

## （二）数据挖掘过程

P5

### 1) 定义挖掘目标

清晰明确的挖掘目标是进行数据挖掘的前提，同时也是能够最大限度地发挥数据挖掘作用的关键。

### 2) 数据取样

针对数据挖掘目标如何选取样本。在选取样本时有三大原则，即时效性、可靠性和相关性。必须保证选择的样本数据是最新的、真实可靠并且与挖掘目标是高度相关的。同时选择数据的简明性，精选数据。

### 3) 数据探索

进一步对数据进行分析探究，为了保证预测质量需要对选取的样本数据进行探索、审核和必要的加工处理是必要的。数据探索主要有：相关分析、异常分析、周期性分析、缺失值分析和样本交叉验证。

### 4) 数据预处理

由于样本数据量一般较为庞大，数据结构较为复杂，样本数据维度过高，有缺失值，重复记录，不一致等等。对样本进行预处理也是必要的。

### 5) 模式发现

数据预处理之后，开始构建挖掘模型，在建模之前要考虑本次的目标是数据挖掘哪方面的应用，针对具体的应用类别选取合适的算法。

### 6) 模型构建

模型构建包括选择什么挖掘算法，模型构建思路，具体操作过程是怎么样的

### 7) 模型评价

对数据挖掘的结果进行评价，对其中的不足之处以及可取之处进行总结，并不断改进。

统计调查是为了获取信息，进而辅助决策，当然统计调查也有其局限性。

P6

有关统计调查的职位名称及职责

- 1、市场研究副总裁，高级职务，负责制定市场研究部的目标
- 2、研究总监，高级职务，负责所有研究项目的制定与执行
- 3、助理研究总监，上一级行政助手，指导与监督
- 4、项目经理，全面负责研究项目的设计、执行和管理工作
- 5、统计师/数据处理专家，统计技术、理论应用，实验设计，数据处理和分析
- 6、高级分析师、分析师、初级分析师，数据搜集、分析、撰写报告、项目进度与成本控制、问卷设计与测试、二手数据分析、问卷编辑
- 7、现场经理，负责调查员和督导的征选、培训和监督
- 8、作业督导，负责有关作业的监督与指导，有时也直

接参与

P9

中国调查行业构成

- GMR 系统：1988 年 7 月 1 日成立广州市场研究公司（GMR）；1990 年，成立了国内第一家私营企业—华南市场研究有限公司，1992 年初成立了远东市场研究公司，后又成立东方、卓越、佳景、大通、业绩等公司。
- 社会调查系统：1986 年产生北京社会调查所（后为中国社会调查所），并孕育了零点、顿邦、大视野、世纪蓝图等。
- 广告研究系统：1983 年上海广告公司通过市调为雀巢公司创造了“味道好极了”这一甜美的口头广告语
- 高校系统：中山大学的广州思路市场研究有限公司、北京广播学院的调查统计研究
- 统计系统：具有统计背景的公司包括北京华通、精诚兴、中国经济景气检测中心、塞诺、中怡康、上海恒通。
- 海外系统：ACNielsen、Gallup 等
- 其他：P&G 系统、央视调查公司

大数据分类——按数据来源划分  
—阿里巴巴

P14

- 目前保存的数据量为近百 PB
- 拥有 90% 以上的电商数据
- 交易数据、用户浏览和点击网页数据、购物数据—百度
- 2013 数据总量接近一千个 PB
- 中文网页、百度推广、百度日志、UGC
- 以 70% 以上的搜索市场份额坐拥庞大的搜索数据—腾讯
- 总存储数据量经压缩处理以后在 100PB 左右
- 数据量月增 10%
- 大量社交、游戏等领域积累的文本、音频、视频和关系类数据—电信行业
- 用户上网记录、通话、信息、地理位置
- 运营商拥有的数据数量都在 10PB 以上
- 年度用户数据增长约数十 PB—金融与保险
- 开户信息数据，银行网点和在线交易数据、自身运营的数据
- 金融系统每年产生数据达数十 PB
- 保险系统数据量也接近 PB 级别—电力与石化
- 仅国家电网采集获得的数据总量就上 10 个 PB 级别
- 石油化工、智能水表等领域每年产生和保存下来的数据量也达到数十 PB 级别。
- 公共安全

