


# 问卷调查中的抽样设计与样本代表性

- 报告人：郝一炜
- 汇报时间：2021.11.18



# 目录

- 前言

---

Introduction

- 思维导图

---

Mind Mapping

- 抽样设计

---

Sampling Design

- 样本代表性

---

Representativeness

- 总结

---

Summary



# 01

---

前言

Introduction

# 前言

市场

分析

调查

## 说文解字

- 1** 服务于企业的市场决策or从社会效益角度讨论某些热点问题
- 2** 调查环节既是获取数据的过程，又是连接案例背景与统计分析的纽带，包含诸多复杂的技术要点
- 3** 分析环节既要熟练运用统计方法工具，又要深刻理解所研究的问题

# 前言

市场

分析

调查

## 抛砖引玉

- 1** 本次分享的内容以样本代表性为入口，带领大家进入抽样技术的广袤天地
- 2** 抽样调查技术是统计学知识体系中古老而又极具魅力的一个学术分支（理论+实际）
- 3** 本次分享的内容在系列公益课程中起到承上启下作用，大家在学习后面各节课程时，可反复思考本次课程的内容



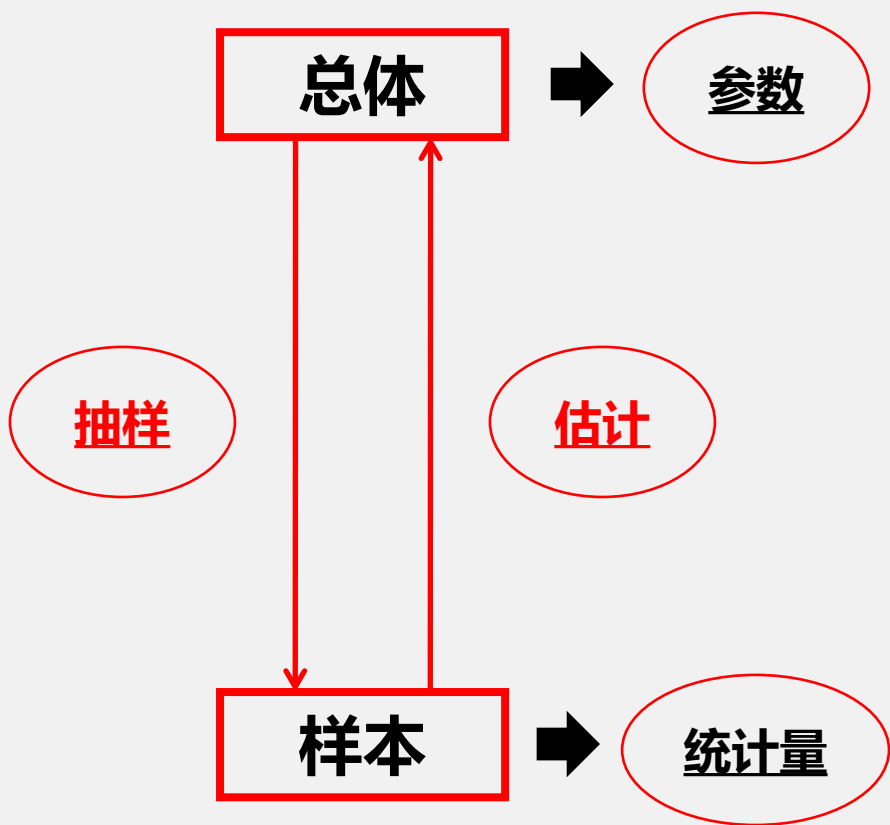
02

---

思维导图

Mind Mapping

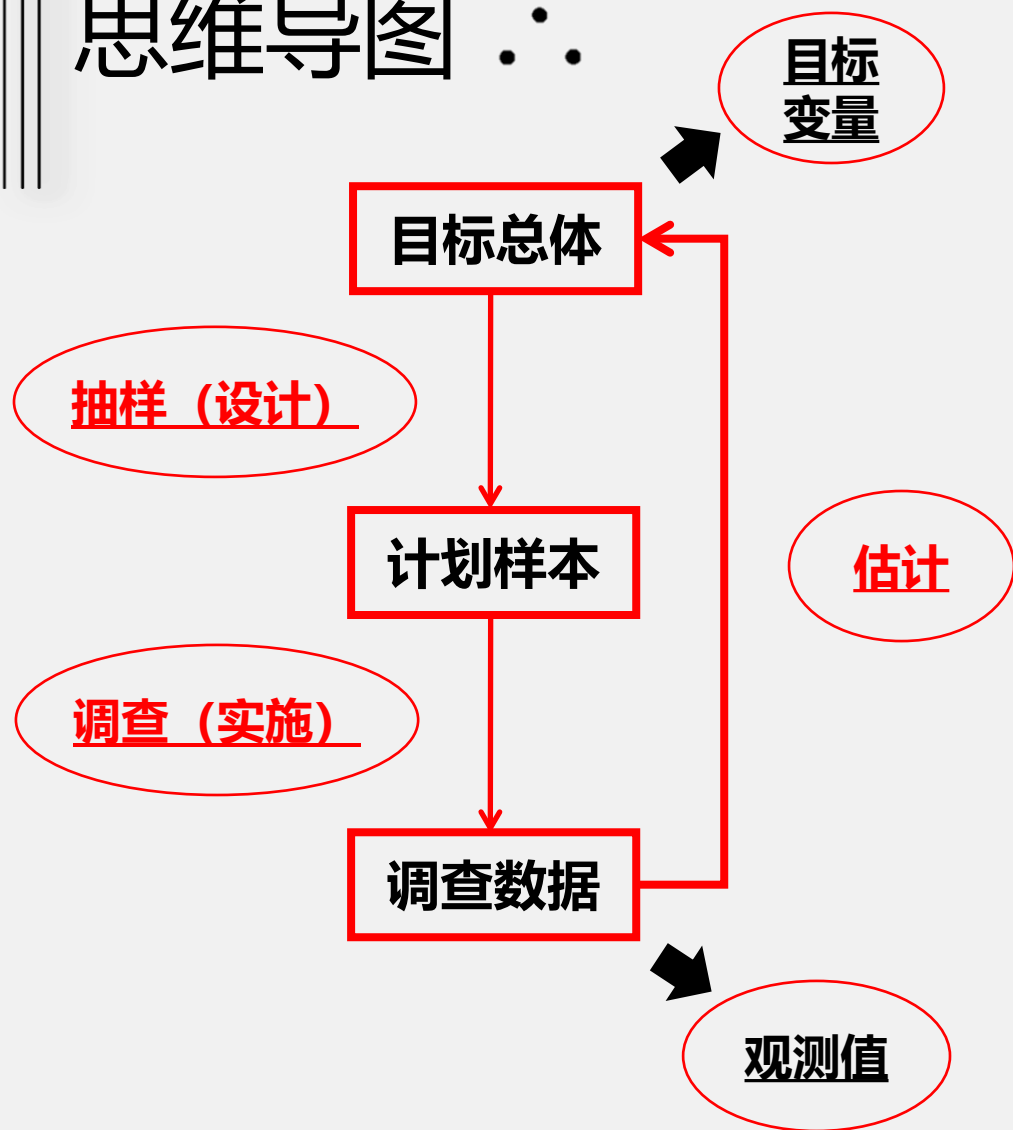
## 思维导图 ∴



## 普适理论框架

- 1 统计推断的本质：从总体中获得样本（抽样），并基于样本信息对总体特征加以推断（估计）
- 2 总体的参数是固定的量，样本的统计量本质上是一个随机变量
- 3 注意区分数理统计中的“样本”和抽样技术中的“样本”是不同的概念

