目录

[1.大数据会对数据收集产生什么样的影响？ 1](#_Toc89620783)

[一、大数据的特征价值 1](#_Toc89620784)

[（一）大数据的的特征 1](#_Toc89620785)

[（二）大数据的类型 1](#_Toc89620786)

[（三）大数据挖掘价值方法 1](#_Toc89620787)

[二、大数据收集方式及其对传统数据收集的影响 1](#_Toc89620788)

[三、大数据时代下数据收集的问题 1](#_Toc89620789)

[（一）大数据存在覆盖漏洞，需用传统方法互补 1](#_Toc89620790)

[（二）法律法规不全面，导致暴力收集数据和隐私泄露 1](#_Toc89620791)

[2.设计一份统计调查方案应该考虑哪些内容？P25 2](#_Toc89620792)

[一、调查方案的概述 2](#_Toc89620793)

[（一）调查方案设计的概念P25 2](#_Toc89620794)

[（二）设计调查方案的基本原则 2](#_Toc89620795)

[（三）调查方案设计的作用P25/26 2](#_Toc89620796)

[二、调查方案设计的内容 2](#_Toc89620797)

[三、统计调查方案的评价 2](#_Toc89620798)

[（一）调查方案的可行性研究 2](#_Toc89620799)

[（二）调查方案的总体评价 3](#_Toc89620800)

[3.你平时收集资料(数据)的渠道有哪些？ 3](#_Toc89620801)

[一、原始资料、次级资料 3](#_Toc89620802)

[二、次级资料的方法及渠道 3](#_Toc89620803)

[（一）次级资料的方法 3](#_Toc89620804)

[（二）次级资料采集的实际渠道 3](#_Toc89620805)

[（三）次级资料的评估 3](#_Toc89620806)

[三、原始资料的收集 3](#_Toc89620807)

[（一）定性数据、定量数据 3](#_Toc89620808)

[（二）定性数据的收集 4](#_Toc89620809)

[（三）定量数据的收集 5](#_Toc89620810)

[4.调查方法与调查组织形式有区别吗？各有哪些？ 8](#_Toc89620811)

[课堂ppt 8](#_Toc89620812)

[一、调查方法和组织形式的概念 8](#_Toc89620813)

[二、调查方法分类 9](#_Toc89620814)

[（一）直接观察法 9](#_Toc89620815)

[（二）报告法 9](#_Toc89620816)

[（三）采访法 9](#_Toc89620817)

[（四）通讯法 9](#_Toc89620818)

[（五）问卷法 9](#_Toc89620819)

[（六）网络调查法 9](#_Toc89620820)

[（七）电话调查法 9](#_Toc89620821)

[（八）文献法 10](#_Toc89620822)

[三、调查组织形式 10](#_Toc89620823)

[（一）统计报表 10](#_Toc89620824)

[（二）普查 10](#_Toc89620825)

[（三）重点调查 11](#_Toc89620826)

[（四）典型调查 11](#_Toc89620827)

[（五）抽样调查 11](#_Toc89620828)

[5.大数据时代还需要抽样调查吗？ 12](#_Toc89620829)

[一、变革：从抽样到大数据 12](#_Toc89620830)

[（一）大数据时代的抽样面临一些实际困难 12](#_Toc89620831)

[（二）大数据对抽样的替代作用 12](#_Toc89620832)

[二、互补：大数据仍需抽样 12](#_Toc89620833)

[（一）大数据获取信息的局限性 12](#_Toc89620834)

[（二）抽样也要积极应对所面临的挑战 12](#_Toc89620835)

[三、选择：大数据抑或抽样 13](#_Toc89620836)

[（一）问题的类型 13](#_Toc89620837)

[（二）事件的类型 13](#_Toc89620838)

[（三）研究目的 13](#_Toc89620839)

[四、策略：大数据与抽样的融合 13](#_Toc89620840)

[6.你认为选择调查组织形式最重要的依据是什么？ 13](#_Toc89620841)

[？7.统计调查与实验设计有什么区别及联系？ 13](#_Toc89620842)

[一、联系 14](#_Toc89620843)

[二、含义、特点 14](#_Toc89620844)

[（一）统计调查 14](#_Toc89620845)

[（二）实验设计 14](#_Toc89620846)

[三、分类 15](#_Toc89620847)

[（一）统计调查 15](#_Toc89620848)

[（二）实验设计 15](#_Toc89620849)

[四、步骤 17](#_Toc89620850)

[（一）统计调查 17](#_Toc89620851)

[（二）实验设计 17](#_Toc89620852)

[五、研究目的 17](#_Toc89620853)

[（一）统计调查 17](#_Toc89620854)

[（二）实验设计 17](#_Toc89620855)

[？8.在进行抽样设计时，样本容量如何确定？ 17](#_Toc89620856)

[一、样本估计与推断方法P210 17](#_Toc89620857)

[二、影响样本容量的因素P211 17](#_Toc89620858)

[三、样本容量的确定P212 17](#_Toc89620859)

[1.成数估计法及变形 17](#_Toc89620860)

[2.经验主义方法 18](#_Toc89620861)

[3.实际的样本容量确定 18](#_Toc89620862)

[？9.你认为是调查的具体实施重要，还是方案更重要？ 18](#_Toc89620863)

[一、统计调查方案P25 19](#_Toc89620864)

[（一）调查方案设计的概念 19](#_Toc89620865)

[（二）设计调查方案的基本原则 19](#_Toc89620866)

[（三）调查方案设计的作用 19](#_Toc89620867)

[（四）调查方案重要性 19](#_Toc89620868)

[二、总结 19](#_Toc89620869)

[？三、调查误差 19](#_Toc89620870)

[（一）抽样误差 19](#_Toc89620871)

[（二）非抽样误差 19](#_Toc89620872)

[四、误差控制 20](#_Toc89620873)

[（一）抽样框误差 20](#_Toc89620874)

[（二）无回答问题及处理 20](#_Toc89620875)

[五、提高回答率： 20](#_Toc89620876)

[10.调查阶段要提高数据质量需注意哪些方面？ 20](#_Toc89620877)

[一、统计数据质量的重要性 20](#_Toc89620878)

[（一）统计数据的性质 20](#_Toc89620879)

[（二）数据审核、质量性评估重要性 21](#_Toc89620880)

[二、影响数据质量的因素 21](#_Toc89620881)

[（一）误差的类型 21](#_Toc89620882)

[（二）影响数据质量的因素 21](#_Toc89620883)

[三、现场调查的质量控制 21](#_Toc89620884)

[（一）调查准备阶段的质量控制 21](#_Toc89620885)

[？（二）调查实施阶段的质量控制 22](#_Toc89620886)

[（三）整理资料阶段的质量控制 22](#_Toc89620887)

[四、大数据背景下统计数据质量的影响因素与解决对策： 23](#_Toc89620888)

[（一）流程维度： 23](#_Toc89620889)

[（二）技术方面的解决对策： 23](#_Toc89620890)

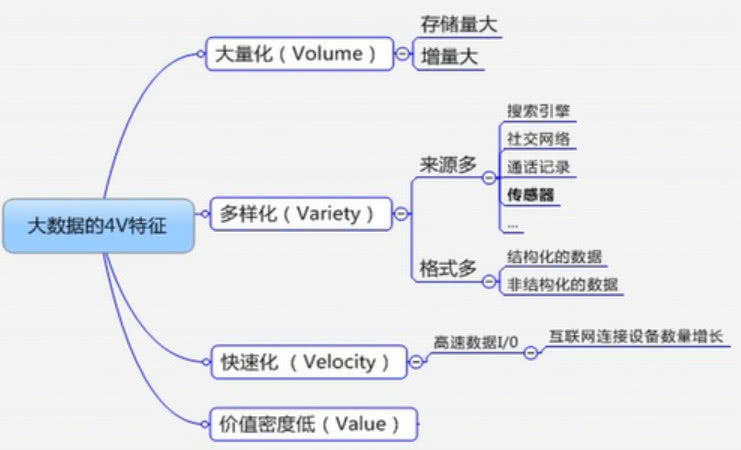
[（三）管理方面的解决对策 23](#_Toc89620891)

# 1.大数据会对数据收集产生什么样的影响？

## 一、大数据的特征价值

### （一）大数据的的特征

我们似乎可以从大数据的特征入手来进一步了解大数据。一般我们认为大数据有4V特征，即**Volume:体量大、Variety:样式多、Velocity:速度快、Valueless:价值密度低**。



### （二）大数据的类型

1.传统企业数据：包括 CRM 系统的消费者数据，传统的ERP数据，库存数据以及账目数据等。

2.机器和传感器数据：包括呼叫记录，智能仪表，工业设备传感器，设备日志，交易数据等。

3.社交数据：包括用户行为记录，反馈数据等。可以看出由于数据资源丰富程度的不同和从业人员的不，大数据的价值主要体现在第三产业的相关行业中。

### （三）大数据挖掘价值方法

1.客户群体细分，然后为每个群体量定制特别的服务。

2.模拟现实环境，发掘新的需求同时提高投资的回报率。

3.加强部门联系，提高整条管理链条和产业链条的效率。

4.降低服务成本，发现隐藏线索进行产品和服务的创新。下图是麦肯锡公司列出的各个行业利用大数据资源的难以程度以及发展潜力，其中横轴是发展潜力，纵轴是难易程度。

## 二、大数据收集方式及其对传统数据收集的影响

**1.收集到的数据类型更加丰富。**传统数据收集得到的一般是结构化数据。大数据收集的对象为一切可记录、可存储的信息，不仅包括结构化数据，还包括非结构化和半结构化数据等形式，这些多样化，难以通过传统统计指标或者图表表现，借助现代技术可以挖掘其中的价值。

**2.数据收集方式发生改变。**相比传统数据收集的抽样调查方法而言，大数据采集技术通过传感器数据、社交网络数据、移动互联网数据等方式获得各种类型的结构化、半结构化及非结构化的海量数据，大数据来源更加多样且易于获取，内容也更加完整丰富，由此可以得到与研究对象相关的海量数据，数据收集效率大幅提高。

**3.数据不确定性原因改变。**传统数据收集的不确定性来源于调查方案的设计、抽样的随机性、观测的偏差等，而大数据的收集数据几乎全覆盖，它的不确定性更多源自于收集数据时的数据形式多样性、内容真伪性、数据质量优劣性。如何有效的比较甄别以提取出真正需要的数据是数据收集的主要任务。

**4.更注重事物之间的相关关系。**在统计预测时，传统统计偏爱因果关系，大数据时代倾向更简单的相关关系。另外，由于大数据的数据收集技术的应用，对于捕获相关关系也更容易。

**5.统计数据的中心偏移。**传统统计实务的重心在于如何收集数据，大数据时代， 则更在于如何在海量数据中去伪存真、去芜存菁选取出我们想要的数据。

## 三、大数据时代下数据收集的问题

### （一）大数据存在覆盖漏洞，需用传统方法互补

大数据覆盖了数据的大部分区域，但由于信息的延展性，大数据集永远存在“数据黑暗地带”或“数据阴影区域”。再结合针对性、安全性和成本等比较问题，我们既要继续采用传统的方式方法去收集特定需要的数据，又要善于利用现代网络信息技术和各种数据源去收集一切相关的数据，并善于从大数据中进行再过滤、再选择，互相补充，以得到最为接近现实的数据。

### （二）法律法规不全面，导致暴力收集数据和隐私泄露

大数据时代多数数据产生于移动手机端，数据收集的低门槛大大降低数据收集的难度、成本和时间，加上法律缺少关于明确数据收集主体资格与责任、限定数据信息收集的对象与范围等内容，避免不良商家过度挖掘隐私，恶意掌握个人行为与习惯，进一步可能演化为网络诈骗等违法行为。

# 2.设计一份统计调查方案应该考虑哪些内容？P25

## 一、调查方案的概述

### （一）调查方案设计的概念P25

调查方案设计就是统计调查组织和实施者在进行实际调查之前，根据调查研究的目的和调查对象的性质，对调查工作总任务的各个方面和分阶段进行的整体设计，提出相应的调查实施方案，制定出合理的工作程序。