- 北京：50 万个监控摄像头，每天采集视频数据量约 3PB
- 整个视频监控每年保存下来的数据在数百 PB 以上
  - 医疗卫生
- 仅广州中山大学医院 2013 年数据量为 1000 个 TB
- 整个医疗卫生行业一年能够保存下来的数据就可达到数百 PB
  - 交通
- 航班往返一次能产生数据就达到 TB 级别
- 列车、水陆路运输产生的各种视频、文本类数据，每年保存下来的也达到数十 PB。
  - 气象与地理
- 中国气象局保存的数据约 4-5PB，每年约增数百个 TB
- 各种地图和地理位置信息每年约数十 PB
  - 政务与教育
- 北京市政务数据资源网涵盖旅游、教育、交通、医疗等门类，一年上线公布了 400 余个数据包
- 政务数据多为结构化数据
  - 制造业——存储数据类型：
    - ①产品设计数据：以文件为主，非结构化，共享要求较高，保存时间较长。
    - ②企业生产环节的业务数据：数据库结构化数据，生产监控数据：数据量非常大
  - 其他传统行业：
    - 线下商业销售、农林牧渔业、线下餐饮、食品、科研、物流运输等行业数据量剧增
    - 行业数据量还处于积累期，整个体量都不算大，多则达到 PB 级别，少则百 TB 甚至数十 TB 级别。

P14

- 系统日志采集方法
 

很多互联网企业都有自己的海量数据采集工具，多用于系统日志采集，如 Hadoop 的 Chukwa，Cloudera 的 Flume，Facebook 的 Scribe 等，这些工具均采用分布式架构，能满足每秒数百 MB 的日志数据采集和传输需求。
- 网络数据采集方法：对非结构化数据的采集
 

网络数据采集是指通过网络爬虫或网站公开 API 等方式从网站上获取数据信息。该方法可以将非结构化数据从网页中抽取出来，将其存储为统一的本地数据文件，并以结构化的方式存储。它支持图片、音频、视频等文件或附件的采集，附件与正文可以自动关联。除了网络中包含的内容之外，对于网络流量的采集可以使用 DPI 或 DFI 等带宽管理技术进行处理。
- 其他数据采集方法
 

对于企业生产经营数据或学科研究数据等保密性要求较高的数据，可以通过与企业或研究机构合作，使用特定系统接口等相关方式采集数据。

P14

在当今大数据时代，收集、存储、分析海量数据，挖掘

数据之间相关关系，洞察数据变化规律和趋势特征，是经济社会统计研究的重要内容。

一方面，大数据以其高频率、细粒度、多样化的优势，为政府统计开辟了新的数据源，成为政府统计数据的重要补充来源，提升了统计服务能力；另一方面，大数据获取难度大、数据质量问题多、统计应用难度大，作为政府统计数据源表现出较大的脆弱性，面临着技术、方法和实际操作层面的诸多困难和瓶颈。

从发展趋势看，大数据在政府统计中应用潜力大，前景广阔。近年来，各国政府统计机构积极研究探索大数据的统计应用，大力推进大数据与政府统计工作融合，积累了许多有益经验，取得了很多成功的应用案例。联合国全球大数据工作组（UNGWG）、世界银行、国际货币基金组织、欧洲经济委员会、欧盟统计局等有关国际组织也分别组织开展大数据统计应用试验合作项目，研制大数据质量评估标准，建立大数据统计应用案例共享平台，举办应用技能培训，共同解决大数据统计应用的世界性难题。

P16

## （二）主要方法思路

大数据应用不能再现传统统计生产过程，也不能运用传统的统计方法来处理，而主要依赖各种算法来挖掘发现大数据的规律特征，通过建模方法进行统计估算和推断。大数据统计推断的技术可行性、方法合理性以及结果的有效性，是评估判断大数据统计应用是否成功的重要依据。在认定大数据的统计价值和成本效益之后，大数据的统计应用通常须经过 3 个阶段：