# 思维导图 ∴

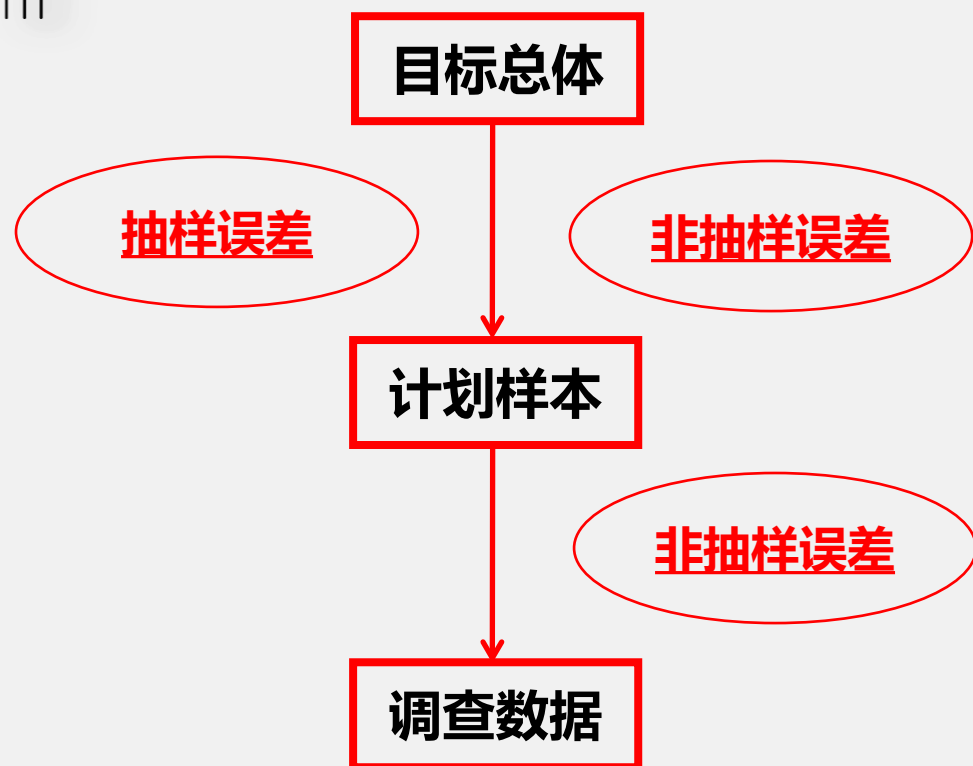


## 抽样技术理论框架

- 1 抽样（设计）环节设定了从总体中选取样本的规则和机制
- 2 调查（实施）环节通过调查访问从样本中的被调查对象收集目标变量的样本观测值，汇总为调查数据
- 3 利用调查数据进行统计推断，对总体参数进行估计



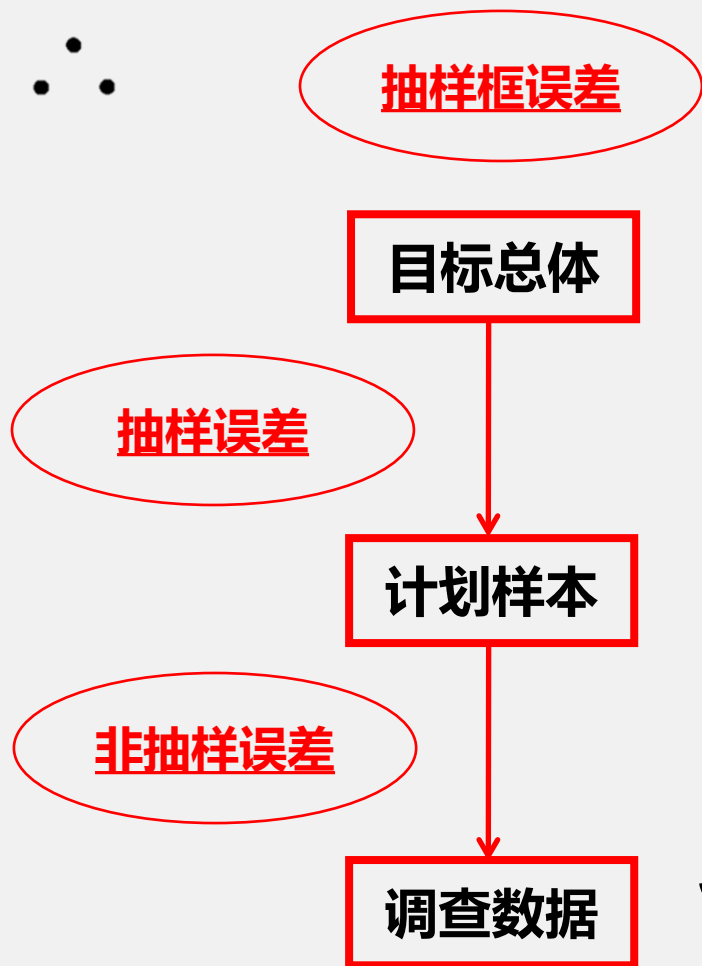
## 思维导图 ∴



## 从误差的角度考虑

- 1** 抽样误差的类型：系统误差（准确度）、随机误差（精度）
- 2** 非抽样误差的类型：抽样框误差、无回答误差、测量登记误差
- 3** 在抽样（设计）和调查（实施）环节均会产生误差，本课程关注的是抽样（设计）环节的抽样误差

