



中國人民大學
RENMIN UNIVERSITY OF CHINA

统计学院
SCHOOL OF STATISTICS



第12届市调大赛第二期公益培训第二讲:

抽样调查的误差分析和质量控制

主讲人: 蒋妍

中国人民大学调查技术研究所

日期: 2021/11/18

为什么要重视调查中的误差和质量控制？

- 市场调研的评价：有价值、创造社会经济效益
- 隐形的敌人：误差
- 控制误差的胜利也是隐形的，但它直接影响调查的生命
- 误差的控制受成本约束，~~零误差~~是不客观、也不实际的
- 结论要靠谱，误差控制至关重要（必要非充分条件）

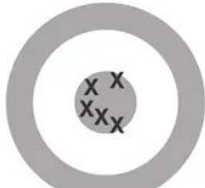
例如：医疗手术的杀菌消毒
火箭发射的可靠性



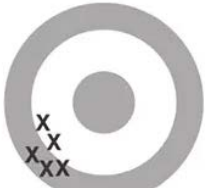


调查目标：靶心 (what, why, how)

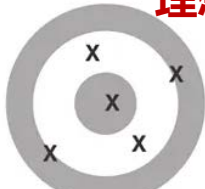
序号	分项	小项	分
			分项得分
1	研究选题	选题新颖	10
		结合实际	
2	文献研究	理论综述	7
		文案研究	
3	方案设计	调查方案设计	28
		样本代表性	
		问卷或其他技术设计	
4	调查实施	调查组织科学合理	20
		调查过程执行完整	
		数据处理严谨	
		调查质量控制严格	
5	分析与结论	选择方法正确	20
		数据运用充分	
		分析支持结论	
6	报告文本	结论实用性/适用性	15
		完整性和阐述充分性	
		格式规范性	
		文字表述清晰	
7	报告合计		100



A 队



B 队



C 队



D 队

理想队

偏差队

噪声队

偏差噪声共存队

误差可测、有控制



C 目录 ontent

1

定义：什么是调查误差

2

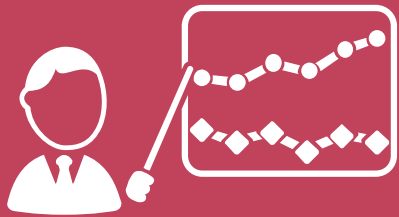
类型：调查误差从哪里来

3

控制：如何减小误差影响

4

建议：市调大赛作品的质控



PART ONE

什么是调查误差



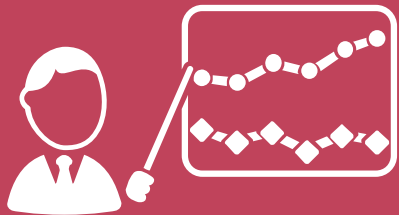
中國人民大學
RENMIN UNIVERSITY OF CHINA

統計學院
SCHOOL OF STATISTICS

调查中的误差

- 人类判断的错误=偏差+噪声
(丹尼尔.卡尼曼《噪声》)
- 视角：不是一次性的随机结果，而是可能的重复调查结果对真值的偏离
- 调查中的误差=系统性偏差 + 变异性
(有方向) (无方向随机偏离)
- 要靠谱，关键得有谱：尽量量化误差
- 误差的测量：均方误差 $MSE = E(\text{样本估计值} - \text{总体参数})^2$
 - 有些误差是可以估算的，例如随机样本的抽样方差
 - 有些误差的测量很难，例如回答的准确性、无回答偏差