大数据统计应用的方法很多也很复杂，具体方法取决于不同的专业领域和不同的数据类型。实际应用时要突破传统统计理论的框框，创新统计方法和统计思维。

在价格统计方面，扫描价格数据和网络抓取价格数据已成为许多发达国家编制价格指数的新数据源。面对采价产品数量规模大、更新换代快的全量数据，需要突破“比较不同时期固定数量篮子同质可比的产品和服务价格”这一传统价格指数理论框架。为此，英国、荷兰、比利时等统计局专门针对网络抓取数据和扫描数据提出比较固定类群相对同质可比的产品价格，观测反映消费者购买同质同类产品群的价格变化。

英国提出采用大型数据集聚类价格指数方法（Clustering large datasets into price Indices，简称 CLIP）计算基本分类以下不同时期相同产品类群的价格之比，而不是具体产品价格之比。其前提条件是要对巨量的采价产品进行聚类，最大限度地增加群内产品的高度同质性和相似性，以保证群内产品没有显著差异，以降低价格指数的偏差。

欧盟统计局提出采用动态方法和多边比较方法（GEKS

法、TPD 法、GK 法等）计算分类价格指数，以解决采价产品更新快的问题。显然，这些理论框架方法是对传统价格指数理论方法的拓展和延伸。

国际货币基金组织计划修订《消费者价格指数手册：理论与实践》，增加新数据源收集和应用的理论方法。大数据统计应用的理论突破和方法创新必须进行公理检验，具有经济含义的可解释性和统计推断的有效性，充分体现方法的无偏性和科学严谨性。

#### 四、大数据在政府统计中应用的推进路径

P17

• 政府统计机构对统计信息的垄断时代已不复存在，只有通过“设计的”传统数据源与“发现的”大数据源的相互融合组合，才能产生强大的统计信息系统，从而增强政府统计提供高效而有效的统计服务能力。推进大数据统计应用是未来政府统计的发展方向 and 必然趋势。自政府统计建立形成以来，曾经历了从普查到抽样调查的统计范式转变。可以预见，大数据统计应用是政府统计的又一次范式大变革，政府统计机构将从单纯的数据收集者转变为兼具不同数据源的融合者、数据源质量的评估者。大数据与政府统计工作融合预期将经历 3 个渐进的变化阶段：从短期看，普查、抽样调查等传统统计调查仍是政府统计数据的主要来源，而大数据则是政府统计的重要辅助来源；从中期看，在政府统计信息系统中，大数据的影响逐渐上升，而传统调查数据的影响有所减弱；从长期看，大数据源将部分（而不是全部）替代传统调查项目，最终实现大数据和传统调查数据优势互补、相互融合的目标。

• 近年来，我国政府统计机构确定了“总体设计、牵头攻关、先易后难、专业突破”的总体思路和“打造政府统计数据来源第二轨”的工作目标，扎实推进大数据在政府统计中的应用，已取得初步成果。

P19

• 随着科学技术的发展，各种先进的仪器、仪表等手段被逐渐地应用到市场调查中。

• 市场调查人员可以借助摄像机、收视计数器、交通计数器、监测器、闭路电视、计算机等设备来观察或记录被调查者的行为，以提高调查的准确性。

P28

• 例如，一个公司在—个地区播放广告后，测量其销售量的变化。这类设计对于外在因素几乎不加控制，可能发生历史因素及成熟因素的影响，从而影响实验的效度。

• 举例：对于测量广告效果的公司来说，广告播放前先测量一下产品的认知度、美誉度等指标，在广告播放后再重新测定这些指标。O1 和 O2 之间的差异就是对实验处理影响的度量。实验中也会受到成熟因素、丢失效应和测量效应的影响，但在时间和成本上具有优势。