统计调查无论范围大小，都会涉及到相互联系的各个方面（指对市场调查的横向设计）和各个阶段（指的就是对市场调查纵向方面的设计）。

### （二）设计调查方案的基本原则

1.科学性原则：为了提高数据的质量，就必须采用科学合理的方法。在调查中必须根据调查要求选择一种或少数几种调查方法。

2.可行性原则：只有操作性强的调查方案才能真正成为调查工作的行动指导。

3.有效性原则：是指在一定的经费约束下，调查结果的精度可以满足调查目的的需要。

### （三）调查方案设计的作用P25/26

1.是统计调查从定性认识过渡到定量认识的开始阶段；

2.在统计调查过程中起着统筹兼顾、统一协调的作用；

3.在实践上适应现代统计调查发展的需要。

客观上不存在唯一最优的统计调查方案，每个方案都有其优缺点，最终的统计方案实质上是调查成本与调查数据质量要求之间的权衡。

## 二、调查方案设计的内容

调查方案是否科学、可行，会直接影响整个调查的成败。因此，在对调查方案设计时，调查组织和实施者一般都会按照调查工作的各个阶段进行整体考虑，并拟定出统计调查方案。

1.调查目的：明确调查目的是确定调查的范围、内容和方法的前提

2.调查对象和调查单位：调查对象是根据调查目的和任务所确定的调查范围及所要调查的总体，由某些性质上相同的若干调查单位组成。调查单位是调查对象中的每一个具体单位，它是调查中要调查登记的各个调查项目的直接承担者。需要注意：以科学的方法为指导确定调查对象的内在涵义；调查对象和调查单位之间的联系；调查单位和填报单位的区别；不同的调查方式会产生不同的调查单位。

3.调查内容：调查内容指调查组织和实施者对调查单位进行调查的主要内容。确定调查内容，就是要明确调查组织和实施者需要向被调查者了解哪些问题。需要注意:确定的调查内容应既是调查任务所需，又是便于取得调查答案的问题。调查内容的表达方式必须明确，备选答案应有明确的表示形式，如数字式、是否式或文字式等。确定调查内容时，应尽可能围绕某一需要调查的核心问题而设置一些相互关联的调查项目。调查内容的涵义明确、肯定，必要时可附加调查项目解释。

4.调查提纲和调查问卷：调查项目确定后，即可将调查项目进行分类、排列，在此基础上形成调查提纲或调查问卷，以方便调查登记和汇总。

5.调查时间和调查工作期限：调查时间指调查资料所要求的时间；调查期限指整个统计调查工作的开始时间和结束时间。

6.调查地点：在调查方案中，还要明确规定调查地点。调查地点与调查单位通常一致，当不一致时，有必要规定调查地点。

7.确定调查方式和方法：调查中所采用的调查方式和方法不是固定不变的，根据调查目的和其他一些约束条件的允许，调查方式和方法也可以进行灵活处理。

8.调查资料整理和分析的方法：调查资料的整理包括定性资料整理和定量资料整理两个方面。常用的分析方法有：频数分析、交互分析、回归分析、相关分析、聚类分析、因子分析等。每种分析方法都有其自身的特点、适用性和相关的技术要求，在分析时应根据调查的要求，选择最佳的分析方法，并在方案中进行规定。

9.调查报告内容和提交的方式：调查报告内容主要包括调查报告书的形式和份数，报告书的基本内容、报告书中图表量的多少等。调查报告提交的方式指统计调查组织和实施者以何种形式向用户提交调查报告，一般有电子资料方式、光盘资料方式、纸质资料方式等。

10.调查的组织计划：调查的组织计划指为确保实施调查而制定的具体工作计划。主要包括调查机构的设置、人员的选择和培训、工作步骤及其善后处理等。

11.调查经费预算：统计调查费用的多少通常应根据调查的目的、调查的范围和调查的难易程度而定。

## 三、统计调查方案的评价

### （一）调查方案的可行性研究

1.**逻辑分析法**：检查所设计的调查方案的内容是否符合逻辑和情理。

2.**经验判断法**：组织一些经验丰富的人士，对调查方案加以初步判断，说明方案的可行性。

3.**试点调查法**：对于大规模调查尤为必要，帮助发现问题，完善方案。

试点调查的**任务:**（1）对调查方案进行实地检验；（2）作为实战前的演习。试点调查的**注意事项**:（1）建立团队；（2）选择调查对象；（3）灵活调查；（4）做好总结工作。

### （二）调查方案的总体评价

调查方案一般都应满足以下三个要求：方案设计是否体现调查目的和要求；方案设计是否科学、可行和有效；方案设计能否使调查质量有所提高。

# 3.你平时收集资料(数据)的渠道有哪些？

## 一、原始资料、次级资料

原始资料:是调研人员为满足特定目标而收集的第一手资料。

次级资料:也叫二手资料、已有资料，是其他人或机构为了其他目的而收集、记录和整理出来的有关资料

## 二、次级资料的方法及渠道

### （一）次级资料的方法

1.参考文献查找法：利用有关著作、论文所列的参考文献目录，以此为线索追踪、查找有关的文献资料。

2.检索工具查找法：手工检索:目录、索引、文摘；计算机检索:各类搜索引擎、数据库

3.报刊剪辑分析法：是指调查人员平时从各种报刊上所刊登的与企业经营和市场有关的文章、报道中，分析和收集情报信息的一种方法。

4.情报联网法：是指相关机构在全国范围内或国外有限地区内设立情报联络网，使情报资料收集工作的触角伸到四面八方的一种方法。

5.广告收集法：指通过收集其他厂商免费赠送产品目录、说明书等资料获得“文案”资料的方法，企业还可以根据自己需要的内容通过书信索取、询问、现场搜集、接受赠阅等获得这些资料。

6.索取和购买:参加一些研讨会、展示会议等索取免费资料；向一些中介机构或专业信息公司购买。

### （二）次级资料采集的实际渠道

1.官方机构：各地统计局--统计数据、分析报告2.学术单位：学校图书馆、地方图书馆--统计出版物、期刊、数据库3.行业资料：行业报刊、杂志等、行业专业人士访谈或研讨会、行业专项调查、现成的研究报告、其他市场研究公司资料服务研究等4.网上资料：搜索引擎、网络数据库（资讯行）、专业论坛（人大经济论坛）、开放式文库（百度文库、豆丁文库、新浪爱问）等

### （三）次级资料的评估

1.质量评估的问题

是谁收集的资料?（Who）资料的收集目的是什么?（Why）收集的是什么资料?（What）资料是什么时候收集的?（When）资料是如何收集的?（How）所得资料与其他资料是否相一致?（Coherence）

2.次级资料的具体评价

相关性评价：即所收集的资料与调查项目主题关系是否密切;资料的专门程度够不够格;准确性评价：即所收集的资料是否有误；资料所涉及的时期是否适当，有没有事过境迁;与第一手资料的接近程度如何; 系统性评价：即所收集的资料是否全面，资料是否针对与研究主题最有关的各个方面。

3.次级资料需满足的要求：准确性、相关性、时效性、完整性、适用性、一致性。

## 三、原始资料的收集

### （一）定性数据、定量数据

**定性调查**主要应用于探索性调查，调查的目的是考察被调查者的态度、感觉、动机、反应等。在定性调查中，调查员根据讨论提纲或讨论指南进行调查，所收集的数据大多是难以量化的定性数据。

**1.定性调查的优点**：（1）定性调查一般只对小样本进行调查，这在很大程度上降低了调查的成本。（2）能更好地了解被调查者内心深处的态度、动机。（3）可以提高定量调查的效率。

**2.定性调查的局限性**：（1）不能反映细微的差别。（2）样本的代表性较低。定性调查的样本量一般较小，对总体的代表性较差。（3）对调查员的要求较高。调查员须具备心理学、社会学等多方面的知识储备才能胜任定性调查。

**3.定性调查的应用：**定性调查主要应用于探索性调查，其应用领域十分广泛，几乎可以应用于所有的行业和市场。近年来，定性调查越来越受到人们的关注，并且出现了专门从事定性调查的调查公司。

**定量调查**主要应用于描述性、因果性和预测性调查，调查的目的是对总体的数量特征进行推断。定量调查中，调查员一般根据调查问卷进行调查，所获得的数据大多是定量数据，因此在数据分析中主要使用频数分析、相关分析、因子分析、聚类分析等统计分析方法。

定性调查与定量调查的区别



### （二）定性数据的收集

**（1）观察法**

**1.观察法概述**：观察法是指研究者根据一定的研究目的、研究提纲或观察表，用自己的感官和辅助工具去直接观察被研究对象，从而获得资料的一种方法。

使用观察法收集信息需具备的条件：（1）所需信息必须是能观察到的，或者能从观察到的行为中推断出来（2）所要观察的行为必须具有重复性或者在某些方面具有可预测性（3）所要观察的行为必须在相对短的时间内完成

**2.观察法的类型**

控制观察和无控制观察：控制观察是根据观察目的事先规定观察范围，并在实施观察时将观察手段、观察技术、观察程度和记录方式标准化，一般用于目的性和系统性较强的观察；无控制观察是对观察项目、程度等不做严格规定，记录也可采取随意的方式。通常用于探索性或有一定深度的专题调查。

完全参与观察、不完全参与观察和非参与观察：完全参与观察：观察者较长时期地生活在被观察者之中，甚至改变自己原有的身份。不完全参与观察：观察者不改变自己原有的身份，只在调查时置身于调查事项中取得资料。非参与观察：观察者不参与调查活动之中，而以局外人的身份去观察事项发生和发展的情况，客观但无法深入

纵向观察、横向观察和纵横结合观察：纵向观察：在不同时期加以观察，取得一边串的记录，能了解调查对象发展变化的过程和规律。横向观察：在某一特定时间内对若干个调查对象所发生的事态同时加以记录，有利于不同对象间的对比分析。纵横结合观察：对不同时期的多个调查对象同时进行观察、记录。

人员观察、机器观察:人员观察法就是由调查人员实地观察受访对象以了解情况的常用形式，包括神秘购物法、单向镜法、顾客观察法、人文调查、实际痕迹观察；机器观察法是通过机器设备观察受访对象。

**3.观察法的优缺点**

优点：（1）直接性和可靠性（2）不依赖于语言交流，减少了误会和干扰。（3）观察法简便易行，灵活性强，可随时随地进行调查；

缺点：（1）只能反映客观事实的发生经过，不能说明发生的原因和动机，只看表象致使观察深度不够。（2）需要大量观察员到现场作长时间观察，调查时间较长，调查费用支出较大。因此，常会受到时间、空间和经费的限制，比较适用于小范围的市场调查。（3）对调查员的技术水平要求较高，人员素质引发观察结论误差。

**（2）小组座谈法**

**1.小组座谈法概述**

小组座谈法也叫焦点小组访谈法或小组讨论法，是挑选一组具有代表性的被调查者，在一个装有单向镜或录音、录像设备的房间中，采用小型会议的形式，由主持人引导对研究主题进行讨论，从而获得信息的一种调查方法。

**2.类型**

电话小组访谈法:被调查者会在一个特定的时间进行电话讨论。优点是被调查者不用到测试室就进行调查，时间安排比较灵活，从而使一些在传统焦点小组访谈中很难征集到的被调查者能够参加访谈。

电视会议小组访谈法:是在配有摄像机和麦克风的会议室里进行的。优点是观察访谈的人员不用在现场就可以进行观察，可以允许大量的观察人员对访谈过程进行观察。

在线小组访谈法:以互联网为媒介展开的焦点小组讨论。优点是减少了调查的成本，便于甄选到合适的被调查者，委托方和主持人可以实现双向的充分沟通。不足是缺乏群体动力、无法获得非言语信息、被调查者的可靠性较低、注意力容易受到干扰等。

**3.优缺点**

优点：（1）资料收集速度快，效率高;（2）取得的资料较为广泛和深入;（3）能将调查与讨论相结合;可以进行科学监测。

缺点: （1）对主持人的要求较高;（2）容易产生偏差;（3）不宜对敏感性问题进行调查;（4）不利于资料的整理和分析。

**（3）深层访谈法**

**1.深层访谈法的含义和特点**

深层访谈指调查员和一名被调查者在轻松自然的气氛中围绕某一问题进行深入的讨论，目的是让被调查者自由发言，充分表达自己的观点。深层访谈法与小组座谈法的区别在于，深层访谈是一对一进行的，而小组座谈是一对多进行的。深层访谈是一种无结构的、直接的、一对一的访问，虽然调查员在访问前会准备一个粗略的调查提纲，并试图按提纲来访问，但提问时的具体措词和顺序可以根据访谈进行的具体情况进行调整。

**2.优缺点**

优点: （1）消除了群体压力，被调查者能够提供更诚实的信息;（2）一对一的交流使被调查者感到自己是被关注的焦点，因此乐于表述自己的观点、态度和内心想法;（3）在被调查者访谈的时间充分，可以鼓励他们吐露更多的信息，这样做常常能够有意外的收获。

缺点: （1）失去了群体动力，无法产生被调查者之间观点的相互刺激和碰撞;（2）成本较高，尤其是当受访人数较多时;（3）访谈员一天内访谈的人数有限。

**（4）德尔菲法**

1.德尔菲法的原理与特色

原理：德尔菲法在我国也被称为专家预测法或专家调查法，是专家们采用书面的形式，背靠背地回答调查人员提出的问题，并经过信息反馈，多次修改各自的意见，最后由调查人员进行综合分析，得出调查结论的方法。

