# 统计调查中的抽样设计理论与方法

Theory and Methods of Sampling Design in Statistical Survey

赵俊康 著

中国统计出版社
China Statistics Press

# 统计调查中的抽样设计理论与方法

## Theory and Methods of Sampling Design in Statistical Survey

赵俊康 著

中国统计出版社
China Statistics Press

**（京）新登字 041 号**

**图书在版编目（CIP）数据**

统计调查中的抽样设计理论与方法/赵俊康著.
－北京：中国统计出版社，2002.5
ISBN 7－5037－3727－1
Ⅰ.统…
Ⅱ.赵…
Ⅲ.抽样调查－设计…
Ⅳ.C811
中国版本图书馆 CIP 数据核字（2002）第 019955 号

统计调查中的抽样设计理论与方法

责任编辑/张美华
责任校对/沙美华
出版发行/中国统计出版社
通信地址/北京市西城区月坛南街 75 号　邮政编码/100826
办公地址/北京市丰台区西三环南路甲 6 号
电　　话/（010）63459084、63266600－22500（发行部）
印　　刷/科伦克三莱印务（北京）有限公司
经　　销/新华书店
开　　本/850×1168mm　1/32
字　　数/254 千字
印　　张/10.125
印　　数/1－3000 册
版　　别/2002 年 5 月第 1 版
版　　次/2002 年 5 月北京第 1 次印刷
书　　号/ISBN 7－5037－3727－1/C·1852
定　　价/21.00 元

# 前　言

抽样调查是国际上通行的统计调查的主要方法之一。抽样调查的成功与否，在很大程度上取决于抽样设计的好坏，抽样设计在抽样调查中占有举足轻重的地位。抽样设计是一项系统工程，涉及到对研究对象、研究目的、资料限制、抽样方法、经费预算、人员素质等诸多因素的分析研究。国内外现有的该领域的文献资料，不论是著作还是论文，都是就各种抽样或估计方法分别进行介绍和研究，没有对抽样设计从整体上进行系统地分析研究，不利于人们对抽样设计的整体把握和抽样调查的实际应用。面对不同的研究对象和众多的抽样估计方法，如何根据实际情况选择最适合的抽样估计方法或通过对抽样估计方法的改造、创新，使它适合于研究对象，是抽样设计人员面临着的最大难题。

本书从"设计"的角度对抽样调查的理论和方法进行系统研究，研究的重点是如何根据研究对象的特点和研究目的的要求，科学合理地进行抽样方案的设计，以达到用最小的成本获得最大的估计精度。主要研究内容包括抽样方法的设计、抽样单位的设计、抽样框的设计、估计方法的设计、辅助变量的设计、样本轮换的设计、多主题抽样的设计、样本容量的设计、敏感性问题调查的抽样设计、调查问卷的设计等。通过对这些问题进行系统全面的分析研究，从更高的层次上提炼出抽样设计的思想和方法，为抽样调查方案的设计提供一套较系统和完整的理论和方法论体系，以指导抽样调查实际工作。这对于推动该领域的理论研究和指导抽样调

查人员把抽样技术正确地应用于抽样调查实践都具有重要的意义。

本书在写作过程中，参考了大量有关抽样技术方面的文献资料，在此，对这些文献的作者致以真诚的谢意。

本书的出版，得到了山西财经大学有关领导的关心和支持，得到了山西财经大学统计学专业重点学科建设经费的资助，作者表示深切的感谢。

本书力求在研究角度、体系以及具体内容方面有一定程度的创新，但由于水平所限，在许多方面还未能如愿。书中难免有不妥甚至错误的地方，恳请读者批评指正。

作　者

2001 年 10 月 30 日

# 目　录

# 第一章　抽样设计概述

抽样调查是一项系统工程，为什么调查？调查什么？怎么调查？这些问题在实施调查之前，需要做出科学合理的安排，这种安排就是抽样设计。这一章从一些基本概念入手，讨论抽样设计的目标、准则、内容、步骤等一般性的问题，并对我国抽样设计中面临的问题和解决思路做一些分析，为后续内容的研究奠定基础。

## 第一节　抽样设计的意义

抽样调查的目的是用较少的费用取得较高的估计精度。抽样设计就是为抽样调查的实施提供一个指导性的文件，以实现抽样调查的目的和任务。

### 一、抽样调查的概念与特点

（一）概率抽样与非概率抽样

抽样调查是从研究总体中抽取部分单位进行调查，以此推断总体指标数值的一种统计调查方法。在抽样调查中，根据从总体中产生样本的方法不同，有概率抽样和非概率抽样两种不同的抽样方法。