PART TWO

调查误差从哪里来



中國人民大學
RENMIN UNIVERSITY OF CHINA

統計學院
SCHOOL OF STATISTICS

例如直播营销方式的效果研究（从测量的视角）

研究目标

- 概念：直播营销
- 效果（认知、态度、购买意愿、购买行为）
- 口径：主播的类型、电商平台
- 理论模型，控制变量

测量方法

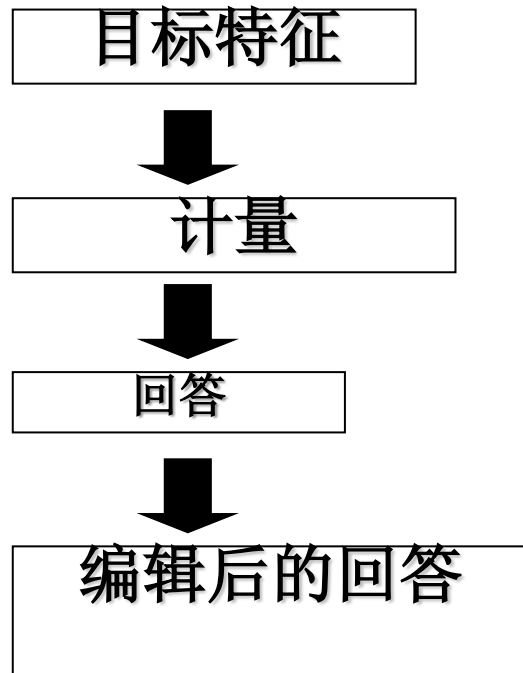
- 数据采集方式（场景的设计）
- 量表设计（信效度）
- 民意感受不等于真实的影响