• 例如：研究者在两个班上进行教学实验，实验班引入

新的教学方法(X)，称之为实验组，控制组沿用传统的教学方法，一学期后进行测验，并对两组物理测验成绩进行比较，看哪种教学方法更有效。

G1: 1 班 X 新的教学方法 O1 后期测验

G2: 2 班 — 传统方法 O2 后期测验

• 例：一位理疗师对—组病人实施—项为期 8 星期的康复计划。前 4 周小组成员每天都接受传统治疗方法，每周末根据—份身体能力测试表进行测试。从第 5 周开始施行了一种实验治疗法，同样每周末进行身体能力测试。

P30

• 例如，单因素完全随机设计中测试单元的分配如下图所示，该实验有 1 个自变量 X，其包括 3 个处理水平，每个处理组有 4 个测试单元，测试单元随机分配并接受—个处理水平。

• 例如，—家企业想了解消费者对其产品价格的反应，如果认为消费者的收入会影响到其反应，则事先将被调查者按收入水平分成若干同质的区组，然后将每一区组内的消费者随机分配给实验处理组。单因素随机区组设计中测试单元的分配如下图所示：

P33

基本单位名录库是基础，统一的数据采集处理软件系统是平台，联网直报系统是手段，而企业—套表制度则是“四大工程”的核心。

建设“四大工程”，是统计理念的重大革新，是统计调查流程的系统再造，是统计数据生产方式的深刻变革，对于推进统计数据采集、传输、汇总、加工环节的科学化和规范化，提高基础数据质量具有十分重要而迫切的意义。而“企业—套表”改革是推动统计工作规范统一、改革创新、公开透明的重大举措，是四大工程的核心，是当前中国统计系统的“第—号任务”。

P37

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—般小于等于 1；整群或多阶抽样设计，设计效应—般大于等于 1。

P39

• 抽样调查从设计开始就必须注意尽量避免非抽样误差，抽样设计方案对于抽样结果是否精确起着很重要的作用。—个设计很差的抽样方案将会带来足以使抽样失败的非抽样误差，《文摘》杂志的民意测验则是一个有说服力的实例。

• 设计带来非抽样误差有很多表现为抽样框出了问题，《文学摘要》的抽样框明显地偏向于某—部分人而忽略了另—部分人，从中产生的非抽样误差使得抽样结果不能反映总体（全体选民）的意见。这种致命的错误来源于抽样框“丢失”了总体中占有—定比例的单元，如果这—部分被“丢失”的群体在调查关心的参数方面有其

独特的一面，那么这种“丢失”引起的非抽样误差使推断或估计带有明显的偏性。

• 如果我们的抽样方案是分层抽样，分层抽样不太可能按照总体参数去实施，而常常根据若干辅助信息来进行，当然这些辅助信息与调查变量应当有较强的相关程度。倘若这些辅助变量资料不健全，不准确甚至借用这些辅助变量本身属于判断失误，由此引起的非抽样误差可能会严重威胁到估计的准确性。

• 有时候陈旧的抽样框将带来我们所不希望看到的非抽样误差，因为陈旧的抽样框会造成总体单元的“丢失”现象，一般地它还难以反映近期有关变量的一些变化。以上所述的非抽样误差属于在抽样之前的设计方案工作及编制抽样框中需要认真对待的。

P40

经典统计学观点认为数据分析是通过局部样本开展统计推断，以了解总体的规律性。在收集数据和分析数据能力受限的时代，为有效认识总体特征，抽样应运而生，产生了各式各样的抽样技术。

抽样调查具有经济性、时效性、准确性和灵活性等特征，能够补充、订正全面调查的结果，对总体假设进行检验。虽然具有众多优势和作用，但抽样是在事先设定目的前提下展开工作，不管采用多完美的抽样技术，抽到的只是总体的一部分，难以完全准确地代表总体。

P41

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，譬如在居民收入支出大都是通过银行转帐的地区，银行交易数据即可一定程度上替代居民家计调查数据。

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。

进入大数据时代，能对全体数据进行分析，不一定要抽取样本，但这并不意味着抽样就要退出历史舞台。当所获取的“大数据”总体不能完全代表目标总体时，有一部分个体被遗漏在外，大数据分析可能得到有偏结果。