概率抽样是按照随机原则从总体中抽取部分单位构成样本，以此推断总体数量特征。所谓随机原则是指在抽取样本时，总体

中每一个单位都有一个已知的、并且非零的被抽取的概率。在概率抽样中，如果总体中每一个单位被抽中（在一次或 n 次抽取中）的概率都相等，称为等概抽样；如果总体中至少有一个单位被抽中（在一次或 n 次抽取中）的概率与其它单位不相等，则称为不等概抽样。等概抽样下，样本的性质比较简单、不等概抽样的样本性质比较复杂，但在一定的条件下，具有较高的抽样估计精度。

不符合随机原则的抽样方法都属于非概率抽样。实际工作中，主要是在市场调研中采取的非概率抽样方法包括以下一些基本类型：

1.便利抽样。也称为任意抽样或偶遇抽样，是指调查者为了方便而任意抽取样本单位的一种抽样方法。比如，要了解某种商品使用的普遍程度，调查员在道路、商店、学校等场所随便选择部分人作为样本进行调查，这种调查称为"拦截式调查"，它是便利抽样的一种。

2. 判断抽样。选择对研究总体有代表性的单位进行调查，即样本是总体中的一些有代表性的单位，或叫做典型单位。如对一个包含单位数量少而单位之间差异大的总体，抽样者检查了整个总体，然后选一部分"典型单位"，即接近于他对总体平均数的印象的那些单位进行调查。如果判断比较准确，这种抽样方法较之于便利抽样可以提高样本的代表性。

3. 定额抽样。也称为"配额抽样"。将总体按若干标志分类，掌握总体中各类单位数所占的比例，并以此比例确定样本中各类单位的比例。然后，由调查者主观确定样本单位。这种抽样方法，较之于判断抽样加强了对样本结构在"量"的方面的质量控制，使样本结构更接近于总体结构，可以保证样本对总体有较高的代表性。定额抽样与概率抽样中的分层抽样有类似之处，不同点在于，分层抽样的样本单位是随机抽取的，而定额抽样中的样本单位是由调查者主观确定的。

4. 滚雪球抽样。它是通过使用初始被调查者的推荐来挑选另

外的调查者的抽样方法。有时候，对于一些少见的总体，也称为低发生率的总体进行调查，要找到这些个体需要付出很大的代价。比如，对患某种疾病的人进行调查，要全部找到这些人，再从中抽取样本，是相当困难的。如果事先掌握了符合条件的一部分人，由他们推荐来扩大样本单位，就相对容易的多了。

非概率抽样方法的优点是调查容易实施，可以大大节约调查成本，然而，这种成本的节约是以调查质量的降低为代价的。整个样本很可能有偏差，结果是样本可能不能很好地代表总体。

这里特别指出，根据我国的习惯叫法，抽样调查仅指概率抽样，因此，本书在后面谈到抽样调查时，如无特别说明，均指概率抽样。

（二）概率抽样的特点以及它们之间的关系

与全面调查和非概率抽样相比较，概率抽样具有三个基本特点：

1. 按照随机原则抽样。这是概率抽样区别于非概率抽样的根本特点，也是概率抽样科学性的根本所在。按照随机原则抽样，不仅排除了选择样本单位时主观因素的影响，使样本对总体有较高的代表性，更重要的是使估计量具有了某种已知的概率分布，为抽样推断提供了科学基础。

2. 用部分推断总体。这是抽样调查区别于全面调查的基本特点。抽样调查用总体的一部分单位作为样本，用样本数据推断总体的指标，不仅扩大了统计调查的应用范围，而且可以大幅度节约统计调查费用，加快统计调查的速度，提高统计调查的经济效益和时效性。

3. 抽样误差可以进行计算并加以控制。这也是概率抽样区别于非概率抽样的一个重要特点。抽样调查是用样本去推断总体，由于样本的分布与总体的分布不会完全一致，因此，抽样误差是不可避免的。但概率抽样由于按照随机原则抽取样本，使估计量服从某种已知的概率分布，从而不仅使抽样误差可以计算，而且可以根据需要把它控制在要求的范围之内。

概率抽样的以上三个特点是密切联系的。部分推断总体产生了抽样误差的可能，提出了计算和控制抽样误差的要求；而抽样误差的计算和控制是以估计量具有已知的概率分布为前提条件的；按随机原则抽样，使估计量具有了某种已知的概率分布（如正态分布），从而使计算和控制误差的要求能够得以实现。

非概率抽样也有计算和控制误差的要求，但是，由于它不能提供估计量的概率分布，因此，这种要求不能实现。

（三）概率抽样的优点及其局限性

概率抽样具有适用范围广、调查速度快、调查费用省和精度高等四个基本优点。

概率抽样具有广泛的应用领域。有些调查活动不能进行全面调查，只能采用抽样的方法进行。比如，产品耐久性或使用寿命的调查，调查活动对产品本身具有损耗性或者破坏性，只能使用抽样的方法进行调查。有些调查活动，虽然从理论上讲可以进行全面调查，比如人口调查、资源调查、社会经济活动的调查等，但在总体单位数量太大时，出于经济和时间等方面的考虑，也大多采用抽样方法进行调查。