受访者的回答

- 会受问卷设计、访问员、受访者、调查环境等影响
- 可能的虚假回答

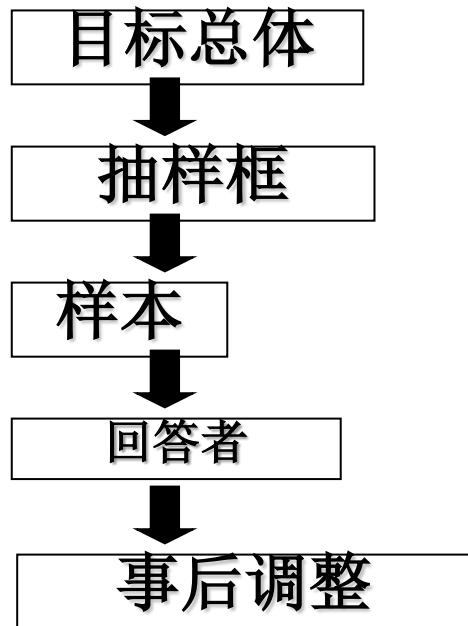
数据的事后编辑

- 异常值的判定

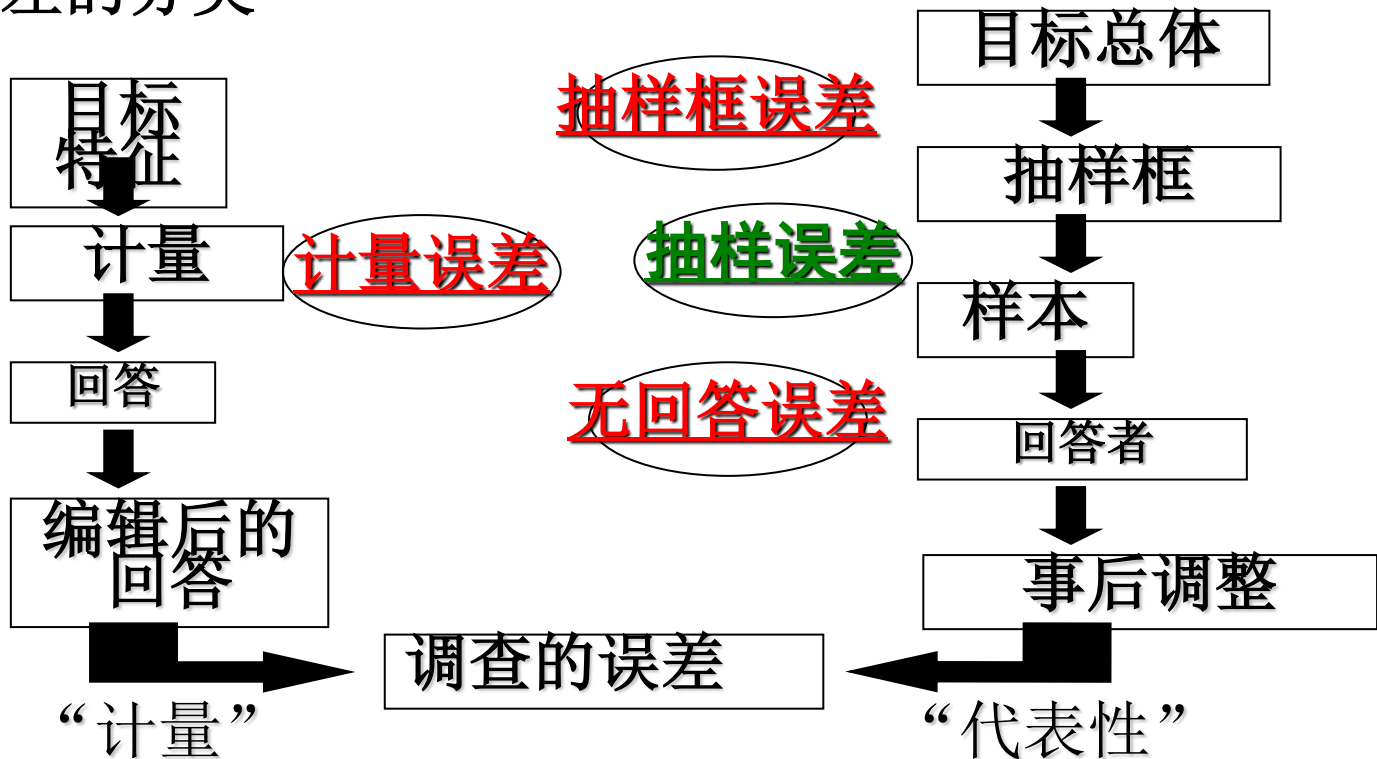


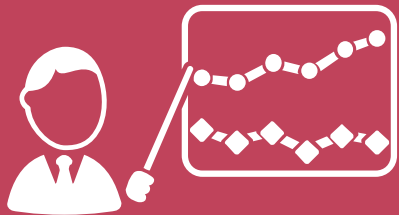
例如直播营销方式的效果研究（从样本代表性的视角）

- **研究单位的集合**
 - 所有的消费者？潜在消费者？（明确）
- **有可能被选入样本的总体单位**
 - 大学生群体的代表性？
 - 某平台的网民？（对总体的覆盖率）
- **待调查的单位集合**
 - 某学校的500个学生？（样本代表性）
- **成功计量的单位集合**
 - 完成调查的所有受访者（有没有同质性）
- **事后加权调整**
 - 权数的选择



