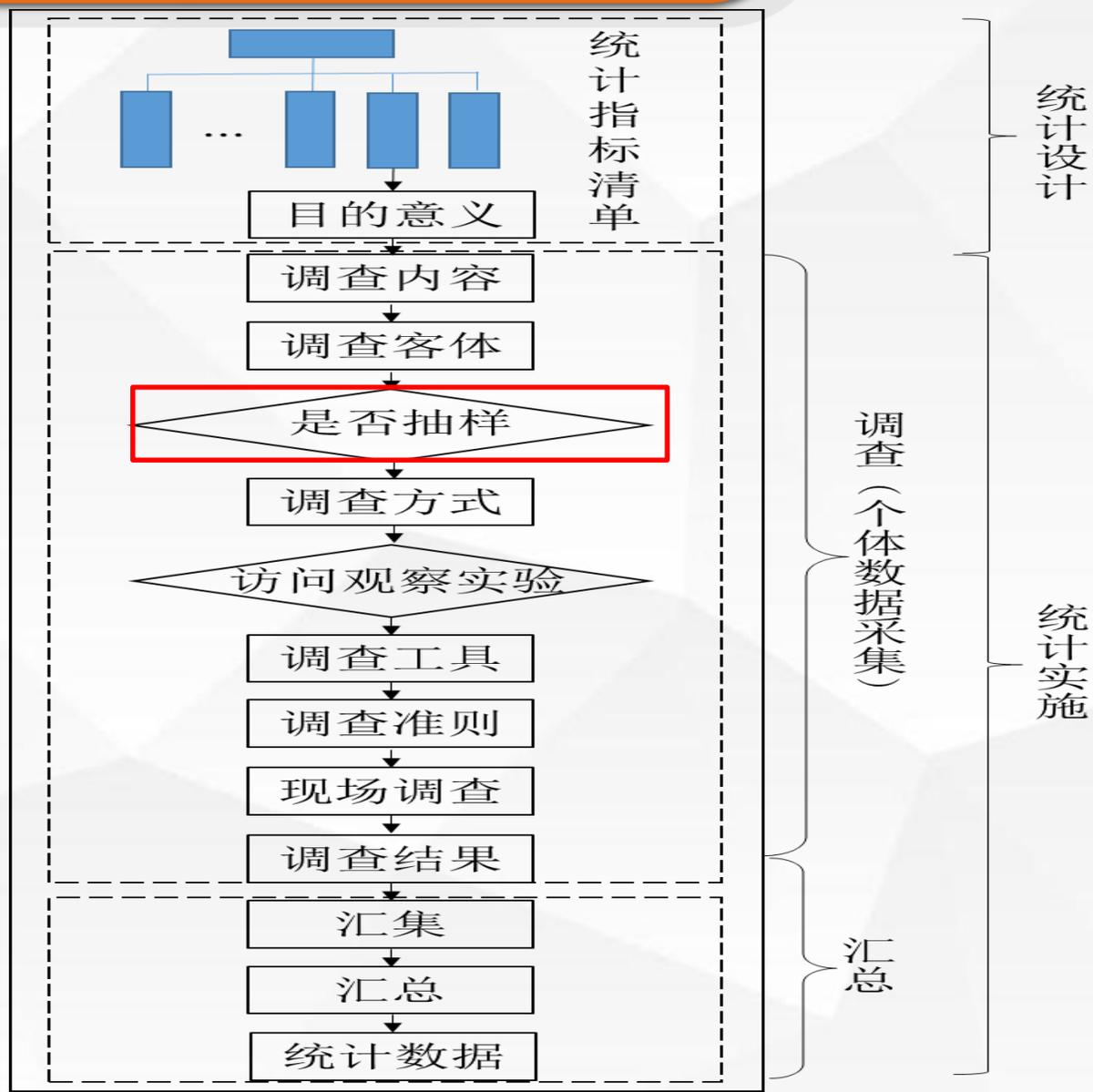
(E) 编辑(E) 视图(V) 数据(D) 转换(T) 分析(A) 直销(M) 图形(G) 实用程序(U) 窗口(W) 帮助(H)