# 思维导图



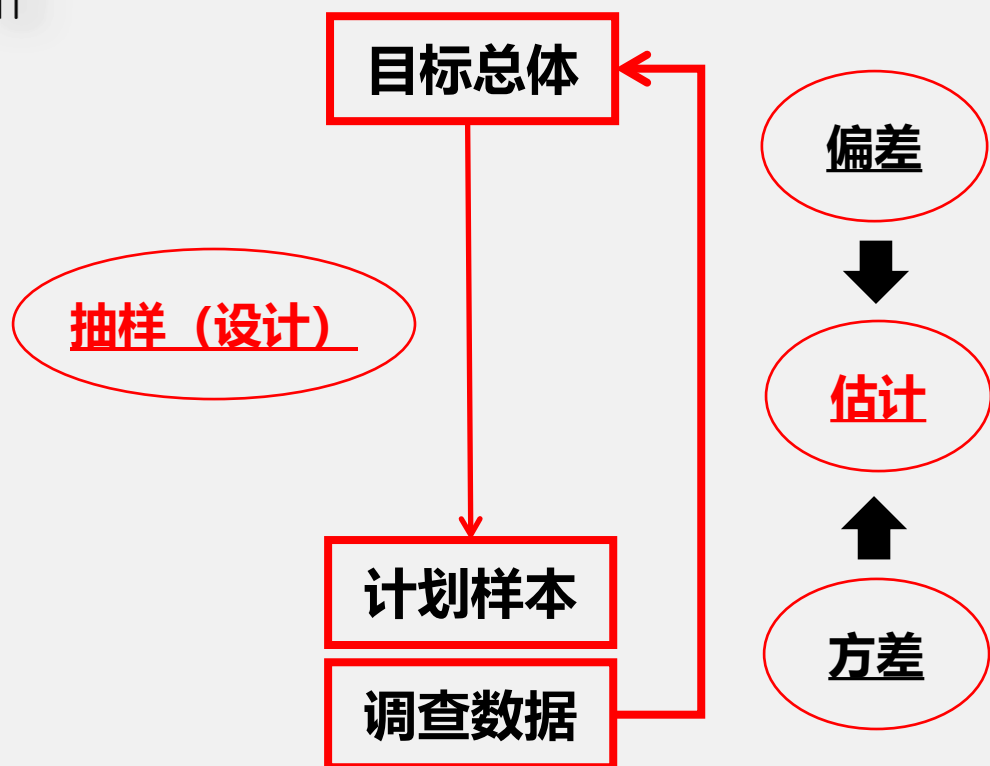
**系统误差：**选择性偏倚  
**随机误差：**样本随机性

**测量系统误差：**仪器失准、问卷效度差  
**测量随机误差：**测量结果的随机波动、作答问卷者心情的变化  
**无回答误差：**部分问题无回答、拒访

## 考虑误差来源

- 1 控制选择性偏倚（幸存者偏差）
- 2 问卷调查前，要通过预实验、因子分析、德尔菲等方法确保调查问卷的效度
- 3 在不考虑非抽样误差的情况下，系统误差和随机误差都会在估计过程中造成影响