误差的分类



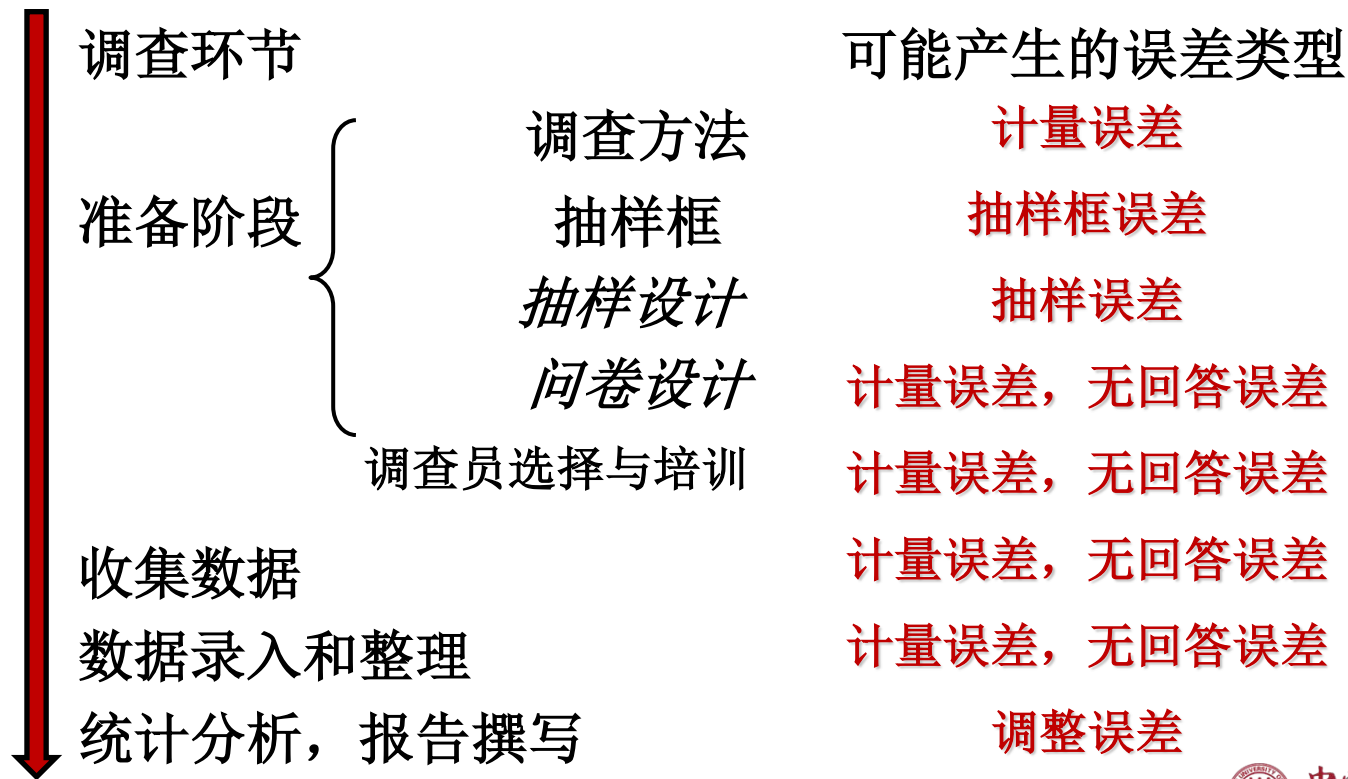


PART THREE

如何减小误差 (质量控制)

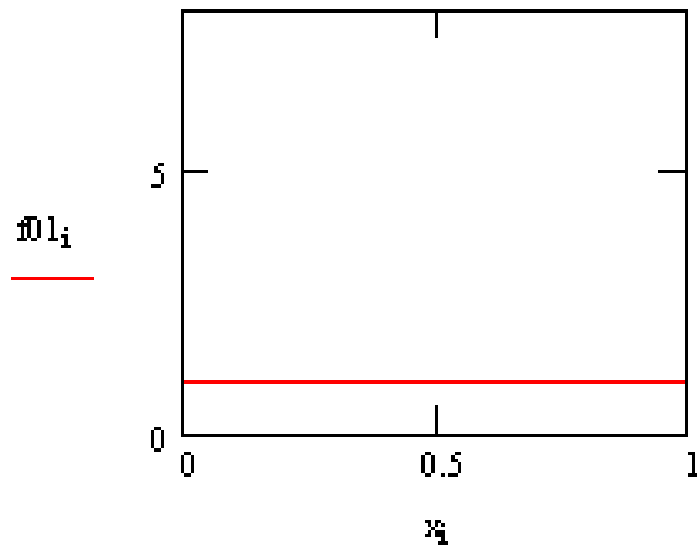


误差的产生环节：全过程质量控制



抽样误差

- 通过**设计**可以控制和降低抽样误差
- 充分利用**辅助信息**
- **样本量**及其分配
- **配额**控制
- 样本代表性（和总体结构做比较）
- 估计量的有效性