	名称	类型	宽度	小数	标签	值	缺失	列	对齐	测量	角色
1	性别	字符串	3	0		无	无	3	左	名义(N)	输入
2	年级	数值	12	0		无	无	6	右	名义(N)	输入
3	学部	字符串	12	0		无	无	12	左	名义(N)	输入
4	请问您最近压力大吗...	数值	12	0		无	无	12	右	名义(N)	输入
5	请问您的主要压力来...	数值	12	0		无	无	12	右	名义(N)	输入
6	请问您的主要压力来...	字符串	12	0	请问您的主要压...	无	无	12	左	名义(N)	输入
7	请问您毕业之后您是否...	数值	12	0		无	无	12	右	名义(N)	输入
8	请问您换工作的原因...	数值	12	0		无	无	12	右	名义(N)	输入
9	请问您换工作的原因...	字符串	27	0	请问您换工作的...	无	无	27	左	名义(N)	输入
10	关于内在素质，您认...	数值	12	0	关于内在素质， ...	无	无	12	右	度量	输入
11	关于内在素质，您认...	数值	12	0	关于内在素质， ...	无	无	12	右	度量	输入
12	关于内在素质，您认...	数值	12	0	关于内在素质， ...	无	无	12	右	度量	输入
13	关于内在素质，您认...	数值	12	0	关于内在素质， ...	无	无	12	右	度量	输入
14	关于通用能力，您认...	数值	12	0	关于通用能力， ...	无	无	12	右	度量	输入
15	关于通用能力，您认...	数值	12	0	关于通用能力， ...	无	无	12	右	度量	输入
16	关于通用能力，您认...	数值	12	0	关于通用能力， ...	无	无	12	右	度量	输入
17	关于通用能力，您认...	数值	12	0	关于通用能力， ...	无	无	12	右	度量	输入
18	关于通用能力，您认...	数值	12	0	关于通用能力， ...	无	无	12	右	度量	输入
19	关于通用能力，您认...	数值	12	0	关于通用能力， ...	无	无	12	右	度量	输入
20	关于专业能力，您认...	数值	12	0	关于专业能力， ...	无	无	12	右	度量	输入
21	关于专业能力，您认...	数值	12	0	关于专业能力， ...	无	无	12	右	度量	输入
22	关于工作生活的平衡...	数值	12	0	关于工作生活的...	无	无	12	右	度量	输入
23	关于工作生活的平衡...	数值	12	0	关于工作生活的...	无	无	12	右	度量	输入
24	请问您所在的单位性...	数值	12	0		无	无	12	右	名义(N)	输入