概率抽样只调查总体中的一部分单位，而且在总体单位数量很大的情况下，用很小比例的调查单位就可以满足估计精度的要求，因此，相对于全面调查而言，在提高调查速度的同时，也节约了调查的费用。

概率抽样可以取得较高精度的估计结果。一般来讲，统计调查都会存在一定的误差。统计调查误差按照其来源可以分为两类：一是登记性误差，即从被调查者得到的数据与真实数据之间的离差；二是代表性误差，即被调查者与需要估计的总体的结构不一致，使估计的总体数据与实际的数据之间产生的离差。全面调查只存在第一类误差，抽样调查两类误差都存在。但由于抽样调查涉及的调查单位数量少，在调查经费一定的条件下，可以对调查员进行培训或挑选，提高调查员的素质，大大缩小登记性误差；同时，

通过科学的抽样设计，可以把代表性误差控制在可以接受的范围内。实践证明，只要科学设计、精心组织，抽样调查中两类误差的和就会小于全面调查中的登记性误差。

抽样调查也有一定的局限性，主要是在样本容量不大时，对全面认识总体内部结构有一定的困难。比如，在人口抽样调查中，如果样本中没有抽到患有某种疾病的人口，那么，这类人口占总人口的比例就得不到合理的估计。

（四）抽样调查与全面调查的结合应用

作为统计调查的两种基本方法，抽样调查与全面调查各有自己的优缺点。为了最大限度地发挥两种方法的优点，克服它们的缺点，实际工作中，往往把两种方法结合起来使用。抽样调查与全面调查有四种基本的结合模式。

1. 时间结合式。即在连续性的调查活动中，一些时间上进行全面调查，另外一些时间上进行抽样调查。比如在人口调查中，每隔一定时间进行一次全面调查，在两次全面调查之间，实施抽样调查。

2. 空间结合式。即对某些单位实施全面调查，对另外一些单位实施抽样调查。比如在工商业统计调查中，对大企业实施全面调查，对小企业实施抽样调查。

3. 项目结合式。即对某些项目实施全面调查，对其它项目实施抽样调查。比如，在残疾人调查中，对残疾人的数量、收入、致残原因进行全面调查，对其婚育状况、家庭人口、就业、年龄分布等进行抽样调查。

4. 用抽样调查资料矫正全面调查的登记误差。

## 二、抽样设计的作用

抽样设计是为抽样调查的实施提供一个指导性的文件，以实现抽样调查的目的和任务。它的作用体现在以下几个方面：

（一）使调查费用控制在预算范围之内

抽样调查是一项实践性工作，它的实施需要相应的费用做支持。一项抽样调查，事先都有经费的预算。如果调查过程中的费用开支突破预算，使调查的后期工作失去财力支持，就会使整个调查前功尽弃。抽样设计的一个作用就是根据预算经费的限制，合理地确定整个抽样调查各个环节上的费用，保证抽样调查的顺利实施。

（二）使调查误差控制在要求的范围内

抽样调查是用样本去推断总体，必然存在抽样误差，同时，还可能存在非抽样误差。概率抽样的一个基本优点是抽样误差可以控制在一定的范围内，它是通过抽样设计来实现的。抽样设计，通过对抽样方法的选择、样本容量的科学计算等，可以把抽样误差控制在要求的范围内。同时，抽样设计可以对抽样调查的组织与实施提出具体的要求和必要的措施，比如对抽样调查的组织机构、调查员的培训、数据的记录、整理、计算机录入等方面提出要求，最大限度地控制非抽样误差。

（三）使调查时间控制在要求的范围内

任何一项调查都有一定的时间期限要求，在规定的调查时间内完不成调查任务，不仅会增加调查费用，而且会使调查资料的价值大打折扣。抽样设计的一个内容是为调查提供一个具体的日程表，指导调查工作按预定的时间要求进行，保证在规定的调查期限内全面完成调查的各项工作。

## 三、抽样设计的目标与准则

抽样调查的优点是建立在科学的抽样设计基础之上的，没有科学的抽样设计作基础，抽样调查的优越性就得不到发挥，甚至会给决策带来负面影响。

抽样设计的目标是实现抽样调查的低成本与高效率。低成本，就是使抽样调查的费用达到最小，高效率就是使抽样调查的估计误差达到最小。为了实现这个目标，抽样设计必须遵循以下准