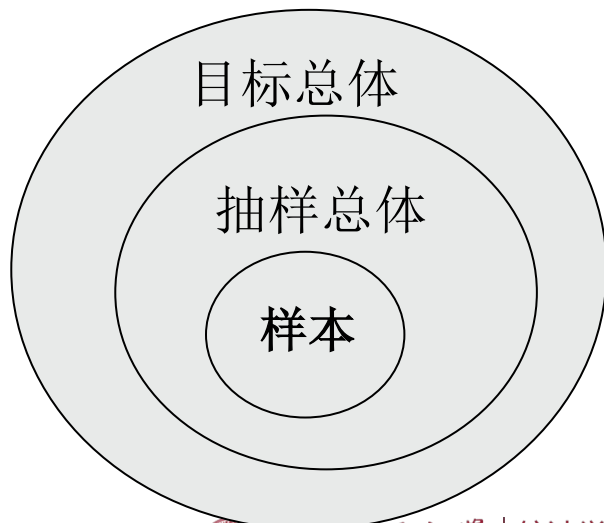


NonNormal Distribution of X



抽样框误差

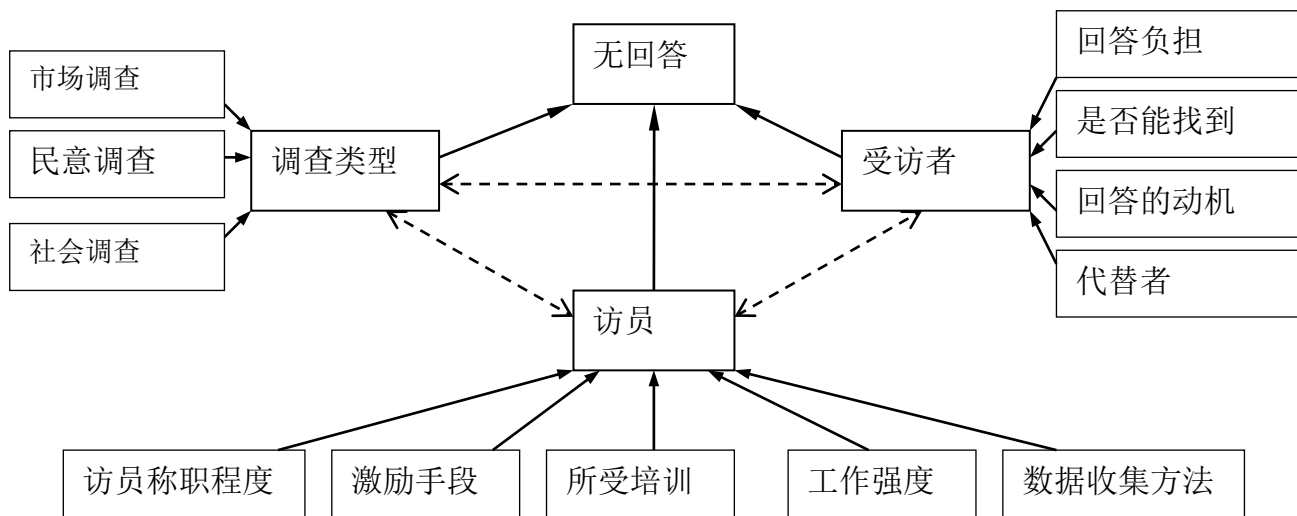
- 要随机抽样，一般需要抽样框
- 抽样框不完备会导致目标总体和抽样总体不一致
- 抽样框建设有困难
 - 可以考虑多阶段抽样名录库
 - 地图、行政区记录数据
 - 多个抽样框组合
 - 大数据辅助：样本抽选和估计



