第一章　广义抽样调查概论

纵观世界统计发展史，人们会发现，在统计学的发展过程中已经出现了两次重要的革命：

第一次是以抽样调查为代表的认识事物数量特征的方法论革命，主要发生在18世纪末到19世纪中叶。当时的抽样调查主要是指概率抽样调查，应用范围主要在技术和工艺领域。

第二次是以国民经济平衡为代表的认识事物关联性规律的方法论革命，即国民经济核算方法的产生。这为我们从事物相互之间的有机联系上来全面认识研究对象的数量关系提供了理论基础，主要发生在20世纪30年代。

其实我们有理由相信，世界统计发展史的第三次革命将会出现，那就是以现代高新技术为支撑的信息技术、空间技术和网络技术在信息搜集、信息传输、信息分析、信息集成等具有统计特征和技术含量的领域的应用，这必将引起统计专业的新的革命，目前已经初见端倪。比如：空间遥感信息观测技术在统计中的应用，经济地理的系统信息集成及微区位理论，主观意向快速公正的调查分析（CATI），大系统多变量的综合评价与监测，投入产出分析技术等，无一不与高技术、信息化、大容量发生联系。这已经昭示，统计作为现代信息技术的核心将发挥重要作用，统计的专业性、广泛性、渗透性、应用性必将充分显示出来。

“统计思维有一天会像读写能力一样成为有能力的公民身份的象征。未来是统计的时代”——韦尔斯。

第一节　广义抽样调查的产生

一、抽样调查产生的渊源

（一）抽样调查产生的渊源

统计在绝大多数情况下都是依据小量数据（样本）所提供的数据以估计预测某研究对象总体的方法。面对不确定情况，统计学为决策制定提供了科学的方法。目前我们所说的抽样调查一般是指随机抽样法条件下的狭义抽样调查，

属于统计推断的一个主要内容而存在。统计推断包括参数推断、非参数推断。统计推断主要产生于对于不确定现象和大系统现象的求知欲，可以追溯到1662年英国数学家J. Graunt在调查伦敦市死亡人数时所使用的方法，这是统计史上最早出现的统计推断技术——抽样调查技术。

从此，社会学家应用统计推断技术研究社会问题；经济学家使用统计推断研究经济问题；自然科学家应用统计推断研究自然科学的若干技术领域；管理者应用统计推断-抽样调查校验各种工艺技术、各种产品质量、各种管理模式的优越性、稳定性和规范性等。

1899年，高尔顿的著作*Nature Inheritance*的出版，标志着现代统计学的诞生。而杰出的统计学家卡尔·皮尔逊还因此书而对统计学产生了浓厚的兴趣，此前，皮尔逊只是伦敦大学的数学教员。当时，是“所有知识都基于统计”这一想法吸引了他。1890年，他在格里辛学院讲授的课程是“现代科学的范围与概念”，也正是这一课程使他越来越强调科学定律的统计基础，最后竟完全致力于统计理论研究。经他的推广，人们越来越深信统计资料的分析能为许多重要的问题提供解答方法。1908年，爱尔兰贵尼斯酿酒厂的一位统计工作者威廉司·来·哥司特，第一个注意到小样本，并认识到从小样本导出可靠数据的重要性及可能性，进而创立了以小样本代替大样本的方法，并以笔名“学生”（Student）发表了许多关于以抽样得来的资料解释总体的文章，也就是现在常用的“t分布”。这一方法不久便为英国剑桥大学教授朗奈德·阿·费暄及其同事所推广，这给抽样统计的发展以及估计抽样带来误差的计算提供了坚实的理论基础。

（二）抽样调查应用的发展

显然，上述提到的抽样调查指的是概率抽样调查。然而随着人们对社会、经济、科技、自然的认识的深入，人们对抽样调查的认识也得到了提高。比如：人口抽样调查、产品质量抽样调查、农产量抽样调查、医学抽样调查、经济统计抽样调查、社会统计抽样调查、民意调查、市场调查、自然科学实验性设计、预警与监测（景气指数）等，进而逐渐形成了比较成熟的、包括一切非全面调查组织形式的、以事物的部分特征来推断总体特征的广义抽样调查方法。