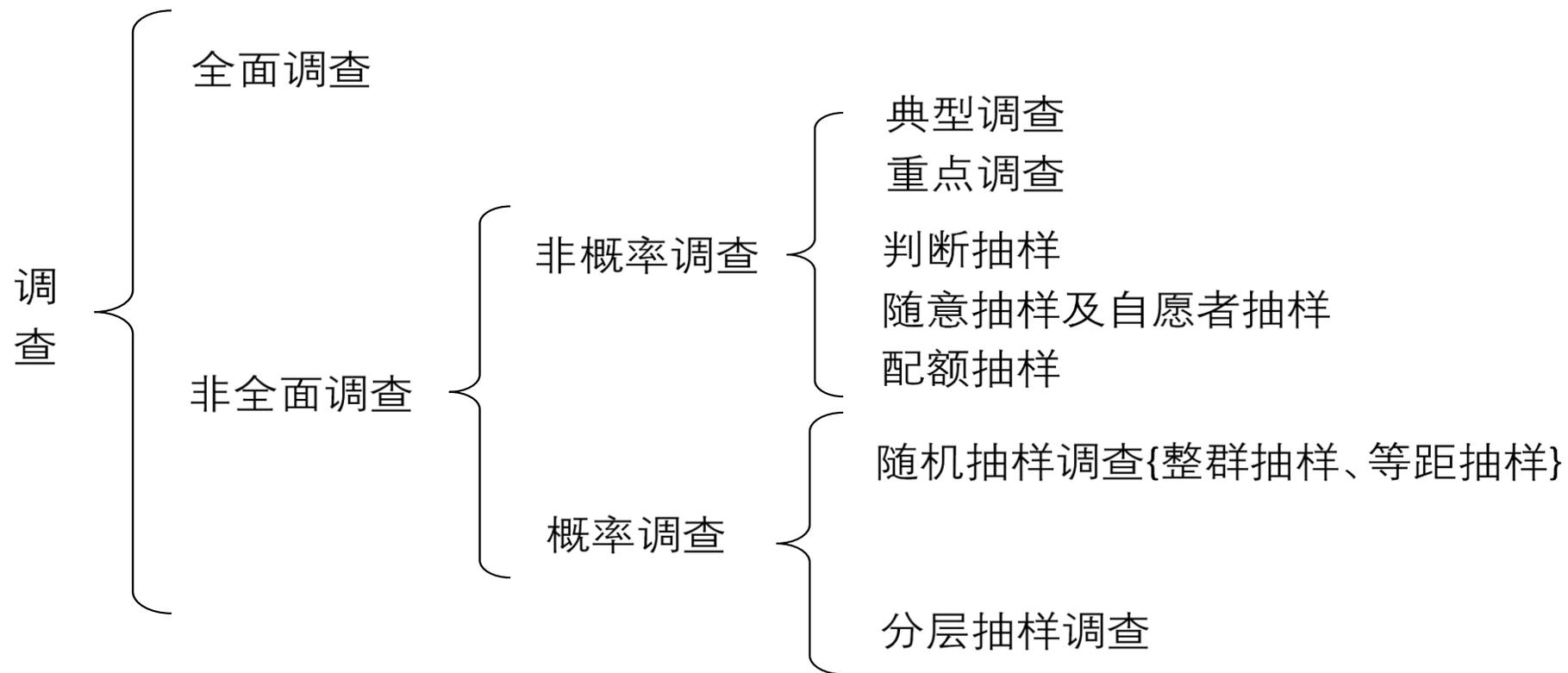
1

数据编辑器 变量视图(V)

IBM SPSS Statistics Process



# 一、统计调查的分类

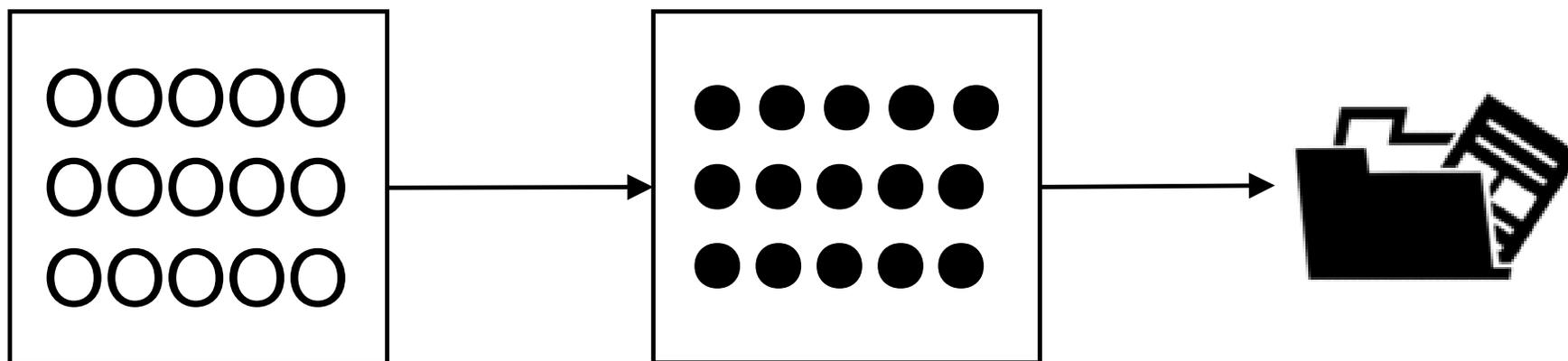


## (一) 全面调查

对总体的全部个体进行信息采集。采取全面调查，如下图所示。

例如要掌握全国人口总数及构成情况，就需要对全国每一户居民进行调查。各种普查和多数定期统计报表都属于全面调查。

全面调查需要耗费较多的人力、物力、财力和时间，因此通常只用来反映最基本最重要的社会经济现象资料。



## **(二) 非全面调查**

统计调查的总体规模一般很大，统计调查也因这一特点特别费时费力费钱，受限于人力物力财力，除非极特别的场合会对总体的全部个体进行信息采集，我们大多只能通过所谓抽样抽取一部分个体进行信息采集。