无回答误差

预防和处理

二重抽样
加权
插补
模型辅助



调查方式
问卷难度、长度
多次尝试
取得信任
激励

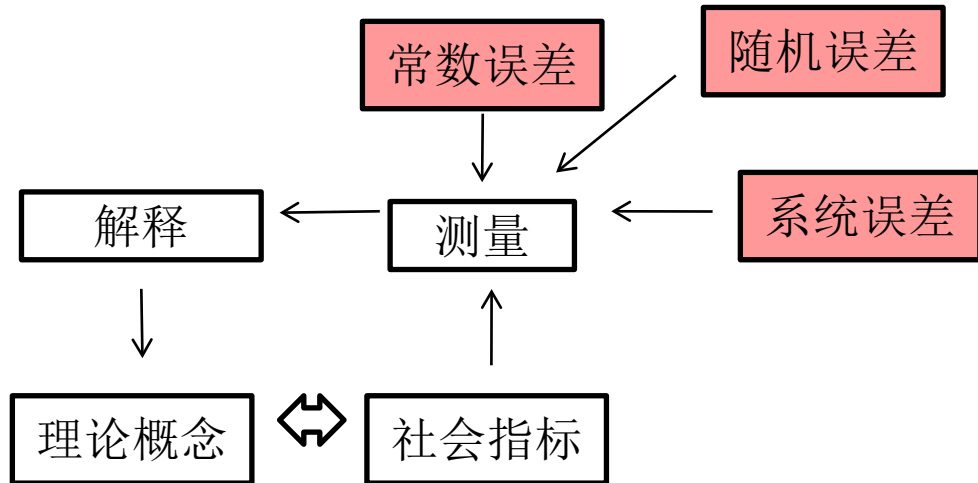
调查方式、访问员培训、工作负担、激励



测量误差

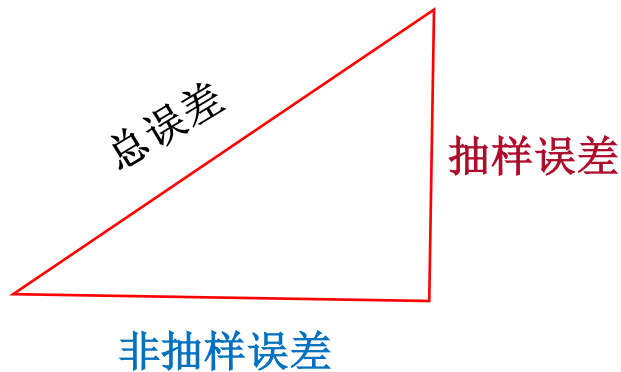
预防:

- 问卷的设计
 - 概念
 - 口径
 - 模型
 - 预调查
 - 信效度分析
- 访问员培训和奖惩制度
- 受访者
- 调查环境设计
- 质量监督方法和效果



调查中质量控制：心中有谱

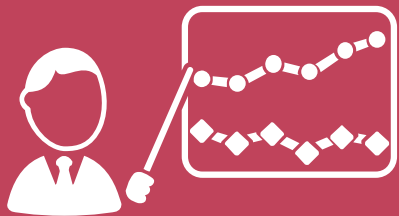
- 客观认知：不要盲目乐观和掩耳盗铃
- 尽量量化客观评价：
 - 抽样误差
 - 问卷信度效度
 - 模型的稳健性等
- 不可直接测度的误差，用辅助手段
 - 无回答误差要报告：回答率
 - 非概率抽样注意样本的代表性
 - 数据审核后的效果
- 控制工作的成效（**设计和执行**）
 - 例如预调查的调整、模型的改进



调查中质量控制：努力控制

- 一定经济水平下（学生小组团队，不同调查主题和设计的成本不同）
- 满足调查需求（区分描述性和验证性主题的调查）
- **预防为主**
 - **全过程**
 - **标准化**
 - 概念、问卷、抽样、培训、数据清洁、编码等
 - **方法的选择**
 - 抽样设计、现代化采集方式、多数据源、实验设计、量表改进等
- **事后处理为辅**
 - 加权、插补、估计量改进





PART FOUR

市调大赛作品的 质控和建议



质量控制不仅限于误差的控制 《国家统计质量保证框架（2021）
真实性、准确性、完整性、及时性以及适用性、经济性、可比性、协调性和可获得性等九个方面

http://www.stats.gov.cn/tjsj/zxfb/202106/t20210617_1818529.html

适用性要求统计数据能够最大限度为用户所用。侧重于对统计用户满意度的评价。

准确性要求统计数据的误差必须控制在允许范围内，能够为形势判断、政策制定、宏观调控等提供可靠依据。侧重于对统计数据生产科学性的评价。

经济性要求统计数据生产应当尽可能降低成本，统计调查、行政记录、大数据等数据资源得到充分利用。侧重于对统计数据成本效益的评价。

可比性：使用规范统一的统计标准和统计原则，侧重于对统计工作标准化、规范化程度的评价



大赛项目中质量控制建议

- **全流程**控制，注意系统化和结构化，不要流水账
- 满足调查需求，质量控制**不仅限于准确性**
 - 区分探索性、描述性和验证性主题的调查
- 不同调查主题和设计下主要误差来源不同，要有**针对性**
 - 例如敏感性的处理、提升测量的准确性、取得信任和激励、模型辅助
- **重点和特点**要突出
 - 网络调查属于非概率样本，注意样本代表性分析
 - 问卷设计注意：文献研究、预调查评价、信效度分析
 - 数据处理：具体方法和效果
 - 建模：变量的选择、理论模型的来源、模型的检验
- 可以用附件展示**标准化**文档和**组织执行力**
 - 问卷、抽样、培训、访谈、编码等



- 任何调查结果都有误差的，尊重客观事实
- 具体要达到什么质量，没有固定统一标准，满足调查需求
- 尽量定量测度和评价调查误差
- 重视误差的预防（全流程、突出重点）
- 重视标准化和技术方法
- 调查质量：可得的资源+方案设计+组织+执行





谢谢!



中國人民大學
RENMIN UNIVERSITY OF CHINA

统计学院
SCHOOL OF STATISTICS