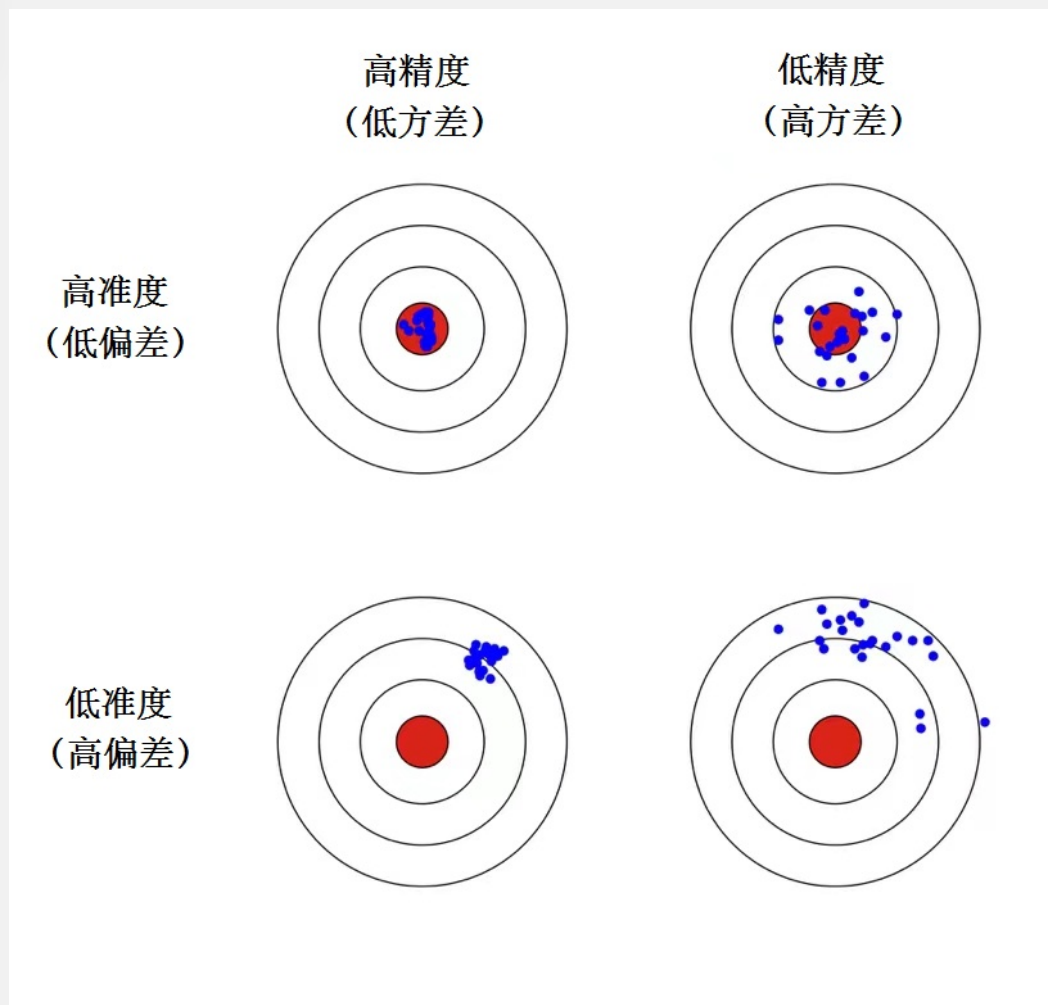
# 思维导图 ∴



## 引入估计阶段

- 1 假定：无非抽样误差，且调查数据能够真实反映计划样本的情况（抽样技术理论探讨）
- 2 评价估计量的统计性质：无偏性、一致性、有效性
- 3 通常考虑估计量的均方误差（方差与偏差结合考虑）

# 思维导图 ∴



## 考察估计量性质

1 均方误差=方差+偏差<sup>2</sup>

2 偏差与方差往往存在此消彼长的关系 (trade-off)

3 不妨牺牲部分无偏性来换取更高估计效率

# 思维导图 ∴


## 抽样设计

一套执行抽样调查的方法论层面的综合方案，包括：确定目标总体（抽样框制定）、选择抽样方法、样本量计算、选择估计方法以及评价估计量的精度和准确度等环节



顾名思义，样本代表性反映了样本能否对总体进行合理科学的刻画，提高样本代表性本质上是通过科学的抽样设计控制抽样误差

## 样本代表性



03

---

抽样设计

Sampling Design

# 抽样设计 ∴

## 抽样设计

抽样设计确定了从“目标总体”中获取“计划样本”的机制，确定了由样本预测总体的路径，是一个古典的统计学范畴  
(抽样+估计)



更为宽泛的概念，包含调查问卷设计、现场访问/线上调查计划制定、成本控制、人力资源管理、数据管理乃至社会心理学的技术细节

## 调查方案

# 抽样设计：抽样框编制 ∴



- 1.能够清晰划定总体单元范围的名录（户籍信息、工商登记、电话号码、人力资源部门等）
- 2.要求：完整而不重复
- 3.编制抽样框时需考虑到采集辅助信息
- 4.非概率抽样无法构建抽样框，但必须划定调查总体的范围和特征



