

第二章 数据搜集

#一、原始数据的搜集方法

1. 统计调查的方式：

- 定期统计报表制度

定期统计报表制度是一种按照国家法律规定**自上而下布置统一的报表**，然后**自下而上地逐级上报汇总报表资料**的调查方式

优点：数据资料具有全面性和连续性；数据资料的统一性和及时性；数据资料的来源和准确性有可靠的基础

不足：报表过多加重基层负担；个别单位可能会瞒报或者虚报某些资料

- 普查

普查是专门组织的一次全面调查，所调查的内容既可以是某一时点的现象也可以是一定时期的过程性现象

优点：它比任何其他调查方式所取得的资料都更全面，更系统，更详尽

缺点：工作量大，需要花费大量的人力、物力、财力和时间，所以不经常进行

- 随机抽样调查

随机抽样调查是一种非全面调查，它是在全部被调查的总体中随机地抽选一部分单位进行观察，并根据从样本得到的数据来推算总体的数量特征。

特点：按照随机原则抽样；目的是利用样本数据从数量上推算总体数量特征和数量表现

保证抽样随机性是抽样调查的第一要务

常用的随机抽样调查方法有：

- **简单随机抽样(纯随机抽样，或者完全随机抽样)**

按照原有的总体状况，不加任何限制，确保每个个体都有同等被抽中的机会的抽样方法。分为放回抽样和不放回抽样，适合总体规模不大，内部差异小

- **等距抽样：**将总体按照某种标志进行排序，然后按照规定的间隔从已经排序的总体中抽取样本。排序方法有：按照与抽样调查内容无关的标志排序和按照与抽样调查的内容有关的标志排序。抽样方法有：直线等距抽样，循环等距抽样

- **类型抽样(分层抽样)**

把总体分为满足**互斥性、完备性和相似性**的要求的群体，然后组中随机选取样本分层的准则是讲总体中性质相同或者相近的个体归为一层，使得各层内的分布比较均匀，形成若干层

各层样本容量确定分为等比例分配和奈曼分配

分层抽样的作用：

可以在不增加样本容量的情况下，减少抽样误差，提高抽样估计的精度；

在既定的抽样估计精度下，可以减少样本单位，节约调查费用；

分层抽样资料既可用于对总体的统计推断，也可用于对层的统计研究；

分层抽样的组织管理比较便利。

- **整群抽样(聚类抽样)**

将总体分为若干个互不交叉的群，然后从所有的群中随机抽取若干个群作为样

本，即对抽中的群进行全面观测的一种抽样组织形式。群内方差尽可能大，群间方差尽可能小

- **非随机抽样调查**

非随机抽样调查是一种非全面调查，这种调查不按照概率均等原则，而是根据人们的主观经验或其他条件来抽取样本

常用的几种方法：

- **典型调查**：典型调查是在调查对象的时候有意识地选取个别或者少数有代表性的单位进行调查

典型调查的首要问题是如何挑选典型单位。选取规则：

如果要了解总体的一般数量表现，可以选择中等水平(平均型或者多数型)的单位作为调查单位

如果要较为准确地估计总体的一般水平，首先对总体中的个体划分若干类型，然后再从各类中按其比例大小选择若干典型单位进行调查

如果要总结经验或者失败的教训，应该选择先进或者落后的单位作为典型，做深入细致的调查

- **重点调查**：重点调查是在被调查总体中选出一部分重点单位进行调查，这些重点单位虽然占一小部分，但是它们在所调查的数量指标上有较大的比重

重点调查确定方法：

确定一个最低标志值，凡是标志值达到或超过最低标志值的个体就是重点单位；

确定一个最低的累计标志比重，如75%，将各个单位按照标志值由高到低排序并依次计算累计比重，当累计比重大于等于所要求的最低累计比重时，被累计的单位就是重点单位。

- **典型调查与重点调查的异同：**

同：都属于非全面调查，非随机抽样调查

异：① 调查单位选取的方式不一样，典型单位是有意识的选取部分有代表性的单位，单位的选择带有主观性，重点单位要选取其数量特征在总体的数量特征总量中占有较大比例的单位，使其能保证有足够的代表性，具有客观性。② 调查目的不同，典型调查的目的是推广先进经验或为吸取落后教训等，重点调查的目的是反映经济现象的基本情况

2. 统计调查方案

设计统计调查方案应该考虑以下主要问题：

- 调查的目的和任务
- 调查对象，调查单位和报告单位
- 调查内容和调查表
- 调查时间和期限
- 调查的空间标准
- 调查组织实施计划

3. 问卷调查法

问卷调查法就是以问卷的方式提问，由被调查者自愿回答的一种搜集资料的方法，又称为民意测验

在设计问卷的时候需要考虑的问题：

根据具体情况，确认问题的提问方式和调查的具体方法

问题必须准确、简洁
要考虑问题的可行性
避免诱导性的问题

#二、次级资料收集

数据的第二个来源是利用已发表的汇编资料。政府机构所编辑出版的统计资料是宏观、微观数据的主要来源，其次一些商业机构或研究所也会为了某种目的而研究和发布各种数据资料。

次级资料来源有：

国家统计局出版的主要统计资料汇编刊物

其他政府机构及研究所出版的主要统计资料汇编刊物

国际组织机构等提供的资料

次级资料的评估：

- 是谁收集的
 - 为什么目的而收集
 - 什么内容
 - 是怎样收集的
 - 什么时候收集的
-

#三、调查误差

登记性误差和代表性误差：

登记性误差(非抽样误差): 指调查人员登记过程的重复、遗漏、记录失误，调查人员提问不当或调查者记忆不清而回答错误，调查人员故意瞒报、虚报等原因导致的误差。(分为偶然性登记误差和系统误差)

代表性误差(抽样误差): 指由于用部分代表总体所必然产生的误差。

误差不可避免，尽量减少

数据质量要求

精确度：最低的抽样误差或随机误差

准确性：最小的非抽样误差或偏差

关联性：满足用户决策、管理和研究的需要

及时性：在最短的时间里取得并公布数据

一致性：保持时间序列的可比性

最低成本：以最经济的方式取得数据