二、广义抽样调查的提出

（一）广义抽样调查的概念

广义抽样调查是指一切使用部分调查结果推算总体特征的非全面调查的调

查组织形式和方法的总称。有时我们也简称为广义抽样、抽样调查（下同），包括概率抽样调查和非概率抽样调查。以后在没有特别注明的情况下，我们所说的抽样调查均指广义抽样调查。

（二）广义抽样调查的种类

广义抽样调查的基本方法体系如表 1.1 所示。

表 1.1　广义抽样调查方法体系分类表

	一级分类	二级分类	三级分类调查
广义抽样调查	概率抽样调查	单一概率抽样调查	纯随机抽样调查
			机械抽样调查
			分层抽样调查
			整群抽样调查
		复合概率抽样调查	阶段抽样调查
			组合抽样调查
	非概率抽样调查	单一非概率抽样调查	重点抽样调查
			典型抽样调查
			方便抽样调查
			判断抽样调查
			配额抽样调查
			雪球抽样调查
		复合非概率抽样调查	目录抽样调查
			随机模型抽样调查
			关联抽样调查
			捕获-再捕获抽样调查

（三）广义抽样调查的特征

由于广义抽样调查是从研究对象的总体中抽取一部分单位作为样本进行调查，并据此推断有关总体的数字特征，因此，抽样调查作为一门技术，具有经济性好、实效性强、适应面广、准确性高的特征。

鉴于广义抽样调查是根据部分实际调查结果来推断总体标志总量的一种统计调查方法，因此属于非全面调查的范畴。它是按照科学的原理和计算方法，

从若干单位组成的事物总体中，抽取部分样本单位来进行调查、观察，并用所得到的调查标志的数据以代表总体，推断总体。

与其他全面调查一样，抽样调查也会遇到调查的误差和偏误问题。抽样调查的误差通常有两种：一种是工作误差（也称登记误差或调查误差），另一种是代表性误差（也称抽样误差）。但是，抽样调查可以通过抽样设计，通过计算并采用一系列科学的方法，把代表性误差控制在允许的范围之内；另外，由于调查单位少，代表性强，所需调查人员少，工作误差比全面调查要小。特别是在总体包括的调查单位较多的情况下，抽样调查结果的准确性一般高于全面调查。因此，抽样调查的结果是非常可靠的。

（四）狭义抽样调查

1. 狭义抽样调查

狭义抽样调查是指按照随机原则从总体中抽取一部分单位进行调查，用这一部分单位的调查的数量特征的结果，在一定的概率保证下对于总体的数量特征进行推断和测算的调查方法。即，概率抽样调查。

2. 概率抽样调查的特点

特别地，当实施一个优秀的概率抽样调查时，其抽样调查的效果往往优于全面调查。概率抽样调查数据之所以优于全面调查并能用来代表和推算总体，主要是因为概率抽样调查本身具有其他非全面调查所不具备的特点，主要是：

（1）调查样本是按随机原则抽取的，在总体中每一个单位被抽取的机会是均等的，因此，能够保证被抽中的单位在总体中的均匀分布，不致出现倾向性误差，代表性强。

（2）以抽取的全部样本单位作为一个“代表团”，并用整个“代表团”来代表总体，而不是用随意挑选的个别单位来代表总体的。

（3）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，因此在调查样本的数量上有可靠的保证。

（4）抽样调查的误差，是在调查前就可以根据调查样本数量和总体中各单位之间的差异程度进行计算的，并控制在允许范围以内，调查结果的准确程度较高。

基于以上特点，概率抽样调查被公认为是其他非全面调查方法中用来推算和代表总体的最完善、最有科学根据的调查方法，概率抽样的结果也明显优于非概率抽样。然而，由于实际中的调查没有一个能严格匹配于经典教科书的概率抽样方法，因此非概率抽样（被抽象化地称为模型抽样，这里的“模型”是

“随机”的对称）能够很好地辅助概率抽样调查。比如，概率抽样中的无回答问题要通过满足一定假设的模型才能获得解决，而模型体现的是一种主观的假定，一种非概率化的操作手段，因此从某种意义上来说，现在大部分的调查均是随机加模型的混合模式。