# 抽样设计：选定抽样方法 ∴



**1.要正确理解“随机抽样”**

**2.概率抽样是抽样调查的金标准**

3.概率抽样：简单随机抽样、分层抽样、系统抽样、整群抽样、多阶段抽样、PPS抽样等

4.非概率抽样：方便抽样、配额抽样、滚雪球抽样

# 抽样设计：选定抽样方法

用什么抽样方法？

1

尽量创造条件实现概率抽样设计，此为上策

2

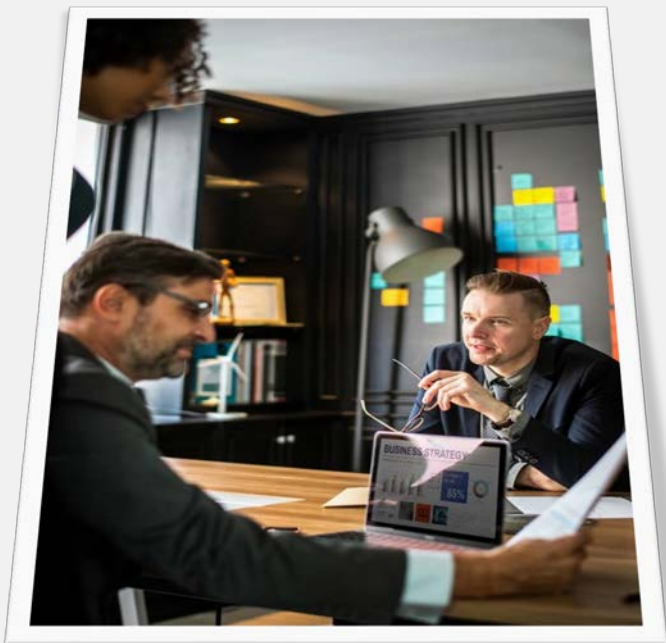
在概率抽样下，依照研究目的选择最佳抽样方法

3

非概率抽样在信息化高度发达的今天，具有节约成本、操作便捷以及短时间汇集大量样本观测的优势

4

非概率样本的估计和估计量性质评价问题属于学术界的前沿热点



# 抽样设计：确定调查样本量.



- 1.不同抽样方法的样本量需求不同
- 2.增加样本量会控制随机误差，但增加成本
- 3.简单随机抽样的样本量计算公式：

$$n_{srs} = \frac{Nz_{\alpha/2}^2 S^2}{Nd^2 + z_{\alpha/2}^2 S^2}$$

- 4.设计效应
- 5.考虑失访率

# 抽样设计：统计推断 ∴

1. 总体总值的估计
2. 总体均值的估计
3. 总体率的估计

总体参数估计

抽样方差估计

1. 由于样本的随机性，进行多次重复抽样得到的样本可能由不同的单元组成，因此估计结果也有波动
2. 估计量本质上是一个随机变量
3. 抽样方差反映了估计效率，也即估计的可靠程度

**估计量：简单估计、比率/回归估计、 $\pi$ 估计**

**总体参数估计量不同，方差估计量也随之不同**

# 抽样设计：统计推断 ∴

## 市场调查，解决什么问题

两个关键要素：调查&分析

调查分析结果给业界/社会创造的价值是什么？



## 市场调查统计推断的特点

**基于调查的目标变量，挖掘复杂的多变量相关关系**

### 目标变量

**目的：**了解现状

**价值：**为业界/社会提供战略/政策参考

**实现途径：**基于科学的抽样设计进行目标变量估计

### 多变量关系

**目的：**深入挖掘影响目标变量的因素

**价值：**提供实现盈利或解决问题的精细参考

**实现途径：**统计建模、多元统计方法、机器学习方法等

# 抽样设计：统计推断 ∴



G. Nathan, D. Holt. “The Effect of Survey Design on Regression Analysis”, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 42, No. 3(1980), pp. 377-386