则：

（一）目的性

整个抽样设计，包括抽样框的设计、样本容量的设计、抽样方法的设计、估计量的设计、调查方法与问卷的设计，都必须以研究目的为依据，服从并服务于研究目的，离开研究目的，抽样设计就失去了方向。

（二）实践性

它指的是能否基本上按预定的设计完成调查。在抽样设计阶段，必须考虑到实施的可行性。

（三）经济性

即用最小的费用实现调查目标。

（四）可度量性

它是指设计能从样本自身计算出有效性的估计值，这个估计值通常是用估计量的方差的估计值来表示的。非概率样本不具有可度量性，概率样本也不自动保证可度量性。比如，在分层抽样中，每层只抽取一个样本单位，或在整群抽样中只抽取一群，这些抽样设计就不具有可度量性，应尽量避免这样的抽样设计。

以上准则往往互相冲突，设计人员必须在四者之间进行权衡，以得到一个好的抽样设计。

## 第二节　抽样设计中的基本概念

抽样设计研究如何用样本数据估计总体目标，构造合适的估计量是实现这一目标的基本手段。而估计量构造的好坏，是通过估计误差的大小来衡量的，衡量的科学基础是估计量的抽样分布。因此，总体、样本、估计量、抽样分布等就构成了抽样设计的一些最基本的概念。深刻理解和熟练掌握这些概念，对抽样设计者来讲是必须的。

## 一、总体

### （一）总体及其数学描述

在抽样调查中，把总体定义为研究对象中所有单位在某一标志上的标志值的集合。包括两个基本构成要素，一是单位，二是标志值。用数学方法表示为：

$$\{y_1, y_2, \cdots, y_N\}$$

其中，$y$ 代表标志值，下标表示单位序号。

### （二）总体中的基本单位与抽样单位

抽样调查中，把构成总体的"单位"分为基本单位和抽样单位两类。基本单位是由研究目的所决定的单位，如研究全国农村住户情况，则"农村住户"为基本单位；抽样单位是抽取样本时的单位，比如，在以农户为基本单位的总体中，可以以"农户"作为抽样单位，也可以以"乡"或"县"作为抽样单位。当总体中的基本单位数不是太大，且有完备的基本单位名单时，就以基本单位作为抽样单位；若基本单位数量太大，或没有完备的基本单位名单，就必须另外确定抽样单位，此时，抽样单位和基本单位就不一致了。如森林材积量调查中，基本单位是"每一棵树"，抽样单位可能是"亩"。抽样调查中的单位通常是指抽样单位，在不引起混乱的情况下，简称为总体单位。

### （三）目标总体与被抽样总体

抽样调查中，把总体划分为目标总体与被抽样总体。目标总体是指所要研究的总体，比如，要研究全国农村住户情况，"全国农村住户"就构成一个目标总体。被抽样总体是从中抽取样本的总体。当目标总体有完备的抽样单位名单时，二者是一致的，有时候，目标总体没有完备的抽样单位名单，或编制这样的名单比较麻烦，只能从与目标总体相近，且具有抽样单位名单的总体中抽样，此时，二者会产生差异。比如，研究全国 7－12 岁儿童的身体情况，为了方便，从小学 1－6 年级的学生花名册中抽样，此时，二者

就有差异。抽样估计的结论只说明被抽样总体，能否说明目标总体，取决于两种总体的接近程度。为了实现抽样目的，要求二者尽量一致。把被抽样总体中的抽样单位按照一定顺序排列起来形成一个抽样单位的名录，这个名录叫做抽样框，实施概率抽样，必须编制抽样框。

（四）总体分布

总体中某一标志上的各个标志值及其对应标志值上总体单位数所占比重。例如：有一总体{2,2,5,7,9}，其总体分布如表1-1所示。

表1-1　一个总体分布例表

| $y_i$ | $P(y_i)$ |
|---|---|
| 2 | 2/5 |
| 5 | 1/5 |
| 7 | 1/5 |
| 9 | 1/5 |

（五）目标量

抽样调查中所要估计的总体指标称为目标量。通常以与总体单位标志相对应的大写字母表示。常用的目标量有：总体均值 $\bar{Y}$、总体总值 $Y$、总体比例 $P=A/N$、总体中具有某一特征的总体单位数 $A$、总体比率 $R=\bar{Y}/\bar{X}$、总体方差 $S^2$、$\sigma^2$。

## 二、样本

（一）样本及其数学描述

样本是从总体中用某种方法抽取的部分单位在某一标志上标志值的集合。用数学方法表示为：

$$\{y_{1i},y_{2i},\cdots,y_{ni}\}$$

其中，$y$ 表示样本单位的标志值，下标1、2、…、$n$ 表示样本单位序号，$1\leqslant i\leqslant N$ 是一个随机变量。为了简洁，通常把样本简记为：

$$\{y_1,y_2,\cdots,y_n\}$$