德尔菲法的特征：匿名性、轮回反馈性、定量分析、调查人员作用突出

2.优缺点

优点: （1）节省费用（2）有利于专家们的独立思考、独立判断（3）提高了调查的科学性

缺点: （1）缺乏客观标准，有一定的主观片面性（2）调查结论有可能接近于中位数或算术平均数（3）调查时间比较长

**（5）投影技法**

1.投影技法概述

投影技法是一种无结构、非直接的调查方法，运用投影技法，调查人员可以挖掘出被调查者潜在动机、态度和情感。投影技法能够穿透人们的心理防御机制，尤其适合于敏感性问题的调查。

2.类型

联想技法：联想技法是在被调查者面前设置某一刺激物，然后询问被调查者最初联想到的事物。

（1）自由联想法：不限制联想性质和范围，回答者可充分发挥想像力。（2）控制联想法:将回答的联想限制在一定的范围内。（3）引导联想法:在提出刺激词语的同时，也提出相关联想的词语。例如，请您就所给的词语按提示写出（或说出）所引发的相关联想。

完成技法:完成技法中，给出一种不完整的刺激情景，要求被调查者完成这一情景。

（1）句子完成法:提出一些不完整的句子，让被调查者完成该句子。（2）故事完成法:在故事完成法中，给被调查者故事的一个部分，被调查者要用自己的话来做出结论。

结构技法:要求被调查者以故事、对话或绘图的形式构造一种情景。

（1）图画回答法:做法是显示一系列的图画，要求被调查者看图讲故事。（2）卡通试验法:在卡通试验中，将卡通人物显示在一个与问题有关的具体环境内。

表现技法:是给被调查者提供一种文字的或形象化的情景，请他将其他人的感情和态度与该情景联系起来。

(1角色表演法：请被调查者以他人的角色来处理某件事，以间接反映其真实动机和态度。（2）第三者技法：在第三者技法中，给被调查者提供一种文字的或形象化的情景，让被调查者将第三者的信仰和态度与该情景联系起来。

3.优缺点

优点:（1）投影技法可以绕过人们的心理防御机制来获取真实的信息.（2）投影技法尤其适用于涉及隐私或敏感性问题.

缺点:（1）需要专门的、训练有素的调查员。（2）调查成本比较高。投影技法对调查员的素质要求比较高，有时候还需要心理学专家参与答案的分析工作，因此增加了调查成本。（3）解释偏差比较大。投影技法对答案的解释比较困难，具有比较强的主观性，因此可能会存在比较大的解释偏差。

**（6）网络定性调查方法**

一对一的网上深层访谈：类似于传统的深层访谈，只是借助于网络工具微信、QQ、电子邮件等

网上小组座谈会：一种为在线交流工具的群体讨论，如微信、QQ的群聊；另一种是在线社区的实时讨论。

网上观察:重点观察被访问者的语言行为，通过观察各种类型的新闻小组、同步聊天的对话，了解被访者的一些想法。

网上文献资料分析：非请求式和请求式，前者是自己利用各种渠道主动收集相应的调查资料，后者是向特定对象请求提供相应的资料。

### （三）定量数据的收集

**（1）电话调查**

1.类型

传统电话调查：经过培训的调查员使用电话，按照调查设计的要求拨打电话，当电话接通时按照准备好的问卷和培训要求，筛选被调查者，然后对照问卷向合格被调查者逐题提问，并及时迅速的将回答的答案记录下来。

样本选取方法:（1）利用现有的电话号码簿作为抽样框，借助随机数字表，随机拨打号码；或按照等距抽样方法，从电话号码簿中抽取号码拨打。（2）按照调查地区的具体情况和抽样方案先确定拨打号码的前几位（通过前几位号码，如区号，可以确定调查的地区），然后按照随机原则确定后几位。

计算机辅助电话调查：在计算机辅助调查中，每一位访谈员都坐在一台计算机终端或个人电脑面前，当被调查者电话被接通以后，问题和选项立即出现在屏幕上。访谈员根据屏幕提示进行提问，并将被调查者的选择直接录入计算机中。在访谈过程中，计算机可以随时提供整个调查的进展情况，统计分析也可以在调查进行的任何阶段进行。研究人员应根据阶段性调查结果及时调整方案，使调查更为有效。

2.优缺点

优点：费用低；样本覆盖面广；时效性好；能够得到更为坦诚的回答；易于控制调查实施的质量

缺点：抽样总体与目标总体不一致；被调查者看不到演示物，应用领域受到一定限制；调查内容难以深入；访问的成功率会受到影响。

3.应用范围

在我国，目前电话调查的应用范围还比较有限，但发展迅速：热点问题或突发性问题的快速调查；特定问题的消费者调查，如新产品的购买意向，新推出广告的到达率，新开播栏目的收视率等；企业调查，如企业管理者对某些问题的看法，对某些产品的评价及购买意向等；特殊群体调查，如新闻记者对塑造企业形象的看法,企业对会员群体的调查等；都市地区某些产品或服务的市场调查，如大城市居民的教育消费意愿等

**（2）面访调查**

1.类型

入户面访调查：是指被调查者在家（或单位）中单独接受访问。入户面访调查时，访谈员按照抽样方案中的要求，到被调查者家中或单位中，依据问卷或调查提纲进行面对面的直接访问。

入户面访调查是一种应用较为普遍的资料收集方式。由于入户面访调查采取私下的、面对面的访谈形式，同时又有访谈员的协助，因此可以对复杂的内容进行调查。从被调查者角度来看，由于被调查者身处自己熟悉的环境，可以放松地参与访谈，从而有可能得到质量较高的调查结果。

优缺点 ：优点：信息获取的直接性；调查组织的灵活性；调查过程的可控制性；调查数据的准确性。缺点：时间限制；调查者的影响；拒访率较高

街头拦截式面访调查：一般是在超市、商贸中心等繁华地段展开。在调查过程中访谈员按照规定的程序和要求选取被调查者，征得其同意以后，在现场或者在附近的访谈室展开访谈。

优缺点：优点：节省费用；避免入户困难；便于对访问员的监控。缺点：不适合内容较多；调查的精确度较低；拒访率较高

中心地调查或厅堂测试：事先选定若干场所，租借好专用的房间或厅堂，然后对拦截对象进行面访调查。

街头/商城拦截式面访调查的调查费用要低于入户面访调查，然而街头拦截式在样本代表性、回答率上比不上入户面访调查，并且由于街头的访谈环境不如入户访谈的环境舒适，被调查者可能会表现出不安、急躁情绪，因此调查的时间不能太长。

计算机辅助面访调查：可以应用于入户面访调查和街头拦截式面访调查。在计算机辅助面访调查中，经过培训的访谈员配备笔记本电脑向被调查者进行面访调查。

2.面访调查的优缺点

优点：可以获得较有代表性的样本；回答率较高；可以对复杂问题进行调查

缺点：费用较高；时效性较差；调查质量的控制难度大；对某些群体的访问成功率较低

3.面访调查的应用范围：消费者研究：消费者行为研究、消费者生活形态研究、消费者满意度研究等；媒体研究：媒体接触行为研究、广告效果研究等；产品研究：产品使用情况和态度研究、新产品的开发研究等；市场容量研究：市场占有率研究、市场总体需求规模研究、未来需求量研究等；其它方面的研究

**（3）自填问卷式调查**

1.类型

普通邮寄调查：将问卷通过邮局寄给选定的被调查者，并要求他们按规定的要求和时间填写问卷，然后寄回调查机构。

优缺点：优点：费用较低；调查者的影响较小；调查区域广泛；应答更确切。缺点：问卷回收率低；缺乏对被调查者的控制；应答者会有选择偏见；时效性较差

留置问卷调查：调查员按面访的方式找到被调查者，说明调查目的和填写要求后，将问卷留置在被调查者处，约定一段时间后再来取回填好的问卷。其是介于邮寄调查和面访调查之间的一种方法，既综合了邮寄调查保密性的优点，又结合了面访调查回收率高的优点。关键之一是保证匿名性。

固定样本邮寄调查：事先抽取一个样本，在征得所选样本同意后由调查机构向这个样本中的成员定期地发送调查问卷，被调查者按问卷要求填写问卷并寄回调查机构。

2.自填问卷式调查优缺点

优点：保密性强；调查区域广；费用较低；无访谈员误差；调查内容可较复杂

缺点：问卷回收率低；时效性差；问卷的质量难以控制；对被调查者的文化水平有一定的要求

**（4）网络调查**

1.概念、特点

概念：网络调查是指以互联网为媒介进行的资料收集活动。特点：（1）网络信息的及时性和共享性（2）网络调查的便捷性和低费用（3）网络调查的交互性能和充分性（4）网络调查结果的可靠性和客观性（5）网络调研无时空和地域的限制（6）网络调研可检验性和可控制性（7）互联网调查的局限性

2.类型

网站/网页调查：是将设计好的问卷放在网站的某个网页上，问卷一般都设计得比较吸引人，而且易于回答。特点：调查结果自动进入数据库，便于快速处理。分类：纯文本格式调查、固定格式互动调查和订制调查

电子邮件调查：是将问卷直接发送到被访者的私人邮箱中，引起被访者的注意和兴趣，主动地填答并发送回问卷。抽样框需目标群体的电子邮箱地址。类型：电子邮件文本调查（数据需要人工输入）、电子邮件软件调查（数据自动采集）和可执行文件调查。特点：类似于传统的邮寄问卷调查，不过速度更快；一般限于平面文本格式，无法实现跳答、随机化、自动查错等复杂的问卷设计；调查质量很大程度上取决于抽样框的完备性和回收率的高低。

弹出式调查：当网民在访问某个网站的过程中，遇到弹出来的窗口，邀请参与一项调查；如果网民有兴趣，点击“是”，则会出现有一份问卷的新窗口，完成网上问卷调查后即可线上提交。特点：网站一般安装有抽取被访者的软件，按照等距、随机或其它方法自动抽取被访者；得到的一般并不是真正意义上的随机样本；保证一个访问者只能填答一次，还需采取跟踪文件。范围：比较适合于网站使用情况调查、对重大时事看法的调查等

网上固定样本调查：是一种将互联网技术与传统调查相结合的方法，通过随机抽样调查或其他有效的方法，征募一个有代表性的固定样本（最好是网民），对这个样本进行固定的网络调查。特点：如果固定样本的抽样和征集保证了质量，这种调查具有较好的代表性；快速、可靠；初期投入费用较高。

新闻组讨论调查：利用新闻组与客户或潜在客户进行交流和讨论以获得信息的调查方式。

3.网络调查的优缺点

优点：调查成本比较低；时效性好；能够高效率地找到具体的调查目标；交互性好，能够实现多样化的问卷设计；较高的质量控制

缺点：样本的代表性问题；无限制样本问题；网络的安全性；问卷长度受到限制

4.网络调查的应用范围

特定人群的消费者研究：文化程度较高、收入水平较高的中青年群体；特定产品或服务的消费研究：电子产品、电信、金融、电子商务等；一些容易借助多媒体技术的特定类型市场研究：广告片测试、产品包装测试、网络游戏测试等；网络行为和态度研究：网络购物、网络广告、网络游戏等；社会热点或突发问题的快速调查：奶粉事件、金融风暴等；内部的员工满意度调查等

**（5）实验法**

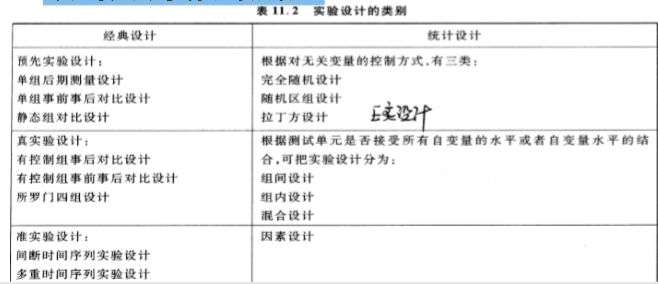
1.实验法概述：实验法是在既定条件下，通过实验对比，对市场现象中变量之间的因果关系及其发展变化过程进行观察分析的一种调查方法，通常又称为因果关系调查。

2.实验法的分类：按场所分：实验室实验、现场实验；按是否随机分组：非随机化实验和随机化实验

3.实验法的应用范围：主要用于对变量间因果关系假设的检验

4.实验法的质量控制：效度，指实验设计能够回答要研究的问题的程度。内在效度：自变量和因变量之间存在关系的明确与真实程度。即因变量的变化在多大的程度上是由自变量引起的，是对自变量进行实验操作的结果，而不是由于其他因素的影响。外在效度是指在与实验条件相类似的情况下实验结果可以推广的范围。一般地说，一项实验的内在效度越充分，实验结果推广的范围越大，实验的价值就越高。内在效度和外在效度并不总是一致的，内在效度和外在效度之间是会有冲突的，要保证一种效度，可能会削弱另一种效度。

5.实验设计总体可分为两大类：经典设计与统计设计。两者的基本区别在于经典设计一次只考虑一个独立变量的一个处理水平的作用，而统计设计可检测一个独立变量的多个处理水平的作用，或者两个及多个独立变量的作用。



6.几种常用的实验方法

前后无控制对比实验：即指事前对正常情况进行测量记录，然后再测量记录实验后的情况。进行事前事后对比，通过对比观察了解实验变化的效果。

前后有控制对比实验：在同一时间周期内，随机抽取两组条件相似的单位，一组作实验组，另一组作控制组（即非实验组，与实验组作对照比较的），在实验的后分别对二组进行测定比较。

控制组、实验组对比实验：即同一时间内对控制组与实验组进行对比的实验调查法。其中，实验组核按给定实验条件进行实验，控制组按一般情况组织经济活动。

完全随机对比实验：譬如：随机地选取试验几个商店做销售实验.