概率抽样调查是经济、管理、自然科学、产品质量分析和市场调查中最常用的调查方式，而非概率抽样调查在市场调查、社会经济调查中使用更加频繁。

三、调查误差分析

一般来讲，观察值不等于真值时即存在误差。但统计研究具有大量性的特点，它不去研究个体单位经观察或测量而产生的误差，而是从大量的平均角度加以探讨。因此可以说，统计调查误差（error of statistical investigation）是指调查结果所得的统计指标与调查总体指标之间的差异。

统计调查误差从主体影响上分析有主观误差和客观误差之分。主观误差是指人们的主观故意或者倾向性认识带来的误差。比如有意识地虚报、瞒报、拒报、迟报、伪造、篡改等行为。客观误差是指由于技术水平、客观条件以及从个体或者部分推算总体时存在的代表性误差等。客观误差一般有三种类型：登记性误差、系统性误差和代表性误差。

1．登记性误差

登记性误差是由于调查过程的各个环节的工作不准确而造成的。如：计量、登录、计算等错误，但这绝不是指故意性行为。有意识地虚报、瞒报、拒报、迟报、伪造、篡改等行为属于违法行为，这是不允许的，所造成的误差不属于登记必误差不属于登记性误差的范围。登记性误差在全面调查和非全面调查中都存在。为了保证统计资料的准确性，应采取措施尽量避免和减少登记性误差。

2．系统性误差

系统性误差是指由于客观条件的影响而出现的趋势性误差。比如，测量仪器的偏差、个人主观思想的影响、对于颜色、声音、亮度等的偏好和角度等的影响而出现的误差。这种误差一般可以通过改变客观条件进行修正和克服。

3．代表性误差

代表性误差，是指用部分调查单位的统计资料计算出的指标值（样本统计量）来估计总体指标值（总体参数）所产生的误差。如：抽样调查中，由随机样本观察值计算出的样本平均数 x 与总体平均数 μ 之间的差异、样本成数 p 与总体成数 P 之间的差异等，都是代表性误差。可见，这种误差是绝对存在的，

是不可避免的，且只出现在非全面调查之中。抽样误差愈小，样本指标与总体指标差异愈小，估计精度愈高，样本的代表性愈强，这也是代表性误差名称的由来。抽样调查中产生的代表性误差也叫抽样误差。虽然其不可避免，但可以进行计算，并通过合理的试验设计，将其控制在一定范围之内，进而达到我们所要求的抽样估计精确度。正是因为这个原因，抽样调查在整个调查方法体系中占据主导地位。

由于全面调查中只存在登记性误差，而非全面调查中既存在登记性误差，也存在代表性误差，那么，是否非全面调查的误差就一定比全面调查的误差大呢？回答是否定的，这是因为当全面调查的总体单位特别多时，所产生的登记性误差也会随之增大。

第二节 广义抽样调查的理论基础

广义抽样调查的理论基础是由概率抽样调查和非概率抽样调查这两种类型的特征所决定的。也就是说：概率抽样调查和非概率抽样调查各自具有不同的理论基础。

一、概率抽样调查的理论基础

（一）概率抽样调查的理论基础

对概率抽样调查而言，其理论基础是概率论和数理统计。

1. 概率抽样调查的学科性质

概率抽样又称随机抽样，它是一种非全面调查的主要组织形式，是按照随机原则从调查对象中抽取一部分单位作为样本进行观察，然后依据所获得的样本数据，对调查对象总体的数量特征作出具有一定可靠程度的估计和推算。

概率抽样所依据的概率原理属于数理统计学的一个重要分支，也是现代统计学的基础。抽样的方法不仅对统计推断、统计检验以及统计决策等理论的发展产生了直接的影响，而且还构成了其他应用性学科如计量经济学、管理会计学等的方法论基础。