04

---

样本代表性

Representativeness

# 样本代表性 ∴

1. 选用恰当的抽样方法
2. 灵活运用辅助变量

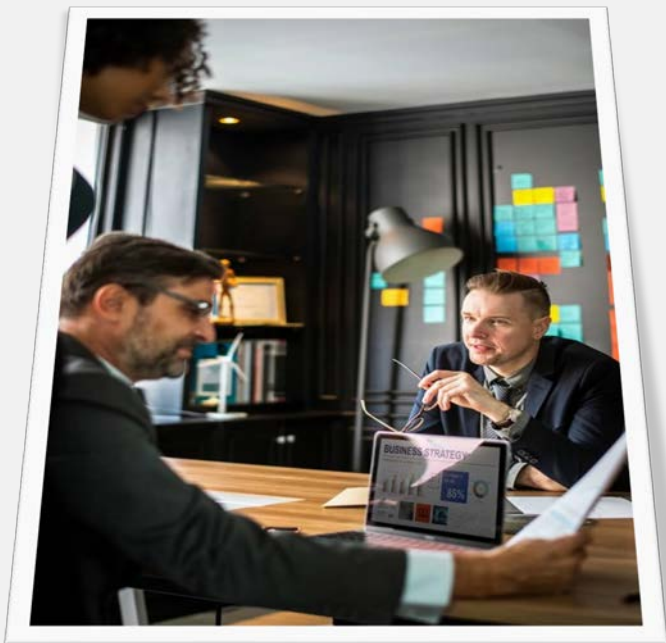
概率抽样

非概率抽样

1. 市调大赛中绝大部分案例都采用非概率抽样进行调查
2. 尽可能使用更多的统计信息对样本代表性进行弥补



# 样本代表性 ∴



## 如何保证样本代表性?

1

充分了解案例背景：  
行业知识+文献研究

2

深入研究总体特征：  
预调查+行业知识+文献研究+爬虫等

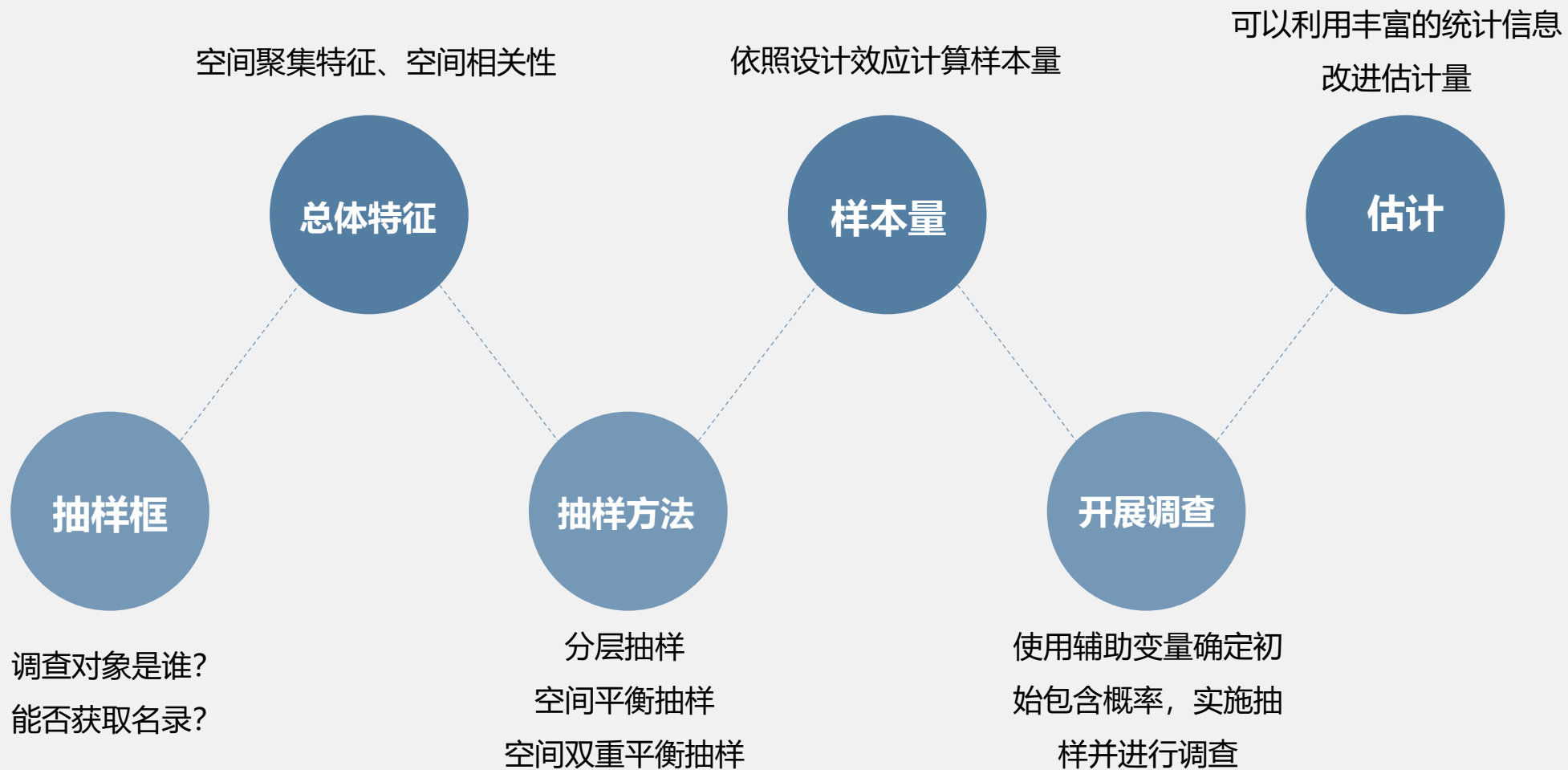
3

挖掘总体中变异性的根源：  
数据探索

4

抽样过程中充分利用统计信息：  
抽样技术

# 案例1：医疗机构急诊诊疗能力调查 ∴



# 案例1：医疗机构应急诊疗能力调查 ∴

## 启示1

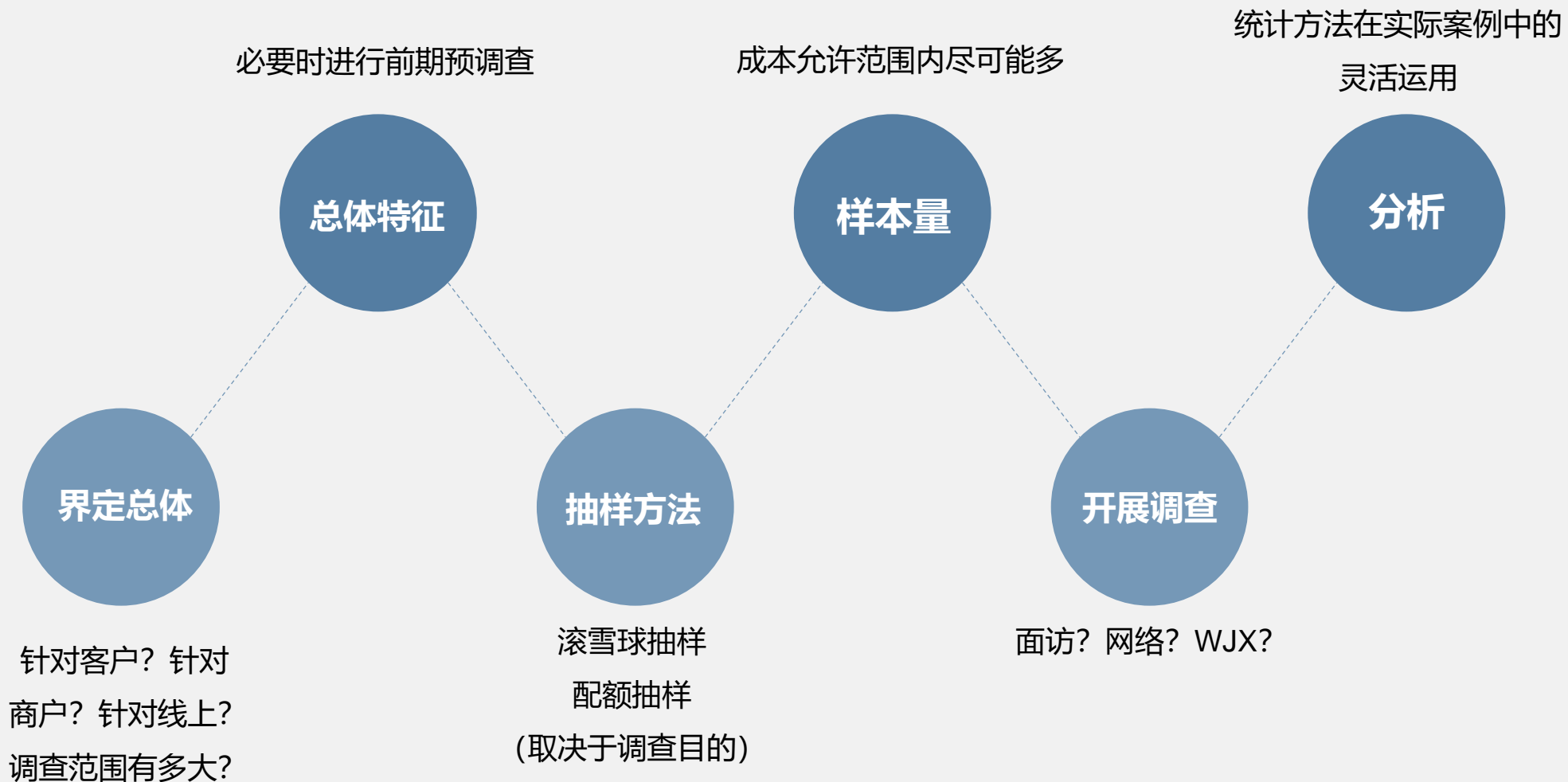
概率抽样能有效控制系统误差的产生，想要获得代表性足够好的样本，还需要深入研究总体特征，借助统计信息控制随机误差



无论调查的目的是为了估计目标参数，还是进行多变量相关关系研究，良好的样本代表性都是统计分析得到可靠结果的基础

## 启示2

# 案例2：剧本杀游戏消费的市场调查 ∴



## 案例3：盲盒消费情况的市场调查 ∴

### 启示1

通过研究总体特征，挖掘出调查推进过程中可能遇到的问题，面对的问题越复杂，样本代表性就越难保证，也越能体现出统计人员的能力和价值



撰写调查报告时，应当将设计调查方案时遇到的问题 and 解决问题的思路予以论述，体现思辨性

### 启示2



05

---

总结

Summary

# 总结 ..



## 制定科学的抽样设计，最大限度保证样本代表性

---

严格意义上的“设计 (Design)”针对概率抽样而言

非概率抽样虽然没有设计，但也可以**“控制”样本结构**

在大赛报告中详细讨论抽样策略和样本代表性的保证

鼓励大家在解决实际问题时应用前沿理论方法

The background is a complex abstract composition of various geometric and organic shapes. It features several overlapping circles and wavy, ribbon-like forms in shades of light blue and white. Scattered throughout are small, simple geometric icons: plus signs, triangles (both pointing up and down), and a circle with diagonal hatching. In the top left corner, there are three vertical parallel lines and a set of three dots. In the bottom right corner, there is a vertical column of small black dots and a white inverted triangle. The overall aesthetic is clean, modern, and professional.

•感谢各位同学!