分组随机对比实验：研究者除了考察基本自变量因素的影响外，还可将某个主要的外部因素控制起来研究。

7.实验法的优缺点

优点：可以检验变量之间的因果关系；这种优点是其他调查方法如面访调查、焦点小组访谈、深层访谈、电话调查等所不能够提供的；便于重复进行，实验结果的可靠性较高；提高决策的科学性；控制调查环境和调查过程；提高调查的精确度

缺点：成本高昂；保密性差；实施、管理困难；易发生不可控的因素和现象；调查的时间较长；调查风险大。费用较高；操作难度大

# 4.调查方法与调查组织形式有区别吗？各有哪些？

**调查方法：**定性调查和定量调查

**组织方式：**

统计报表制度（统计报表制度是以原始记录为依据，按照国家统一规定的表格形式，统一的指标和内容，统一的报送时间和程序，自下而上定期向国家报告基本统计资料的一种调查方式）

专门调查（专门调查是为了研究某些专门问题而对某些社会经济现象专门组织的登记和调查。它是我国统计工作中重要的统计调查的组织形式。专门调查主要包括：普查、重点调查、抽样调查和典型调查四种）

**两者区别**：

调查组织形式首先是抽象的，相当于一个理论。调查方法是具体的，相当于一个实际活动方法。

样本单位确定之后，对这些单位实施调查，即从样本单位那里得到所需要的数据，可以采用不同的方法。这里的方法就是统计调查方法。统计调查组织方式是指组织统计调查、采集信息资源的方式、方法。社会经济统计调查组织方式有统计报表、普查、重点调查、典型调查和抽样调查等种类。统计调查组织形式决定于调查对象的特点和调查的目的和要求。不同情况应该采用不同的组织方式，以达到统计调查准确、及时、完整、系统的目的。

就比如我们调查某一类情况，我们要先选择一个匹配的调查组织形式 ，从而在一个总体中选择出了一个样本，那么对于这个样本中的每个个体，我们要选择匹配的调查方法，从而在这个样本中得到想要的数据。

## 课堂ppt

## 一、调查方法和组织形式的概念

统计调查方法：是指统计机构和统计人员搜集统计资料的方法。具体的调查方法有直接观察法、报告法、采访法、通讯法、问卷法、网上调查法、电话调查法和文献法等

统计调查组织形式：是指组织统计调查，搜集统计资料的方式方法。社会经济统计调查组织形式主要有统计报表和专门调查，其中专门调查又可以分为普查、重点调查、典型调查、抽样调查。

## 二、调查方法分类

### （一）直接观察法

1.定义：直接观察法是指调查人员亲临现场对调查单位的调查项目进行清点、测定、计量，并加以登记，以取得第一手资料的一种方法。

2.优缺点

能够保证所搜集的调查资料的准确性，它不受被观察者的意愿和回答能力影响；需要花费大量的人力、物力、财力和时间；只能搜集到现场资料，无法搜集到历史资料；由于存在调查者与被调查者的主客观因素问题，使调查资料的客观性受到影响。

3.类型：公开观察 是调查人员在调查地点的公开，即被调查者意识到有人在观察自己的言行;隐蔽观察 是指被调查者没有意识到自己的行为已被观察和记录。

### （二）报告法

1.定义：报告法 是指调查单位或报告单位利用各种原始记录、基层统计台账和有关核算资料作为报告依据，按照隶属关系，逐级向有关部门提供统计资料的一种方法。报告法可以促进被调查单位建立健全原始记录和核算资料。

2.特点:统一性和时效性：由于报告法的表格形式、口径、范围、要求及报送程序等都是统一规定的，各报告单位只是按规定执行，所以保证了资料的统一性和时效性；周期性：采用报告法搜集资料，往往是不间断地按相等的时间间隔定期进行，资料具有动态衔接性和可比性；相对可靠性：报告法建立在基层单位的原始记录和核算资料的基础上，故资料具有相对可靠性；灵活性差：自下而上的报告制度需要严密的组织工作，使实际操作中的难度增大，降低了灵活性。

### （三）采访法

1.定义：采访法是由调查人员根据调査提纲向被调查者提出所要了解的问题，然后根据被调查者的答复来取得统计资料的调查方法。

2.特点：由于询问者与被询问者直接接触，逐项研究问题，因而搜集的资料比较准确；调查所需要的人力、费用较多；对调查人员的要求较高，如知识面广，公关能力强，态度及心理素质较好等。

3.分类：个别询问法：是由调查人员对每一个被调查者逐一提出所要调查的问题，以取得统计资料的方法；开调查会法：是召集了解情况的有关人员，以座谈会形式对被调查的问题开展讨论和分析，以取得统计资料的方法。

### （四）通讯法

1.定义：通讯法又称邮寄法，是指将调查表或调查问卷邮寄给被调查者，由被调查者按规定要求填好寄回，以取得资料的一种调查方法。

2.优缺点

优点：可以扩大调查的地域和范围，所需经费相对较少；被调查者有充足的时间思考和回答问题

缺点：被调查者的数量不宜太多，调查项目不宜过于复杂；调查表的回收、调查内容的理解和回答的准确性、可靠性等难以得到有效保证

### （五）问卷法

1.定义：问卷法是调査者运用统一设计好的调查表，向被调查者了解情况、搜集资料的种调查方法。问卷调查多用于非全面调查。

2.优缺点

优点：通俗易懂，实施方便，适用于各种范围与环境；易于对资料进行处理和定量分析，节约时间和人力、财力，能提高调查效率

缺点：调查问卷的设计需要大量经验；问卷的回收率和有效率难以得到保证。

3.分类：自填式问卷调查，按照问卷传递方式的不同，可分为报刊问卷调查、邮政问卷调查和送发问卷调查；代填式问卷调查，按照与被调查者交谈方式的不同，可分为访问问卷调查和电话问卷调查。

### （六）网络调查法

1.定义：网络调查法是通过互联网发布调查问卷来搜集有关统计资料的一种方法。

2.优缺点

优点：组织简单，速度快，效率高，成本低；市场调查对象广泛；客观性强；可视性强，利用多媒体技术，可以使调查更加生动形象、直观。

缺点：受访对象难以限制，针对性不强，样本代表性仍然不够；回答率低；资料安全性较低。

### （七）电话调查法

1.定义：电话调査法是指调査人员通过电话对被调查者进行语言访问来搜集信息的调查方法。由于彼此不直接接触，而是借助于电话这一中介工具进行，因而是一种间接的调查方法。

2.分类：电话调查分为传统电话调查和计算机辅助电话调查。采取计算机辅助电话系统，更有利于访问质量的监控，访问人员的管理更为系统规范。

3.优缺点

优点：电话调查速度快，范围广，费用低；在电话中回答问题一般较坦率，适用于不习惯或者不能面谈的人

缺点：受电话设备的限制，电话调查时间短、答案简单，调查内容的深度远不及其他调查方法；很难判断所获信息的准确性和有效性等

### （八）文献法

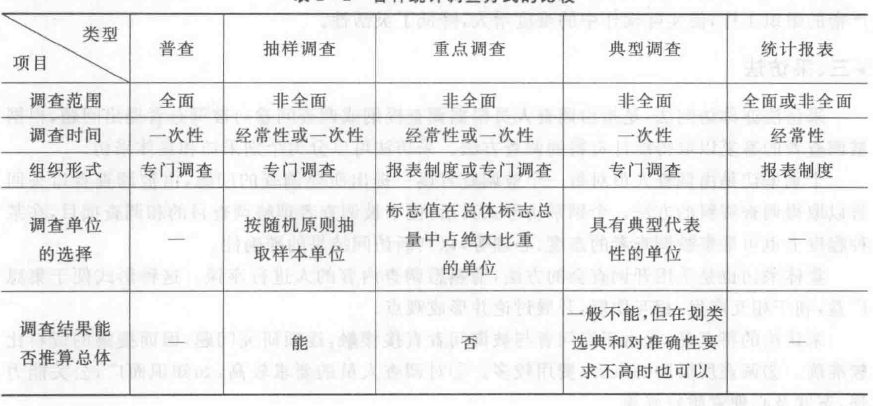
1.定义：文献法也称历史文献法，是根据调査目的，浏览著作、报告、论文、统计或业务报表等，获得所需要的研究信息。

2.优缺点

优点:文献调查省时、省钱、效率高；文献法比较方便、自由；文献调查超越了时间、空间限制，调查范围广泛

缺点：许多文献的质量往往难以保证，常常隐含着作者的主观意图和偏见；许多文献都不是公开的和可随意获得的；文献的整理和编码困难。

## 三、调查组织形式



### （一）统计报表

1.概念：统计报表是我国搜集统计资料的一种重要的方式，它是按照国家相关法律的规定，自上而下统一布置，自下而上地逐级定期提供基本统计资料的一种调查方式。

2.优点：资料来源可靠；便于各级领导了解情况；易于资料的纵向比较 缺点？

3.种类：按照调查范围不同，统计报表可以分为全面通缉报表和非全面统计报表；按报送时间不同，统计报表可以分为日报、旬报、月报、季报、半年报和年报统计报表；按报送范围不同，统计报表可以分为国家报表、部门报表、地方报表；按填报单位不同，统计报表可以分为基层报表和综合报表

### （二）普查

1.概念：是对统计总体的全部单位进行调查以搜集统计资料的工作，是为特定目的而专门组织的一次性全面调查。普查资料常被用来说明现象在一定时点上的全面情况。

2.遵循原则：必须统一规定调查资料所属的标准时点；正确确定调查期限、选择登记时间；规定统一的调查项目和计量单位；普查尽可能按一定周期进行，以便于研究现象的发展趋势及其规律性

3.普查方式：建立专门的普查机构，配备大量的普查人员，对调查单位进行直接的登记，如人口普查等；利用调查单位的原始记录和核算资料，颁发调查表，由登记单位填报，如物资库存普查等

4.特点：（1）普查通常是一次性的或周期性的：由于普查涉及面广、调查单位多，需要耗费大量的人力、物力和财力，通常需要间隔较长的时间，一般每隔10年进行一次；（2）规定统一的标准时点：标准时点是指对被调查对象登记时所依据的统一时点。调查资料必须反映调查对象的这一时点上的状况，以避免调查时因情况变动而产生重复登记或遗漏现象；（3）规定统一的普查期限：在普查范围内各调查单位或调查点尽可能同时进行登记，并在最短的期限内完成，以便在方法和步调上保持一致，保证资料的准确性和时效性；（4）规定普查的项目和指标：普查时必须按照统一规定的项目和指标进行登记，不准任意改变或增减，以免影响汇总和综合，降低资料质量；（5）普查的数据一般比较准确，规范化程度也较高，因此它可以为抽样调查或其他调查提供基本依据；（6）普查的使用范围比较窄，只能调查一些最基本及特定的现象。

5.优缺点

优点：由于是调查某一人群的所有成员，所以在确定调查对象上比较简单；所获得的资料全面，可以知道全部调查对象的相关情况，准确性高；普查所获得的数据为抽样调查或其他调查提供基本依据。

缺点：工作量大，花费大，组织工作复杂；调查内容有限；易产生重复和遗漏现象；由于工作量大而可能导致调查的精确度下降，调查质量不易控制。

6.作用：定期掌握一国或地区重大国情国力的基本统计信息，为政府制定社会经济发展战略和方针政策提供依据；同时，普查取得的资料，可以为抽样调查提供抽样框和各种辅助资料，提高抽样调查效果。

7.目前我国主要有三大普查：人普、经普、农普。（1）全国人口普查。每十年一次，逢0开展；（2）全国经济普查，每5年进行一次，逢3、8开展；（3）全国农业普查。每十年一次，逢6开展。