因此，我们通常将非全面调查称为抽样调查。

## **(二) 非全面调查**

在以下几种场合里，抽样是非常必要的。

(1) 有些场合调查是破坏性的，例如灯泡、手机电池等产品寿命的测试实验；

(2) 有些场合是经费时间精力等条件不允许进行全面调查；有些场合个体差异较小，例如流水线上的手机、电脑、家电及其零部件和元器件等的产品质量检验，由于大多自动化程度高，工艺精密稳定，本来就不需要涉及太多个体；

## （二）非全面调查

在以下几种场合里，抽样是非常必要的。

（3）有些场合调查要求不高，或只是初步试验性的调查；

（4）有些场合调查误差的影响更大，如果资源更多地配置于现场调查环节，集中足够的人力物力财力，确保调查精度高，则总的效果更好。

## (二) 非全面调查

选择部分个体进行信息采集有两种思路：

- ①根据已知的先验信息或知识选择有代表性的一个或一些个体；
- ②按所谓随机原则选择一些个体，即从总体中选择一些个体时不利用任何信息或没有任何信息可供利用。

在非全面调查中，按照是否利用先验信息或知识，选择有代表性的个体来划分，分类为：**非概率调查**和**概率调查**。

## (二) 非全面调查

### 1. 非概率调查

非概率调查是根据既有信息或知识选择有代表性的若干个体进行信息采集。虽然根据样本调查的结果也可在一定程度上说明总体的性质、特征，但不能从数量上推断总体。

包括典型调查、重点调查、判断抽样、随意抽样、配额抽样等。下面我们重点介绍两种类型，典型调查和重点调查。

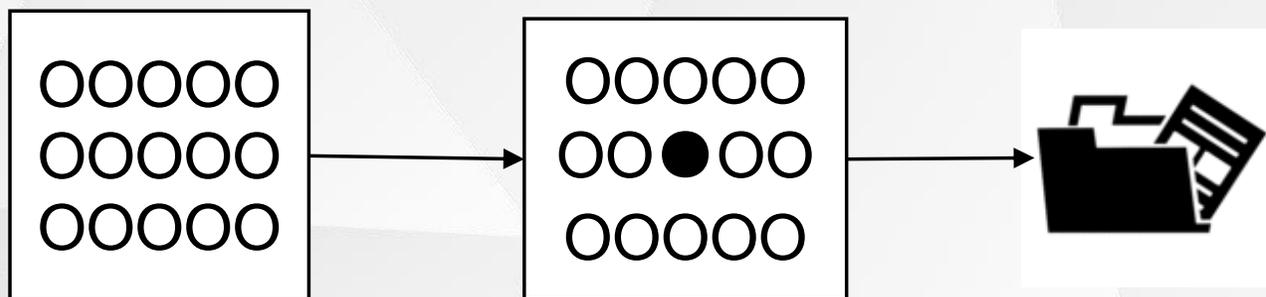
## 二、非概率抽样

### 1. 典型调查

典型调查是根据调查目的，在对研究对象总体进行全面分析的基础上，有意识地从中选取若干个总体单位进行系统周密调查研究的一种非全面调查。

进行典型调查的主要目的不在于取得社会经济现象的总体数值，而在于了解与有关数字相关的**生动具体情况**。

其特点是：调查单位少，并且是调查者有意识选择出来的；调查内容具体细致；调查所需时间短，反映情况快。

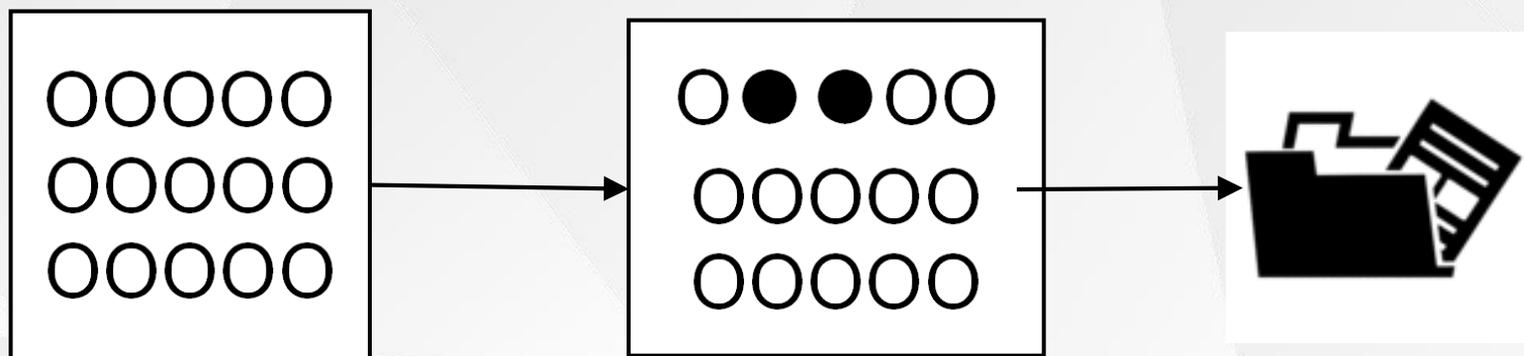


## 二、非概率抽样

### 2. 重点调查

重点调查是一种非全面调查，它是在调查对象中，选择一部分重点单位作为样本进行调查。

重点调查主要适用于那些反映**主要情况或基本趋势的调查**。



## 二、非概率抽样

思考：典型调查和重点调查的区别与联系？

分析：（1）联系：重点调查和典型调查都是非全面调查。

（2）区别：①选择对象的标准不同。典型调查选典型，重点调查选集中性的单位作为调查对象；②调查的主要目的不同，典型调查目的是认识事物的本质和规律，定性调查，重点调查多可以进行定量调查。

### 三、概率抽样

概率抽样，又称为随机抽样，是按照概率论和数理统计的原理从调查研究的总体中，根据随机原则来抽选样本，并从数量上对总体的某些特征作出估计推断，对推断出可能出现的误差可以从概率意义上加以控制。

在我国，习惯上将概率抽样称为抽样调查，统计学最终是随机抽样调查方式。

概率调查包括：简单随机抽样、整群抽样、等距抽样、分层抽样等。

## 三、概率抽样

### 1. 简单随机抽样

简单随机抽样也称为单纯随机抽样、纯随机抽样、SRS抽样，是指从总体 $N$ 个单位中任意抽取 $n$ 个单位作为样本，使每个可能的样本被抽中的概率相等的一种抽样方式。