首先，目前来看并非所有数据都可以通过网络信息系统获得，因为并不是所有产业都已实现智能化，还有很多数据只能通过传统的抽样方式获得；

其次，即使是网络数据，某些情况下对总体进行分析也并非最优选择，例如当面临均匀度很大的总体时，随机抽取部分单位作为样本开展分析就可以得到理想结果，此时并不需要去费时费力分析总体。

大数据带来的既有信息也有噪声，通常噪声居多，使得数据分析容易为假像所迷惑，造成规律丧失和失真。

大数据的生成与采集在人为设计的框架之下，也可能存在系统性偏差，譬如社交网络中的人群数据与真实总体之间可能存在明显偏差。

所以，大数据条件下随着众多缺失、含有噪声甚至是错误的数据进入到数据库中，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。

P41

#### • 大数据抑或抽样

实际当中采取大数据还是抽样的方式，还取决于研究问题的性质。

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。

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，比如测试灯泡、汽车碰撞试验等，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，如个性化推荐、精准营销等。(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。

#### 分层

—按地区：分为 6 大区

—按规模：分为特大、大、中、小和县城

—全国城市和县城被分为 30 层

层内按照城市人均工资排队，采用人口比例不等概率（PPS）方法抽选调查市县

P42

P46

• 将简单、温和、不使人害怕但又必需的问题安排调查表的开始；将困难及使人害怕的问题放在近结束时。

• 有些人口统计学资料可能被认为侵犯性的，如个人收入，应将这些问题放在后面一些，除非这些回答较早用于筛选目的，如用来决定是否纳入调查对象。

• 有关居住、职业或过去情况，应按照从现在到过去或从过去到现在的时间顺序，不要混乱。

• 如果调查表涉及不只一个专题，应在完成一个专题后再转到另一个专题。

• 适当应用过渡性词语转换话题，使应答者转变思路时更觉容易些。

• 恰当安排问题的类型，以增加多样性，减少反应的固定模式。

• 为保证所有可能情况都已包括在内，需为过滤性问题设立一个流程图。

- 尽量使调查表简短，删去冗长的问题，或可能无法分析的问题。
- 被访者容易回答且较为关心的内容先提问，专业性强的具体细节问题和敏感性问题应尽量放在后面；
- 提问的内容应从简单逐步向复杂深化，对相关关联的内容应进行系统的整理，使被访者不断增加兴趣；
- 封闭性问题放在前面，开放性问题放在后面。

### 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。

#### 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、去芜存菁选取我们想要的。

#### 三、大数据时代下数据收集的问题

大数据背景下，我们收集到的数据类型的更加丰富，当万物皆数据的时刻到来时，商业发展的更多新机会将会出现，数据将会帮助我们更好地做出判断。当有些表面上完全不相关的行业数据收集关联起来时，便会产生新的商业价值，行业之间的联系将会变得更加密切。但同时，大数据时代的数据收集也存在缺陷：

- 1.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”。
- 2.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了严重干扰。
- 3.准确度不够。大数据收集的主要途径是通过互联网中的关键词搜索进行相关统计分析，但是在数据的实际收集过程中，研究的复杂问题并不能仅用个别关键词就能真实反映。
- 4.法律法规不全面，导致暴力收集数据和隐私泄露。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。

#### 四、如何进行大数据时代下的数据收集

- 考虑到传统数据和大数据收集的针对性、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。
- 对政府统计部门而言，应整合利用各类大数据资源，建立一套全国性大数据资源库，建立一套快速高效的统计调查平台；解决大数据与传统调查方法，特别是抽样调查的辅助、融合推断问题。政府统计系统在进行大数据采集时，要充分利用统计系统自身的大数据，开发利用好普查数据，加强、完善名录库建设；充分利用好其他政府部门的行政记录数据，加强沟通、协调与合作；与掌握大数据的公司战略协作，争取开放数据接口；利用网络爬虫等技术开展数据采集。
- 引用邱东老师的观点：如果说原来的统计实务的重心在于收集基础数据，难点在于如何获取数据，那么在大数据时代，则更在于如何选择有用数据。如果说原来重心是“做加法”一无中生有；那么现在重心则是“做减



法”——“有中生用”。如果说大数据是一座大矿山，我们要做的就是在这座矿山中去伪存真、去芜存菁选取我们想要的数

④调查方法与调查组织形式有区别吗?各有哪些?