### （三）重点调查

1.概念：是一种非全面调查，是对数据收集对象总体中的部分重点个体进行观测的统计调查方式。重点单位是标志量占标志总量比重较大的少数个体。它是一种非全面调查，既可用于经常性调查，也可用于一次性调查。

2.主要特点：投入少、调查速度快、所反映的主要情况或基本趋势比较准确。

3.调查作用：重点调查的主要作用在于反映调查总体的主要情况或基本趋势。

4.组织形式：根据研究问题的不同需要，可以采取一次性调查，也可以进行定期调查。一次性调查适用于临时调查任务。

5.重点单位选取方式：重点单位通常是指在调查总体中具有举足轻重的、能够代表总体的情况、特征和主要发展变化趋势的那些样本单位;选取重点单位，应遵循的两个原则，一是要根据调查任务的要求和调查对象的基本情况而确定选取的重点单位及数量；二是要注意选取那些管理比较健全、业务力量较强、统计工作基础较好的单位作为重点单位。

### （四）典型调查

1.概念：典型调查是指根据调查目的，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。

2.作用：研究尚未充分发展、处于萌芽状况的新生事物或某种倾向性的社会问题；分析事物的不同类型，研究他们之间的差别和相互关系；在总体内部差别不大，或分类后各类型内部差别不大的情况下，典型单位的代表性很显著，也可用典型调查资料来补充和验证全面调查的数字。

3.特征：主要是定性调查；是根据调查者的主观判断，选择少数具有代表性的单位进行调查；方式是面对面的直接调查；方便、灵活 ，可以节省时间、人力和经费。

### （五）抽样调查

1.概念:抽样调查是一种非全面调查，它是从全部调查研究对象中，抽选一部分单位进行调查，并据以对全部调查研究对象做出估计和推断的一种调查方法。根据抽选样本的方法，抽样调查可以分为概率抽样和非概率抽样两类。

2.特点：调查样本是按随机的原则抽取的，在总体中每一个单位被抽取的机会是均等的；是以抽取的全部样本单位作为一个“代表团”，用整个“代表团”来代表总体，而不是用随意挑选的个别单位代表总体；所抽选的调查样本数量，是根据调查误差的要求，经过科学的计算确定的，在调查样本的数量上有可靠的保证；所抽选的调查样本数量，是根据调查误差的要求，经过科学的计算确定的，在调查样本的数量上有可靠的保证。

3.适用范围：不能进行全面调查的事物；有些事物在测量或试验时有破坏性，不可能进行全面调查；抽样调查方法可以用于工业生产过程中的质量控制；有些总体从理论上讲可以进行全面调查，但实际上不能进行全面调查的事物；利用抽样推断的方法，可以对于某种总体的假设进行检验，来判断这种假设的真伪，以决定取舍。

4.分类

**A.概率抽样**：按照随机原则抽取样本，即总体中的每个个体都有已知的、非零的概率被抽取到样本中。

根据抽样组织形式不同，又分为以下五种。

（1）简单随机抽样（纯随机抽样）：未对总体中的个体进行事先分组或组合，直接从总体中完全随机地抽取样本。是抽样调查最基本的组织形式。

（2）分层抽样：按某一标志将总体的全部个体分层，然后从每层中按照随机原则分别抽取若干个体组成样本。

（3）等距抽样（系统抽样）：将总体的所有个体按某一标志排队，而后按一定的间隔抽取样本单位。

（4）整群抽样：将总体全部单位分为若干“群”，然后随机抽取一部分“群”，被抽中群的所有个体构成样本。

（5）多阶段抽样：以整群抽样为基础，先从总体的所有大群中抽取若干大群，抽中的大群中再抽取若干小群，抽中的小群再抽取若干更小的群，如此下去，最后才抽取所要观测的个体。

**B.非概率抽样**：凭人们的主观判断或根据便利性原则来抽取样本，总体中每个个体被抽取的可能性难以用概率来表示和计算。

（1）任意抽样（随意抽样）：调查者利用现成的名册等资料随意地选取一些个体作为样本，或者利用偶遇的方式选取观测单位进行调查，样本通常由自愿者组成。随意地或任意地从总体中抽取样本。

（2）典型抽样（判断抽样）：调查者凭自己对调查对象的了解和主观判断能力，有意识地从总体中选取若干具有代表性的个体作为样本。

（3）定额抽样（划类选典抽样）：先对总体按一定标志分类，每类分别按一定比例依主观判断抽取若干有代表性的个体作为样本。也就是在每类中进行典型抽样。

（4）流动总体抽样：捕获—标记—放回—再捕获

# 5.大数据时代还需要抽样调查吗？

## 一、变革：从抽样到大数据

### （一）大数据时代的抽样面临一些实际困难

1.抽样框不稳定，随机取样困难.人口流动性加快，企业经营状况变动较快，抽样框快速变动导致随机取样较为困难。

2.事先设定调查目的，会限制调查的内容和范围。抽样调查往往先确定调查目的，调查范围受限；调查会有侧重点，不能全面反映总体。

3.样本量有限，抽样结果不适合细分。样本量限制使得对调查内容进行细分的结果通常因样本量太小而不具代表性。

4.纠偏成本高，可塑性弱。抽样过程中一旦抽样框出现偏误，往往需要重新设计调查方案，成本较高。

随着信息技术、网络技术取得巨大突破，体量大、类型多、结构复杂的海量数据扑面而来。

### （二）大数据对抽样的替代作用

1.大数据主要包括三大领域：一是社会网络数据，微信、微博、Facebook等形成的数据；二是人机交互数据，网购、网游、网娱、交通、医疗等形成的大量记录数据；三是狭义的传感器及机器数据，GPS、智能电表、智能交通、计算与实验数据等。当然，还包含行政记录、监控数据等。

2.大数据的数据采集有别于传统的抽样调查，广泛使用的传感设备、信号识别技术，编译和可扩展的存储系统等，使数据收集的时效性增强，存储和积累更为方便快捷。主要的采集方法有系统日志采集方法、网络数据采集方法和其他数据采集方法等。

3.作为一种新的数据来源渠道，大数据的确会对一些传统抽样数据产生替代作用。大数据条件下，数据收集及处理方式发生重大变革，数据收集将更多地利用现代网络信息技术和各种数据源去收集一切相关的数据，并善于从大数据中进行再过滤、再选择。

## 二、互补：大数据仍需抽样

### （一）大数据获取信息的局限性

1.大数据的覆盖面有限

被互联网、电子设备感知和记录的社会行动者并不覆盖全部的行动者。大数据到底覆盖了哪些主体、对哪些主体覆盖得更严密，没有普查数据或者严格的抽样数据的校验，很有可能出现偏差。尤其当我们关注的主体不是网民、现有用户，而是全部居民和潜在用户的情况下。

2.大数据时代信息的串并成本高

大数据也是有边界的，尽管今天采集和储存数据的成本越来越低了，但仍然是要耗费成本的。现在通过数据开放、黑客攻击、数据交易，很多企业和单位都试图将能够获得的信息都串并起来，增加关于这些主体的信息量（也就是增加变量的个数），但是仍然面临很多的利益的壁垒、技术的壁垒，成本很高。每个单位都有很大的样本量，但是每个样本的变量信息很少。例如，一些数据集中有基本人口特征信息，一些数据中有健康信息，有的数据中有生活方式信息，如果这些信息没有串并起来，要做出新的、更深入的健康研究是很难的。因此，我们可以看到不管是政府部门，还是企业都在强调数据的开放共享，倡导建立数据共享标准，从根本上讲，就是为了将这不同的信息串并起来，将行动主体嵌入到一定的社会网络或者情境中，增加各个研究对象的有效信息量。如果串并不起来，那个“什么叫大数据，情人节订披萨”的段子就不可能出现，也不用担心。

3.大数据时代无效信息过多，信息的精准性下降

如果无效信息所占比重过大，大量的无效信息往往与有效信息混杂，对研究者的统计推断产生严重干扰，则大数据分析并没有很大的意义，数据精准性下降，甚至于会产生误导性的结论，与真实情况大相庭径。因此，如何鉴别和测量大数据中无效信息的影响，有效净化大数据，是摆在深化大数据研究和应用面前需要克服的重要问题。

4.大数据关注相关关系，忽略因果关系

其忽略因果关系追求相关关系这本身就违背了一定的科学性，存在问题。相关关系偏于表象，其分析结果较为脆弱，对于预测分析存在较大偏差。

此时从中抽取部分样本数据更能有效进行数据清洗，以挖掘出数据背后的真规律。总体来看，某些场合下大数据还不能完全代表总体，抽样仍然必要，但抽样环境已发生显著变化。

### （二）抽样也要积极应对所面临的挑战

（1）大数据的大体量、非结构，且来源复杂，使得目标总体的**抽样框难以构造**，而没有抽样框也就难以计算样本单元的入样概率，抽出的样本多属非概率样本，难以应用传统的抽样推断理论，如何利用非概率样本对总体进行推断需要进一步探讨。

（2）针对大数据可能存在的系统性偏差，如何把获取的抽样数据作为大数据分析的**对照基础与验证依据**，以改善大数据的偏倚性，还需寻求有效途径。

## 三、选择：大数据抑或抽样

### （一）问题的类型

关键是要清楚**“何时抽样、何时利用大数**据”？针对不同类别的问题，如何在大数据和抽样之间进行选择，可借助下图来加以说明。

（1）A类问题。如果通过抽样能够显著降低数据处理的复杂程度，同时解决问题的效果（目标函数）没有太大的下降，那么应该采用抽样的方式来开展数据处理。

（2）C类问题。一些数据问题基本上不可能通过处理一小部分数据来达到处理全体数据所能达到的效果，即随着采样率的降低，解决问题的效果会快速下降，此类问题是典型的大数据问题。

（3）B类问题。实践中一些问题的处理效果随着数据量上升有一定提升，但当数据大到一定规模后效果提升并不明显。解决此类问题往往是选取一个有较大规模但并非全体的数据集来处理。

### （二）事件的类型

事件：大概率事件和小概率事件。前者通常选择抽样的方式进行处理，如新政策的支持率、售后服务的满意度等；后者应使用尽可能多的数据以发现这些“罕见”事件，如信用卡诈骗、非法操纵比赛结果问题，只有掌握了全体数据才能找出异常情况。

### （三）研究目的

（1）以**推断总体**为目标，而检测所有个体实际上不可能或者会破坏样本，此时抽样是较好的选择。

（2）以**个体特征分析和应用**为目的，则直接利用所获得的全体数据集进行分析更合适。

（3）一些问题不能采用抽样方式，必须采用全体数据，譬如排名或top10问题、异常值处理等。

## 四、策略：大数据与抽样的融合

大数据时代既要善于利用现代网络信息技术和各种数据源去收集一切相关的数据，又要采用传统的抽样方式去收集与处理特定需要的数据。

**1.充分利用大数据，打造政府统计“第二轨”。**将无所不在的网络数据、物联网数据、行政管理记录、企业销售记录、搜索记录等海量数据逐步纳入统计范畴。大数据拓宽了抽样调查数据采集渠道，大数据为抽样框的及时更新维护提供了便利条件。

**2.切实发挥抽样功能，与大数据相得益彰**。一是统计机构通过抽样调查所获取的数据具有权威性，可作为对照基础和验证依据与大数据进行对比，开展抽样数据对互联网数据的校正与调整。二是把抽样数据看作从混杂的大数据中寻找规律或关系的线索，以更好地进行大数据挖掘和快速探测。

**3.创新抽样方法，适应大数据环境的需要**。针对大数据流环境，探求新的适应性、序贯性和动态的抽样方法，探究大数据的案例抽样方法和基于事件的抽样方法、社会关系网络和图的抽样方法等。大数据背景下的抽样很难构造清晰的抽样框，抽取的样本多为非概率样本，这就需要**解决非概率抽样的统计推断**问题。另外，如何根据已获取的样本逐步调整感兴趣的调查项目和抽样对象，使得最近频繁出现的**热门数据进入样本**，也需要方法上的创新。大数据时代，既要全体又要抽样，有时需分析与事物相关的所有数据，因为大数据更为全面，有时则应分析少量的样本数据，因为抽样更具效率。未来，大数据与抽样将相互补充，携手前行。