简单随机抽样的特点是：每个样本单位被抽中的**概率相等**，样本的每个单位完全独立，彼此间无一定的关联性和排斥性。

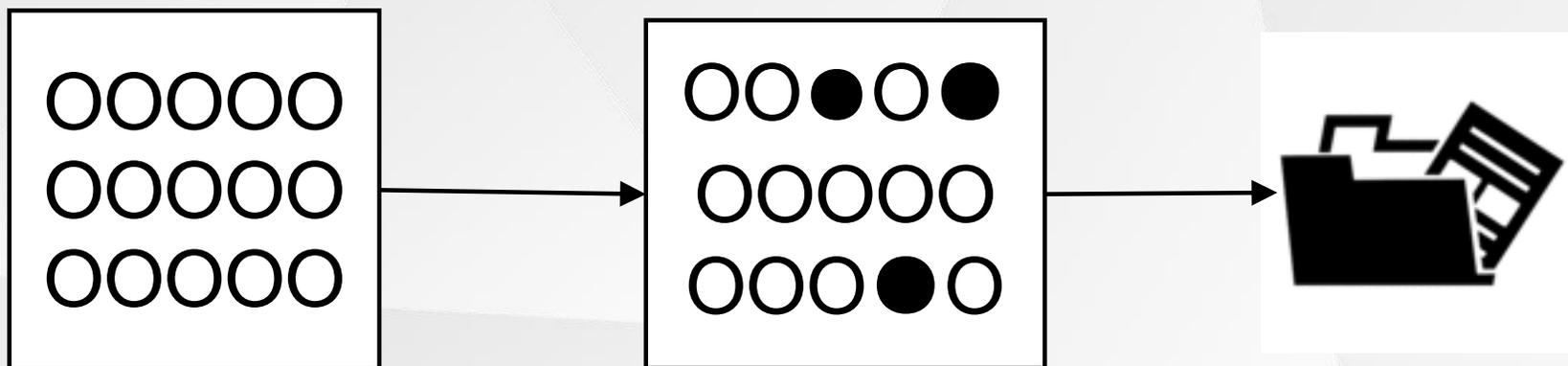
# 如何理解随机

- 最大的不确定谓之随机；
  - 天要下雨娘要改嫁，完全不能干预，随他去吧。
- 不确定性最大谓之随机性；
- 自然界不存在所谓随机现象。
- 随机的原因是完全缺乏任何先验信息。或不利用不试图利用任何先验信息。
- 随机抽样最好的解释是盲抽。
- 随机是统计学的基因！

### 三、概率抽样

#### 1. 简单随机抽样

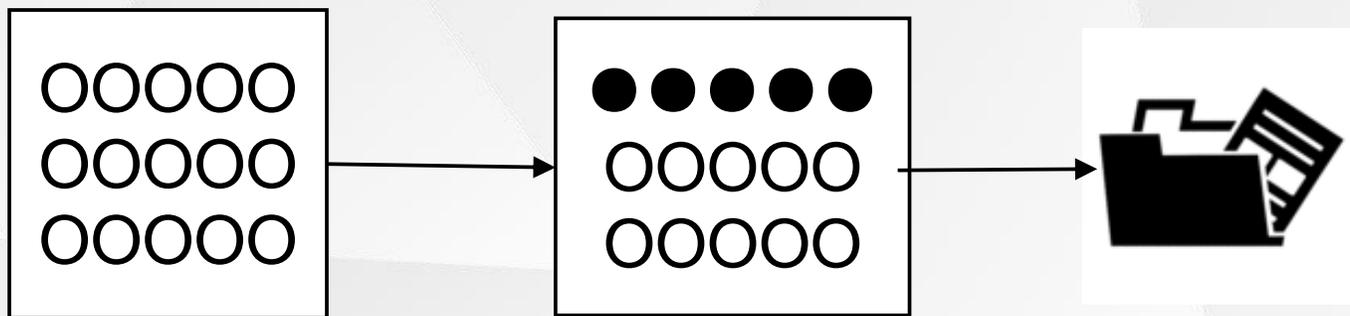