•调查方法是指统计机构和统计人员搜集统计资料的方法，样本单位确定之后，对这些单位实施调查，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。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，统计调查组织形式取决于调查对象的特点以及调查的目的和要求。

•两者存在区别：调查组织形式是抽象的，相当于一个理论；而调查方法是具体的，相当于一个实际活动方法。例如，我们对某一类情况进行调查，首先要选择一个合适的调查组织形式，从而在一个总体中选出一个样本；其次，对于这个样本中的每个个体，我们要选择合适的调查方法，以得到想要的数

•调查方法包括以下几类：

1.直接观察法：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.报告法：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。

3.采访法：由调查人员对被调查者进行采访，根据答复来收集数据的方法。

4.通讯法：把要调查的问题邮寄或电子传送给被调查者，填好后返回。

5.文献法：根据调查目的，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。

•调查组织形式包括以下几类：

1.统计报表：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.专门调查

(1)普查：根据特定的统计研究目的而专门组织的一次性的全面调查，用以收集所研究现象总体的全面资料。

(2)重点调查：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。

(3)典型调查：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。

(4)抽样调查：一种非全面调查，从总体中抽取样本，以样本推断总体。根据抽取样本的方式不同，抽样调查可分为概率抽样和非概率抽样两类。

数据收集方式和方法的选择，关系到我们能否及时、准

确、完整地收集到所需的统计数据，也关系到我们在调查过程中要投入多少人力、物力、财力，因此，要根据研究的目的和我们的能力，选择合适的数据收集方式和方法。

⑩调查阶段要提高数据质量需注意哪些方面?

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，补充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。

正交表  $L_8(2^7)$

试验号	列 号						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

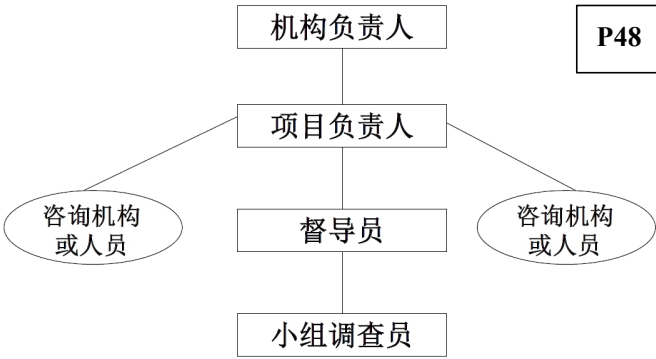
P31

正交表  $L_9(3^4)$

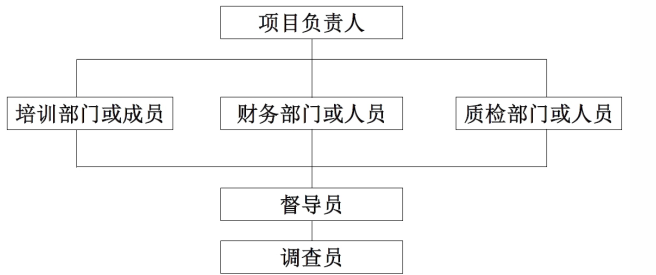
试验号	列 号			
	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

正交表  $L_8(4^1 \times 2^4)$

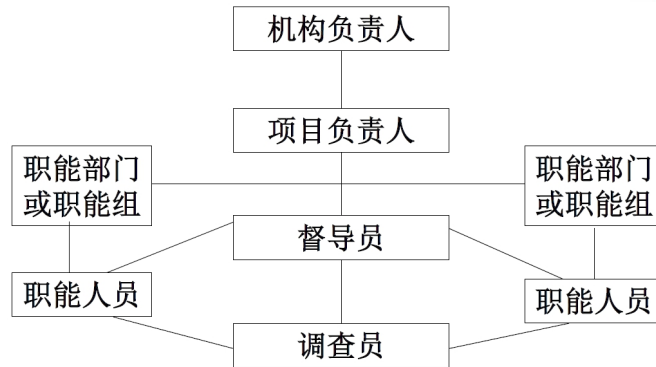
试验号	列 号				
	1	2	3	4	5
1	1	1	1	1	1
2	1	2	2	2	2
3	2	1	1	2	2
4	2	2	2	1	1
5	3	1	2	1	2
6	3	2	1	2	1
7	4	1	2	2	1
8	4	2	1	1	2