# ？6.你认为选择调查组织形式最重要的依据是什么？

调查组织形式概念

对象特点；调查目的；现实条件

1.影响因素

第一，对抽样误差大小的要求（调查目的）。不同抽样方式的误差不同，根据误差要求适当选择。

第二，调查对象本身的特点。取决于对调查对象的了解程度及可控水平。

第三，人力、物力、经费和时间等调查条件。

# ？7.统计调查与实验设计有什么区别及联系？

统计调查概念：统计调查是指根据统计研究的目的和任务，运用科学有效的调查方法，有计划有组织地向调查对象搜集原始资料和次级资料的活动过程。

实验设计，也称试验设计，是对实验进行科学合理的安排，已达到最好的实验结果。

根据实验设计内容的不同，实验设计可以分为专业设计与统计设计。

**区别：**

应用范围：统计调查更加广泛的应用于社会经济现象，几乎所有与社会经济现象有关的统计数据都是观测数据，如GDP、CPl、房价等。实验法更多的是应用于理工农医各领域，通过在实验中控制实验对象以及其所处的实验环境收集到的是实验数据。比如，一种新产品使用寿命的数据，一种新药疗效的数据。自然科学领域的数据大多都是实验数据。

研究方式：统计调查只是研究者“被动”地观察研究对象的实际情况，而未施加任何干预。实验法则是研究者“主动”地对观察对象施加某种人为干预，即对一些外来变量进行控制，而改变自变量来研究其对因变量的影响，而不是消极被动地等待研究对象的发生或改变。

因果关系：实验法可以通过对外来变量的控制，改变自变量而研究其对因变量的影响，并通过与对照组的对比，验证两个变量之间存在相关关系和前后相继关系，且不存在其他的原因，即检验两个变量之间的因果关系。实验法能够直接确立因果关系，这是一般统计调查所不能实现的。若采用统计调查来探究因果关系的话，只能对所获取的数据进一步的进行分析，如进行各种回归分析。

研究结果：相比较而言，实验法的科学性更强，精确度更高。实验法通过有效地控制实验环境，进行反复研究，取得的数据比较客观，具有一定的实用性和可信度，也提高了调查的精确度。

**联系：**相辅相成？

小组ppt

## 一、联系

市场调查作用的充分发挥是一个连续的、跟踪的、反复的调查过程，在这个过程中，调查的主题、内容、范围、形式等是变化的。

实验设计遵循一个基本原则，就是随机对照原则，即必须有处理组和对照组的对比。

总体来说，统计调查与实验设计都是科学研究经常使用也非常实用的两种数据收集方法，二者都是通过对资料的收集、分析，解决一定的问题，为研究提供数据支撑，保证研究的顺利进行。

## 二、含义、特点

### （一）统计调查

1.含义：统计调查是指根据统计研究的目的和任务，运用科学有效的调查方法，有计划有组织地向调查对象搜集原始资料和次级资料的活动过程。

2.特点：统计调查有三个特点，分别是目的性、系统性和科学性。（1）目的性是指统计调查能为决策提供有价值的信息。（2）系统性是指整个调查过程的各阶段相互联系，构成一个有机系统。（3）科学性是指调查的设计、方法必须科学，才能保证信息的真实和价值。

3.原则：（1）准确性，即收集的数据要求准确。（2）及时性，即数据对实际情况的反映要及时。（3）完整性，即收集到的资料要完整。

### （二）实验设计

1.含义：通常是指研究者根据研究目的，对研究对象实施干预措施，观察总结研究效应的结果，回答研究假设所提出问题的整个过程。简单地说，就是将一组随机抽取的受试对象随机分配到两个或多个处理组，观察比较不同处理组之间的效应。

2.特点：（1）研究者可人为设置处理因素；（2）受试对象接受何种处理因素或处理因素的何种水平是随机的。

3.优点：能控制或消除非研究因素对研究结果的影响；误差相对较小，统计效率较高。

3.原则

（1）对照原则

概念：对照是指除了试验组外，再设置一个或多个对照组进行同步试验，以比较试验组的效应情况。

意义：使处理因素和非处理因素的差异有一个科学对比，鉴别处理因素与非处理因素之间的效应差异，消除或减少实验误差。

形式：空白对照、安慰剂对照、标准对照、实验对照、自身对照、历史对照、相互对照。

常见错误：无对照、对照不足、对照多余。

（2）随机原则

概念：是指在抽样时排除主观上有意识地抽取调查单位，每个受试单位以概率均等的原则，随机地分配到实验组与对照组

类型：随机抽样、随机分组。随机抽样是指保证总体中的每一个个体都有同等的机会被抽出来作为样本。随机分组是指保证样本中的每一个个体都有同等的机会被分配到实验组或对照组。

方法：简单随机、区组随机 。一是简单随机化，可通过抛掷硬币、抽签、摸球、查随机数字表或应用操作计算器的随机数字健来完成。二是区组随机化，根据受试者进入研究的时间先后顺序，将其分成内含相等例数的若干区组，而后，区组内的受试者被随机分配至不同组别。

（3）重复原则

概念：重复是指研究样本要有一定的数量，即在保证研究结果具有一定可靠性的条件下，确定最少的样本例数。

作用：有利于充分显示随机变量的统计规律性。

类型：重复实验、重复取样、重复测量。

（4）均衡原则

概念：又称齐同对比原则，指试验组和对照组或各实验组之间，除了处理因素以外，其它一切条件应尽可能相同或一致。

作用：使受试对象受到的非处理因素的影响完全平衡，确保处理因素各水平组间不受其它因素或重要的非处理因素不平衡的干扰和影响，使所考察的处理因素在不同水平条件下对观察结果的影响真实地显现出来。

类型：交叉均衡法：是指在各实验组中又各设立实验和对照的方法，以使两组的非处理因素均衡一致；分层均衡法，是将非处理因素按不同水平划分为若干个单位组（层），然后在各单位组（层）内安排处理因素。

## 三、分类

### （一）统计调查

根据统计调查的性质，统计调查可分为：探索性调查、描述性调查、因果性调查 、预测性调查。

（1）探索性调查

1.概述：探索性调查通常是最无结构和最不正式的调查，进行探测性调查的目的是为了获得有关调研问题大体性质的背景资料。探索性调查通常在项目开始的阶段、正式调查之前进行。

2.用途：获取背景资料；定义术语与概念；阐明问题和假设；确定调研的优先次序

（2）描述性调查

1.概述：当调研的目的只是要了解现状时，可以实施描述性调查。描述性调查通常通过对谁、什么、哪里、何时、怎样等问题的回答来进行。

2.分类：描述性调查可以分为横向研究与纵向研究两大类型。所谓横向研究是指仅在一个时间点上对研究总体进行测定。纵向研究则通过对相同样本的重复测定来完成。

（3）因果性调查

1.概述：因果关系调查是为了了解市场出现的有关现象的之间的因果关系而进行的调查。因果关系调查的主要目的是解决“为什么”。

2.目的：在两个以上的变量中寻找原因与结果关系，确定自变量与因变量，明确变化方向，并建立变化函数。

（4）预测性调查

1.概述：预测性调查是为了预测未来市场的变化趋势及其对企业市场营销活动的影响而进行的市场调研，它着眼于对未来市场状况的调查研究。预测性调查是预测的一个重要步骤，并建立在描述性调查、因果关系调查的基础之上。

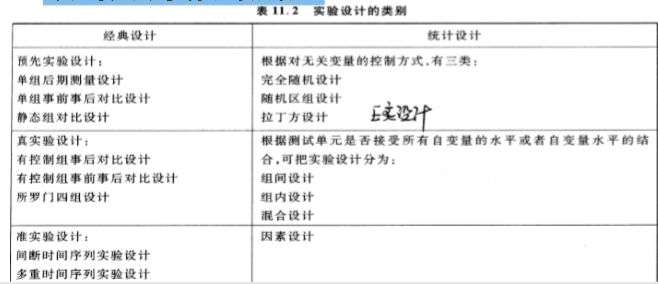
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **比较项目** | **探索性调查** | **描述性调查** | **因果关系调查** | **预测性调查** |
| **目 标** | 发现新想法与新观点 | 描述对象的特征或功能 | 确定变量间的因果关系 | 预测未来市场的变化趋势 |
| **特 征** | 灵活多变；通常是整个研究调查的开始 | 预先提出特定的假设；计划好的结构化的设计 | 操纵一个或多个自变量；控制其他变量 | 着眼于对未来市场状况的调查研究 |
| **方 法** | 专家调查；预调查；二手数据分析；定性研究 | 二手数据；调查法；固定样本组；观察数据等 | 回归分析；  因果关系分析 | 回归预测法；  时间序列预测法 |

### （二）实验设计

（1）按场所分：实验室实验、现场实验

（2）按是否随机分组：非随机化实验和随机化实验

（3）处理水平的数量：经典设计与统计设计。两者的基本区别在于经典设计一次只考虑一个独立变量的一个处理水平的作用，而统计设计可检测一个独立变量的多个处理水平的作用，或者两个及多个独立变量的作用。



**1.经典设计**

**1）前实验设计：**对无关变量不能控制，但是可以操作变化自变量，前实验设计不是严格意义上的实验，实际上只是变量关系的一种描述，所以也叫非实验设计，它能够为真实验设计提供资料。

单组后测设计(单一性个案研究)：特征：只有一组被试且不是随机选择，无控制对照组；实验中只给予一次实验处理；有一个后测。将后测的结果作为实验处理的效应。

单组前后测实验：只有一个被试组且不是随机选择，无控制对照组；仅一次实验处理；有前测和后测，用前后测的平均分做差异检验来说明实验处理效应。

固定组比较设计：在这种设计中，有一个组为控制组，一个组为实验组；在实验中对实验组施以自变量，控制组则做为比较组存在，但两个组不是随机选择，也未加任何控制选择偏向；两组均无前测；实验结束后对各组进行后测，根据实验组与控制组因变量的差异，得出实验的结论。

**2)真实验设计：**按照随机原则选择和分配被试，有控制组对照比较，能够较好地控制内部、外部无效度来源(前面所讲的影响内在、外在效度的因素)使实验得到比较严格的控制的设计。

有控制组后测设计：随机选取和分配被试，控制了选择因素的影响，控制了选择偏差和选择与成熟交互作用两个内在效度影响。

有控制组前后测设计：对于随机等组，如果认为作为等组的依据还不充分，通过前测可以检查随机分组是否存在偏差或者进一步证明两组的等值，如果对于随机分配的等值性把握不大，通过前测可以为证明等组提供依据。

所罗门四组设计：随机等组前后测和随机等组后测组合在一起构成的，随机等组前后测设计和随机等组后测设计都是比较严谨的控制设计。这两种设计模式组合在一起不仅保留了这两种设计的优点，而且还可以获得更丰富的实验信息和更严谨的控制。

**3)准实验设计：**也叫类似实验设计，它与真实验设计有些相似，能够在一定程度上控制一部分无效的变量，但是不能像真实验设计那样随机选择和分配被试。

不对等两组前后测实验设计：通过非随机安排形成两个不完全等值组；前测的目的是为了弥补不能对被试随机选择和分配的缺陷，前测获得两组差异程度的具体数据，为两组后测提供一个进行比较的基础；实验组接受实验方案X，控制组不接受方案X，同时，控制组可以完全不接受任何方案，或接受另一种方案C；用两组前后测结果的差异来说明实验处理的效应。

时间系列设计：时间系列设计基本上是一种单组前后测设计，是对一组被试或一个被试进行周期性的测量，并且在测量过程中引进实验处理，通过实验处理后的变化，推断实验处理是否产生效果。

平衡实验设计：也叫拉丁方设计，所谓拉丁方是指由几个拉丁字母排成的行列相等的方阵。在这种排列中，每个字母在每行每列中只能出现一次，并且必须出现一次。

**2.统计设计**

统计设计可检测一个实验变量的多个水平或多个自变量的效应，它有两个主要特征：一是把处理水平分配到测试目标的实验安排；二是分析实验结果的技巧。

**根据外来变量的控制方式分**

**(1)完全随机设计**：它依据概率的原则，使用随机化方法，随机分配测试单元给各个实验处理，以期各组的外来变量保持相等。由于该方法的假设就是测试单位是随机地分配给各个因素及处理水平的，所以测试单位之间的变异在各个因素及处理水平之间也是随机分布，统计上无差异。

**(2)随机区组设计**：它首先将测试单元区组分类，使每一区组的测试单元尽量保持同质，然后将区组内的测试单元随机分配到各个实验处理中。

**(3)拉丁方设计**：拉丁方设计与随机区组设计的思想相同，但其能够区分两个外来变量，进一步提高实验的精度。该设计是一个含P行和P列，把P个字母分配给方案的管理方案，要求每个字母在每行和每列中都只出现一次，其中P大于等于2。拉丁方设计假设处理水平和外来变量之间没有交互作用，一个外来变量的水平在横行分配，一个外来变量的水平在纵列分析，自变量的水平则分配给方格的每个单元。P＝４时，拉丁方的基本结构如下：