简单随机抽样**最基本的抽样方法**。分为重复抽样和不重复抽样。在重复抽样中，每次抽中的单位仍放回总体，样本中的单位可能不止一次被抽中。不重复抽样中，抽中的单位不再放回总体，样本中的单位只能抽中一次。社会调查采用不重复抽样。



### 三、概率抽样

#### 2. 整群抽样

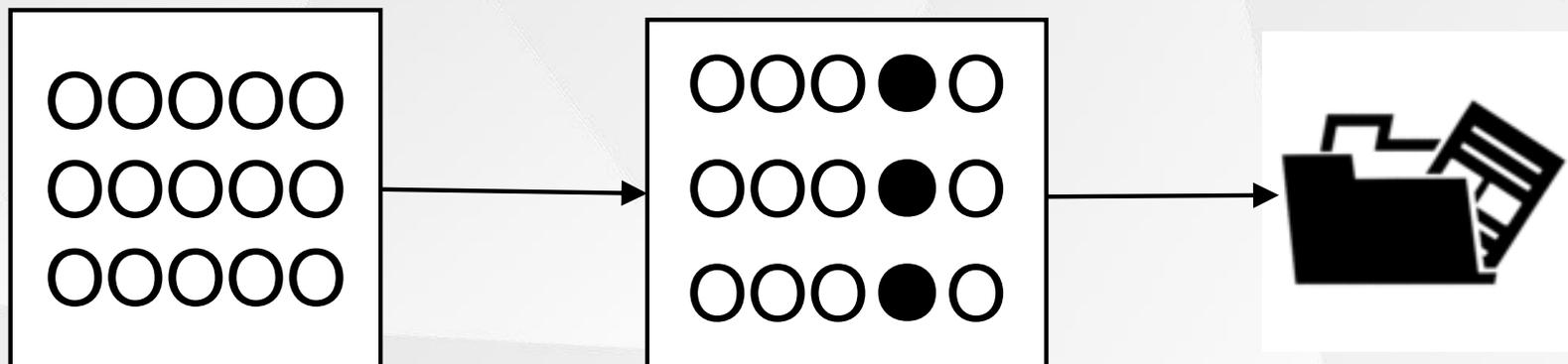
整群抽样是指整群地抽选样本单位，对**被抽选**的各群进行全面调查的一种抽样组织方式。例如，检验某种零件的质量时，不是逐个抽取零件，而是随机抽若干盒（每盒装有若干个零件），对所抽各盒零件进行全面检验。例如，调查中学生患近视眼的情况，抽某一个班做统计；进行产品检验；每隔8h抽1h生产的全部产品进行检验等。