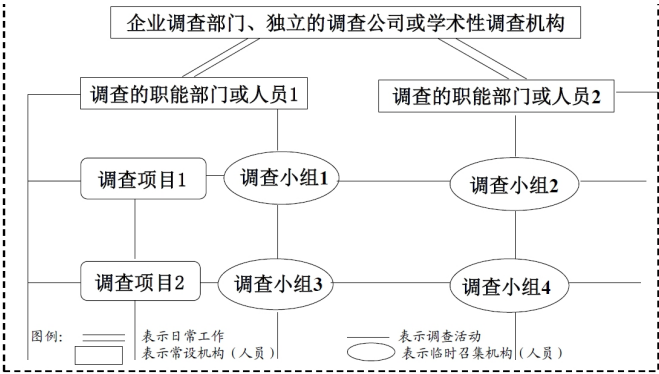
直线式组织结构模式图



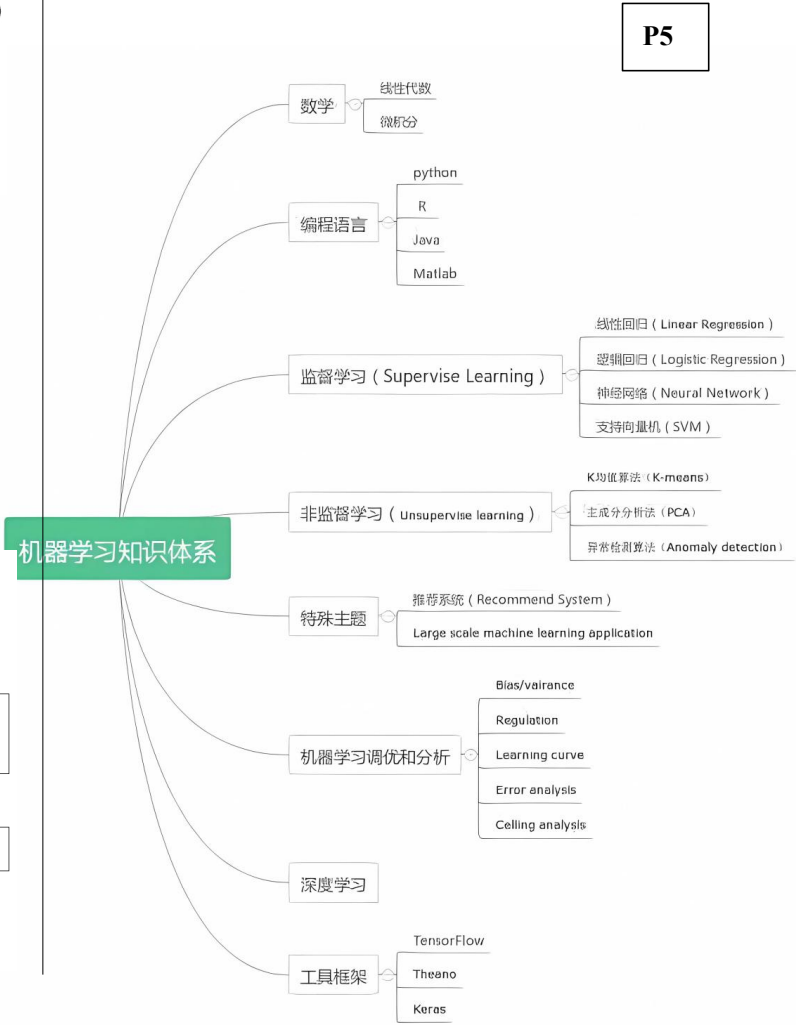