**(4)正交设计：**适合多因素试验，利用正交表科学地安排与分析多因素试验的方法。每个因素的每个水平都相互搭配进行试验。能均匀地挑选出代表性强的少数试验方案；由少数试验结果可以推出较优的方案；可以得到试验结果之外的更多信息

**根据测试单元是否接受所有自变量的水平或自变量水平的组合**

**1.组间设计**：组间设计是指实验中的每个测试单位只接受一种自变量水平或自变量水平的结合。通常把数量相同的测试单位分配到自变量的不同水平或不同的自变量上。

**2.组内设计**：组内设计是指实验中的每个测试单位接受所有自变量水平或自变量水平的结合。

**3.混合设计**：混合设计是指实验设计中既有组间设计自变量，又有组内设计自变量。

（4）前后控制情况

**1.前后无控制对比实验**：即指事前对正常情况进行测量记录，然后再测量记录实验后的情况。进行事前事后对比，通过对比观察了解实验变化的效果。

**2.前后有控制对比实验**：在同一时间周期内，随机抽取两组条件相似的单位，一组作实验组，另一组作控制组(即非实验组，与实验组作对照比较的)，在实验的后分别对二组进行测定比较。

**3.控制组、实验组对比实验**：即同一时间内对控制组与实验组进行对比的实验调查法。其中，实验组核按给定实验条件进行实验，控制组按一般情况组织经济活动。

**4.完全随机对比实验**：譬如：随机地选取试验几个商店做销售实验.

**5.分组随机对比实验**：研究者除了考察基本自变量因素的影响外，还可将某个主要的外部因素控制起来研究。

## 四、步骤

### （一）统计调查

确定调查问题和提出假设——确定调查方案和撰写计划书——实施调查和收集数据——处理资料和分析资料——撰写调查分析报告

### （二）实验设计

确定实验目的、实验对象——确定可比的实验组和对照组——确定把受试对象分配到各处理组中的原则——确定样本含量——确定方法和指标——偏倚及其控制

## 五、研究目的

### （一）统计调查

观察和记录对象本身的固有属性及某些现象，通过对数据的整理分析，比较不同组之间某种现象或相关特征的差异，但是不能说明研究因素与观察现象之间的因果关系。

### （二）实验设计

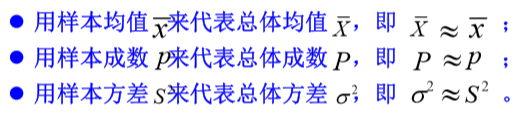
通过积极干预，观察总结研究效应的结果，目的就是说明研究因素与观察现象之间的因果关系，回答研究假设所提出的问题。

# ？8.在进行抽样设计时，样本容量如何确定？

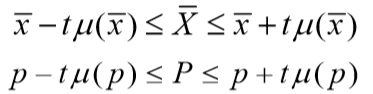
小组展示PPT

## 一、样本估计与推断方法P210

1.点估计：指直接用样本指标作为总体指标的估计值。常见的点估计的形式：



2.区间估计：指在一定把握程度F（t）下，用样本指标和其抽样误差来推算总体相应指标取值的可能范围，即总体指标真值所在的区间。在概率把握度F（t）下，由样本指标x拔、p、所确定的总体参数X拔和P的估计区间即置信区间为：



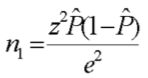
## 二、影响样本容量的因素P211

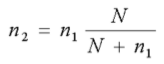
总体的变异程度，即总体方差或标准差。抽样极限误差的大小。概率度值的大小，指对抽样误差范围估计可靠程度的大小。样本的抽选方法。抽样的组织形式。无回答情况。总体规模。预算经费。

## 三、样本容量的确定P212

### 1.成数估计法及变形

对一般抽样设计，给定比例估计的精度，逐步计算样本量（回答率小于100%）。

第1步：计算初始样本量

第2步：对总体大小进行调整：

第3步：如果抽样设计不是简单随机抽样，则用下面公式对样本量进行调整

其中B是设计效应。简单随机抽样设计，设计效应等于1；分层抽样设计，设计效应一般小于等于1；整群或多阶抽样设计，设计效应一般大于等于1。

第4步：根据无回答再次进行调整，以确定最终的样本量，其中r为预计的回答率。

### 2.经验主义方法

经验主义方法主要是基于经验或一些基本的规则如5%规则来确定样本容量,具有主观性。优点：容易确定，简单易行；缺点：无效率、不经济。在置信度为95%的时候，总体规模与样本规模的关系如下：当总体规模大到一定程度时（如100万）,样本规模的增加对抽样误差的影响就微乎其微了。

有关样本量的经验估计：如果是大型城市、省市一级的地区性研究，样本数在500-1000之间可能比较适合；而对于中小城市，样本量在200-300之间可能比较适合；如果是多省市或者全国性的研究，则样本量可能在1000-3000之间比较适合。对于分组研究的每组样本量应该不少于50个。通过试验设计所作的研究，可以采用较小的样本量。如产品试用（留置）调查，在经费有限的情况下，可以将每组的样本量降低至30个左右，好每组在50以上，每组超过100个可能是一种资源浪费。

实验设计中的样本量确定：样本量的大小是控制统计检验功效的主要因素，不同的样本大小所带来的统计功效差异非常大。由此，可以先确定实验或调查所要达到的统计功效（譬如0.8），然后推算出所需要的样本容量。

### 3.实际的样本容量确定

因为市场研究的复杂性和多样性，根据公式计算出的数值只是作为一个参考基础，样本量的确定还需要综合考虑以下几个方面的因素。

**（1）调研的目的。**实际的市场研究中，有些调研目标比较单一，较小的样本量就可以满足要求；而有些调研需要了解全方位、多方面的信息，所需要的样本量就较大。

**（2）调研的产品。**一般而言，对于消费人群比较广泛的产品，如饮料、化妆品等，所需要的样本量较大；而对于消费人群比较集中的产品，如汽车、信用卡等，所需要的样本量相对较小。

**（3）子群分析的需要**。市场研究中，经常要通过市场细分研究了解不同群体的消费习惯和消费需求。因此，设计样本量时，必须保证各子群有足够的样本量可供分析。假定某一调查200个样本是可以满足一定精度和置信度需要的，但如果要了解不同年龄层的消费习惯，如划分成20岁以下、21-25岁、26-30岁、30-35岁、35岁以上5个年龄段，每个子群平均就只有40个样本。（一般的，为了显著性检验的需要，子群分析的样本通常不少于30个，好大于50个）

**（4）调研的区域。**调研区域越大，可能需要的样本量越大。如，一般中心城市的调研，样本量可以设计在300-600个；而对于小城市、县城的调研，样本量可以设计在100-200个。

**（5）调研成本和时间。**企业对于市场研究活动都会有时间和费用方面的计划。时间和费用较为宽松的情况下，可以适当向调查精度方面倾斜，所需要的样本量较大。反之，则设计样本量较小，但是必须保证能够满足调查精度的低要求。

在实际中，在确定样本量时，不考虑时间和费用这两个极为重要的因素是不可思议的。终确定的样本量必须与可获得的经费预算和允许的时限保持一致。

最终样本量的确定需要在精度、费用、时限和操作的可行性等相互冲突的限制条件之间进行协调。它还可能需要重新审查初始样本量、数据需求、精度水平、调查计划的要素和现场操作因素，并作必要的调整。通常，统计调查机构和客户寻求在有效使用费用的基础上（例如缩短访问时间），使用户能对所需的样本量提供经费支持。

注意一个误区：“大城市（区域）多抽，小城市（区域）少抽”，这种说法原则上是不对的。在大城市（区域）抽样太大是浪费，在小城市（区域）抽样太少没有推断价值。

# ？9.你认为是调查的具体实施重要，还是方案更重要？

统计调查是一项复杂细致的工作，具有高度的科学性与群众性。一项全面性的统计调查，往往需要组织许多的人员参加制定详细的调查工作方案与实施计划，有步骤、有组织地开展起来。因此，没有**科学的、严密的调查方案**和**有效的的实施**，要取得预期的效果是不可能的。所以，在组织调查之前以及调查过程中，必须周密地设计调查方案并做好实施，以保证统计凋查任务的完成。

## 一、统计调查方案P25

### （一）调查方案设计的概念

调查方案设计就是统计调查组织和实施者在进行实际调查之前，根据调查研究的目的和调查对象的性质，对调查工作总任务的各个方面和分阶段进行的整体设计，提出相应的调查实施方案，制定出合理的工作程序。调查总体方案是否科学、可行，将影响到整个调查工作能否顺利进行。

### （二）设计调查方案的基本原则

1.科学性原则2.可行性原则3.有效性原则

### （三）调查方案设计的作用

1.是统计调查从定性认识过渡到定量认识的开始阶段；2.在统计调查过程中起着统筹兼顾、统一协调的作用；

3.在实践上适应现代统计调查发展的需要。客观上不存在唯一最优的统计调查方案，每个方案都有其优缺点，最终的统计方案实质上是调查成本与调查数据质量要求之间的权衡。

### （四）调查方案重要性

1.调查目的与任务：只有明确调查目的和任务，才能确定调查的对象、内容和方法，才能保证调查具有针对性。

2.调查的对象和调查单位：只有正确、科学确定调查对象，才能划清要研究的总体界限，这对于保证调查资料准确反映实际情况是十分重要的。

3.调查表：搜集市场调查资料的重要工具。

4.调查方法和组织方式：不同的问题采用的调查方法和组织方式不同，所以选择合适调查方法、方式是很重要的。

5.调查经费预算:能够避免因经费不足等问题从而影响调查进度。

6.可行性研究:方案是否可行是重点，经过可行性研究，尽可能地避免后续实施及调查结果的低效性。

（1）调查方案的可行性研究的方法

逻辑分析法，是检查所设计的调查方案的部分内容是否符合逻辑和情理的可行性分析法。

经验判断法，即组织一些具有丰富调查经验的人士，对设计出的调查方案加以初步研究，以判断方案的可行性。

经验判别法能够节省人力和时间，在较短的时间内做出结论，但也存在一定的局限性，主要是因为人的认识是有限的，有差异的，事物在不断发生变化，各种主客观因素都会对人们判断的准确性产生影响。

试点调查法，通过在小范围内选择部分单位进行试点调查，对调查方案进行实地检验，来检验调查方案的可行性。对于大规模市场调查来讲，是整个调查方案可行性研究中的一个十分重要的步骤。通过试点调查获取所需信息进行分析，发现可能存在的风险，进而对方案进行修改、补充，使其更加科学、完善，使后续正式调查工作更加合理、有效。并且还可以为正式调查取得实践经验。

（2）调查方案的总体评价

调查方案的总体评价可以从不同角度来衡量。但是，一般情况下，对调查方案进行评价应包括四个方面的内容

1、调查方案是否体现调查目的和要求；2、调查方案是否具有可操作性；3、调查方案是否科学和完整；4、调查方案是否具有调查质量高、效果好。

## 二、总结

调查方案的具体实施是调查方案设计的应用。无论方案设计多完美，具体实施工作做不好也是没用的。

在实际调查过程中，除了要遵守工作计划一步步进行，还要根据实际情况进行灵活变通，尽可能围绕调查目的以“不变”应“万变”。

设计调查方案是实施统计调查的前提，科学、周密的调查方案，是统计调查得以顺利实施的可靠保证。

顺利的实施过程是统计调查取得成功的关键因素。

## ？三、调查误差

调查误差是指调查的结果与客观实际情况不一致所造成的误差，分为抽样误差和非抽样误差。

### （一）抽样误差

指使用样本指标值对总体指标值进行估计所造成的误差，也称为随机误差或代表性误差，可以分为实际误差和平均误差两类。影响抽样误差的**主要因素**:总体的标志变异程度、样本容量、样本的抽选方法、抽样调查的组织形式

### （二）非抽样误差

指基于抽样误差之外的各种原因引起的误差。主要存在于抽样设计阶段和现场调查阶段。按来源、性质和处理方法的不同可以分为以下三类：

1.抽样框误差：指由不完善的抽样框引起的误差。

2.无回答误差：在调查过程中由于各种原因而无法对抽取的样本单元进行计量，获取不到有关这些单元的数据所造成的估计量的偏误。

3.计量误差：指抽样调查所获得的数据与其真值不一致所产生的误差，也称回答误差，包括调查人员误差和受访者误差

## 四、误差控制

### （一）抽样框误差

1.抽样框误差的产生

抽样框是有关总体全部单元的名录或地图等的框架，是抽取样本单元的依据。理想的抽样框（也称抽样总体）应该同所研究现象的总体（也即目标总体）一致，但在实践中，抽样总体与目标总体常常不一致，由此产生的误差就是抽样框误差。考察抽样框误差的成因，大致有以下几个方面：（1）丢失目标总体单元。（2）包含非目标总体单元。（3）复合联接。（4）不正确的辅助信息。（5）抽样框陈旧