## 三、概率抽样

### 3. 等距抽样

先将总体的全部单元按照一定顺序排列，采用简单随机抽样抽取第一个样本单元(或称为随机起点)，再顺序抽取其余的样本单元，这类抽样方法被称为等距抽样。等距抽样又称为机械抽样、系统抽样。等距抽样往往不能给出估计量的估计方差。

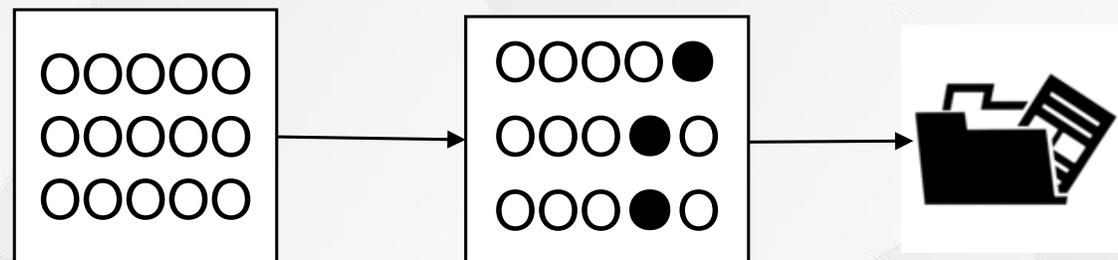


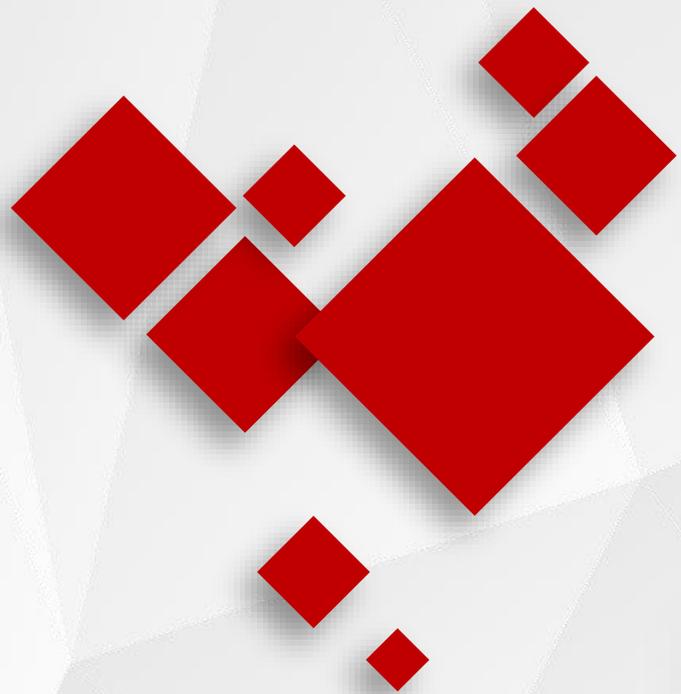
### 三、概率抽样

#### 4. 分层抽样

理想的条件总体具有相应变量在一个条件总体内内部差异小，不同条件总体之间差异较大的特点。具有这一特点的条件总体称为层，或层总体；将总体剖分为层总体的过程称为分层。

如在总体的每个层里独立进行抽样，则称为分层抽样。若记层数为 $L$ ，则要进行 $L$ 次抽样，这意味着要编制 $L$ 个抽样筐，明确 $L$ 个层总体的规模（抽样的前提之一是明确知道总体规模）。





2020

**感谢您的参与**

主讲人：杨世鉴