职能式组织结构模式图



直线职能式组织结构模式图



矩阵式组织模式图



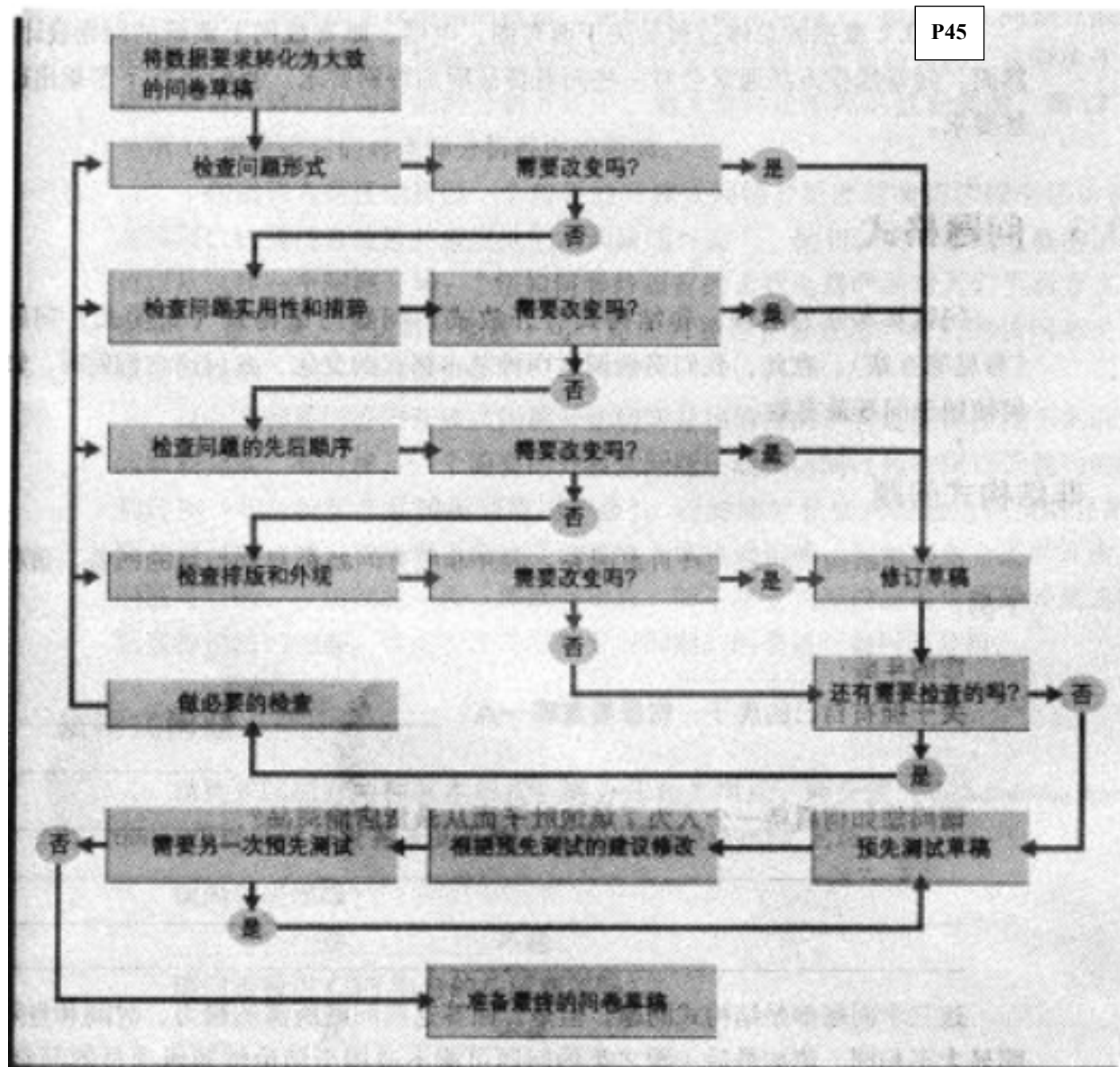


图 10.2 问卷设计过程