2.抽样框误差的补救

（1）丢失单元抽样框的补救基本思路有两条：一是想办法将丢失的目标总体单元纳入到不完善的抽样框中，二是对产生于不完善抽样框的数据进行调整。

具体方法有：对丢失单元实行联接。采用辅助抽样框。利用有关资料进行推算。用复查结果调整。

（2）其它类型不完善抽样框的补救

### （二）无回答问题及处理

在设计方案相当合理，编制的抽样框令人十分满意的情况，无回答现象的发生是非抽样误差表现的主要形式之一。无回答类型通常可归结为：

**（1）遗漏。**由随机抽样所确定的调查单位出现“找不到”现象，或者由于客观存在的一些困难，诸如交通极端不便，气候异常恶劣等而无法找到确定要去访问的对象。（四）无回答问题及处理

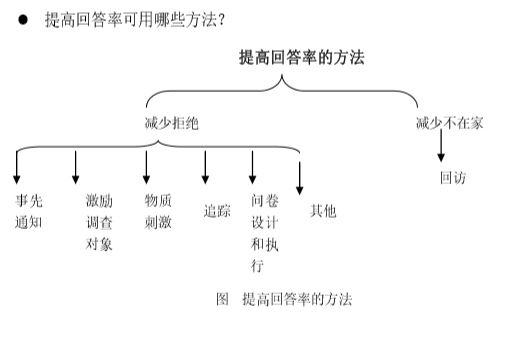
**（2）被调查对象不在家。**有时候可以请家庭中其他人甚至就近换一家进行访问，这样的做法虽然方便，但是在一定程度上破坏了随机性。有时候为了确保抽样的随机性，“不在家”就造成了“无回答”。

**（3）不能回答。**某些被访问者对于问卷中的若干问题缺乏有关资料或者出于各种各样的原因而不愿意回答

**（4）坚决拒绝调查。**这是由于各种原因造成的，尤其是有些问题涉及到个人隐私等等，由此产生的偏差一般较难消除。

## 五、提高回答率：

在调查方案设计和抽样设计的过程中，要认识到问卷的**回收率问题**，并采取有效措施尽可能地提高回收率，减少无回答率。



调查的具体实施是将调查方案的实现过程，好的调查方案能够降低具体实施过程中产生的误差，调查的具体实施效果直接关系到调查数据的质量

PPT第二讲（统计调查方案的概念、作用）

第八讲（调查实施中可能出现的误差）

# 10.调查阶段要提高数据质量需注意哪些方面？

## 一、统计数据质量的重要性

### （一）统计数据的性质

**1.准确性：**准确性是统计数据质量在统计信息客观真实性方面的体现，是统计数据使用者的首要要求

**2.及时性：**及时性是统计数据质量在统计信息的时间价值上的体现，是对统计数据形成和提供的高速度、快节奏、强效率的要求。

**3.完整性：**完整性是统计数据在统计信息的内容含量上的体现，就是要求统计部门提供的统计数据在内容上应该包括使用者所需的所有项目，不能残缺不全。

### （二）数据审核、质量性评估重要性

在统计数据来源多样化的情况下，统计数据的可靠性和有效性直接影响到统计分析和预测能否得出正确的结论。所以统计数据的审核与质量评估显得尤为重要。数据审核是指应用各种检查规则来辨别缺失、无效或不一致的录入，防止发生数据记录的潜在错误，达到防患于未然的目的。数据审核的方法包括调整口径范围、对计算方法和计量单位不统一的换算、重新进行分组和运用各种统计方法对数据质量进行审核与估算。审核的目的就是要保证最后调查所得数据的完整性、一致性和有效性。质量评估是指对调查、整理完毕的统计数据的准确性、代表性、可用性等方面进行评估与检验，目的是为了对它的误差大小做到心中有数。统计部门和统计资料的使用者都可以对统计数据的质量进行评估。例如统计工作中的数据虚假问题，这是最常见的统计数据质量问题，也是危害最为严重的数据质量问题。加强对统计数据的审核与质量评估，有助于避免统计数据使用后负面效应的发生，从而保证统计数据质量评估的权威性。

## 二、影响数据质量的因素

### （一）误差的类型

统计调查中的误差从本质上可分为系统性误差(即非随机误差)和随机误差。

**（1）系统性误差**

1.概念：系统性误差(即非随机误差)是在调查或抽样技术的运用过程中,因与主观因素相关的种种因素而引起的误差，其后果是导致调查结果的失真和估计结果的不准确，使统计资料发生质量问题。

2.特点：这种误差难以计量；引起系统性误差的因素很多较难防止其发生；不具随机性；与样本容量无关。

3.分类：按其来源可分为:抽样框误差、无回答误差和调查员误差等；按产生的原因也可分为：设计误差、估计量的误差、调查误差和编辑误差。

设计误差，是指因在调查方案的设计过程中的错误而导致产生的系统性误差。在抽样方案的设计中，涉及到抽样框的编制目标量和估计量的设计样本容量的确定和抽样方式的选择等。抽样方案的设计还直接关系到在样本的抽取过程中是否严格地遵循了随机性原则，若破坏了随机性原则，则必然会产生系统性误差。

估计量的偏倚，是指所选的估计量（实则是估计方法）破坏了估计的无偏性准则致使产生系统性误差.通常也称其为偏差。从技术上看,很多有偏的估计量可以计算出其偏差的大小，也能有效地把握住其大小.但有的则不能。

调查误差，是指在取得样本数据资料的过程中产生的误差,这部分误差通常与调查者、回答者、资料搜集方式和问卷（或测具）等因素有关，他们会形成在调查过程中出现无回答和回答出现偏误等情况，进而形成系统性误差。

编辑误差，是指在取得样本原始资料之后,在整理、汇总、归类、计算、录入等过程中形成的误差。这部分误差难以把握,但有时可用逻辑的方法检验出来。

**（2）随机/抽样误差**

1.概念：随机误差，主要来源于抽样误差和随机测量误差。因为主要是抽样误差，故通常笼统称之为抽样误差。抽样误差是由于抽样的非全面性和随机性所引起的偶然性误差，即因抽样估计值不同所造成的误差。

2.特点：这种误差可以计算和衡量；它随着样本容量的增大而趋于0。



### （二）影响数据质量的因素

**1.调查的内容：**（1）调查内容的敏感性；（2）问题的难度；（3）问题的兴趣。

**2.调查的条件和环境：**（1）调查的开展时间；（2）调查地点；（3）被调查社区对调查的态度。

**3.应答者的特征：**（1）应答者的社会特征；（2）应答者对入户访问的态度和合作程度；（3）应答者对提问的理解能力；（4）应答者的回答能力及方式: 礼貌偏误、迎合偏误、社会期望偏误、回答问题简单化。

**4.调查者的特征：**

（1）调查员的社会特征；（2）调查技术的掌握程度；（3）调查员的主观能动性和工作责任心。

## 三、现场调查的质量控制

现场调查是经过培训的调查人员在现场(入户调查、拦截访问和观察法)或在办公室 (电话调查、邮寄问卷、电子邮件调查和网上调查)最终完成的调查工作。

### （一）调查准备阶段的质量控制

1.抽取有代表性的样本

2.根据研究目的设计调查问卷:（1）有填表说明；（2）问题的意思应具体明确，没有歧义；（3）问题尽量口语化，通俗易懂（考虑被调查者的教育程度）；（4）问题尽量采用封闭式的选择题；（5）敏感问题排在后面。

3.调查员的选择与培训

（1）要求：对调查员进行严格的选择和培训，须使每个调查员都能准确的理解调查的目的和要求，且在调查实施之前须对被调查群体的特征有初步了解，比如调查的背景等。

（2）培训的方式：采用讲解、模拟训练培训合格调查员。

（3）培训的内容：调查员必须知道的信息；对调查员态度的培训

调查员必须知道的信息：1.问题:为什么要做这次调查?调查员需要此信息调动积极性。2.总体:培训应描述研究总体，使调查者了解将向什么人调查，以及选择这些人为调查对象的主要依据。3.主持者:调查员知道何人、什么单位主持这项调查4.现场调查基本知识:有关的现场调查基本知识。5.抽样:使调查员了解如何抽样，调查过程中哪些行为会破坏样本的随机性。6.标准化:使调查员了解在现场必须有统一的、标准化的行动。如询问问题、记录均应标准化。7.调查员偏误:使调查员了解自己行为所致的偏误。8.应答者的偏误:使调查员了解有应答者为讨好调查员，会发生伪造的答案。9.访问程序:掌握问卷及资料记录的程序。10.向调查员宣布时间要求、工作纪律等。

对调查员态度的培训：1.认识自己工作的重要性，强调本次调查的重要性，对社会、对人民将有何贡献。2.强调必须按规定行动，强调偏误能导致资料无用，教育调查员懂得个人的“经验”与“观察”会使资料“染色”，从而影响调查质量。3.调查中不能激怒对方，必须尊重对方。4.任何情况下，不能与应答者争吵，要记住“应答者就是调查员的顾客”、“顾客永远是正确的”。5.涉及私事必须为应答者保守秘密。

4 组织试调查

### ？（二）调查实施阶段的质量控制

小组ppt

1.调查初期，要及时小结，讨论遇到的问题，统一解决的标准；

2.与被调查者建立密切融洽的关系；

（1）调查员态度要端正，要让被调查者明白你的来意及其个人信息将会得到保护；（2）做好被访者的工作；（3）保持中立，避免诱导和臆测；（4）尽量询问每个成员及家庭主要能说清情况者，减少无回答和回答出现偏误的误差，设法提高应答率。

3.做好初审、复审，及时发现错、漏项，予以改正、补充。在访问调查中，调查员应掌握好面访的时间，合理安排每天的面访次数，并在两次面访之间留出足够时间，补全访问中未完成的记录，以及完成所必须的审核工作。 入户调查中，可以采取电话复核和实地复核等做法。

4.注意减少环境因素带来的影响。

老师课件

三、现场调查的质量控制

**1.对现场工作人员故意误差的控制：**（1）监督：是指派管理人员对现场工作人员的工作进行监督，如电话访问时采用监听。在私人访谈中，监督者可以跟随访问者，在管理现场调查问卷时观察访问者的行为。（2）证实：是指对访问者的工作进行核实。针对篡改和欺骗问题的一种方法是监督者重新与被调查者取得联系来证明他们是否接受了调查；另一种方式是检查问卷，以发现访问者问卷中的某种模式，这种模式会引起监督者对其是否做假的怀疑，那些伪造调查问卷的访问者不可能总是小心地模仿真正的被调查者而做到天衣无缝。

**2.对现场工作人员非故意误差的控制**：监督者或委托方可以进行选择、培训、角色训练，或要求每进行一定数量或一定时间的调查后休息一下等。

**3.对被调查者故意误差的控制：**(1)控制说谎。主要包括:匿名是指向被调研者保证他们的名字不会与他们的回答联系在一起；保密――是指向被调查者保证他们的回答将不予公开；激励――是向被调查者许诺在其参与调查的情况下将给予现金支付、提供礼物或其他有价值的东西；证实检查――即对被调查者所提供的信息进行确认，如一个关于治疗脱发药品的家庭调查中，可以要求看一下被调查者的药品以求得证实。(2)控制不响应。主要包括:物质激励；允许不回答。

**4.对被调查者非故意误差的控制：**(1)周密设计的问卷说明和范例是一种常用的避免被调查者混淆的方法，可以减少误解；(2)对于长的问卷可采用一些提示语，如我们快完成了，现在是最难回答的问题，或策略性地安排一些措辞来鼓励被调查者继续接受调查。

### （三）整理资料阶段的质量控制

1 人工审核问卷上的数据；主要是检查出明显错误及数据不一致;

2 编码: 把填写的文字信息转换为数字代码，以便于分组处理； 发现有缺项与错项应要求所属地区负责核查更正，如无法更正者，则将其作废;

3 数据采取两次录入，完全一致的数据才能进入推算；

4 计算机逻辑检错:有效性审核是检查问卷是否有缺失、数据是否在允许值之内等；一致性审核是检查指标间的关系是否符合逻辑;分布审核是通过数据的分布来确定某一记录是否是离群值。

## 四、大数据背景下统计数据质量的影响因素与解决对策：

### （一）流程维度：

1.建立科学的统计数据质量评估体系；2.建立基于数据仓库的统计数据集成平台；3.尊重客观统计规律，拓新统计技术与方法

### （二）技术方面的解决对策：

1.研究高效率、低成本的数据存储方式；2.将数据质量检测技术贯彻到数据整个生命周期；4.将大数据分析技术运用到统计工作的各个方面。

### （三）管理方面的解决对策

1.树立科学合理的统计观念；2.加强统计法制建设；3.完善数据标准；4.完善统计制度体系