2. 概率抽样的基本原则

概率抽样依据的基本原则是随机原则。所谓随机原则就是在抽选调查单位

的过程中，完全排除人为的主观因素的干扰，以保证使现象总体中的每一个个体都有一定的可能性被选中。换句话讲，哪些单元能够被选作调查单位纯属偶然因素的影响所致。这里需说明几点：（1）随机并非“随意”。随机是有严格的科学含义的，可用概率来描述，而“随意”仍带有人为的或主观的因素，它不是一个科学的概念；（2）随机原则不等于等概率原则；（3）随机原则一般要求总体中每个单元均有一个非零的概率被抽中；（4）抽样概率对总体参数的估计有影响。同时，按随机原则抽样可以保证被抽中的单元在总体中均匀分布，不致出现系统性、倾向性偏差；在随机原则下，当抽样数目达到足够多时，样本就会遵从大数定律而呈正态分布，样本单位的标志值才具有代表性，其平均值才会接近总体平均值；按随机原则抽样，才可能实现计算和控制抽样误差的目的。因此，随机原则是抽样调查所必须遵循的基本原则。

3. 概率抽样调查的理论依据

概率抽样调查的理论依据主要有概率统计中的大数定律和中心极限定理。

（1）大数定律。

大数定律又称平均数定律或大数法则，它所描述的是当样本充分大时，样本统计量的极限行为。即在充分大规模的抽样下抽样平均数和总体平均数间的离差可以为任意小这一可能性的概率可以尽量接近于1，即接近完全的精确性。大数定律可以用切比雪夫定理加以证明。若从逻辑意义、哲学意义来阐明的话，它是大量现象和过程的规律性，而且一般只有在充分大量观察时，才会显露出现象和过程在某种具体历史环境中具有代表性的主要特征。

大数定律的具体表现如下：

第一，只有掌握足够多的单位数目或足够多的情况时，大量现象的规律性及大量过程的倾向性才能很好地显示出来。也就是说，只有在掌握足够多单位数目或足够多的情况时，对这些大量现象和过程，才能很好地进行研究。

第二，只有在平均数形式上，这些规律性与倾向性才能被表现出来。正因为如此，大数定律又称为平均数定律。

第三，研究大量现象和过程时，如果抽取更多的单位，那么从这些单位的标志值所计算出来的平均数越能够正确地表现出这种现象或过程的规律性。

第四，如果我们研究足够多的单位数目或足够多的情况，以平均数为中心，各个单位或情况向正反两方向的离差往往互相均衡化，或者互相抵消。对大量现象或过程来说，这些离差当然不是由于本质的差异所引起，而是由于偶然的状态所发生的。

大数定律的理论和方法，对科学地安排统计试验和制定抽样调查方案是十分重要的，它使抽样法的应用获得充分的数学依据，同时也为抽样结果的精确推断提供了充分的可能性。所以说，大数定律是统计抽样调查的数理基础，也给统计中的大量观察法提供了理论和数学方面的根据。因此它要求在运用抽样调查时，必须注意：① 遵循随机原则。只有在随机原则下进行抽样，样本中各单位才能均匀分布在总体中，使样本具有代表性，这样，样本指标才可以用来对总体指标作出估计和推断。② 抽样必须注意观察现象的大量性。在同一总体中进行随机抽样，每个被抽中的样本单位的标志值或偏大或偏小，纯属偶然，并不代表总体的数量特征，而通过大量观察，根据大数定律的原理，消除偶然因素的影响，用抽出的单位组成样本综合的结果，才能把总体的数量特征接近准确地反映出来。

（2）中心极限定理。

中心极限定理的基本内涵是：一组独立同分布的变量的和或平均值当 n 充分大时近似地具有正态分布。它分别由德莫佛尔-拉普拉斯和林德伯格-勒维所证明。以下仅对德莫佛尔-拉普拉斯中心极限定理做一简单介绍。

设随机变量 $Y_1, Y_2, \cdots, Y_n$ 相互独立，服从同一分布，且有有限的数学期望值 μ 和方差 σ^2，又使 $\overline{Y}_i = \dfrac{Y_1 + Y_2 + \cdots + Y_n}{n}$，则随机变量 $Y_n = \dfrac{\overline{Y}_i - \mu}{\sigma/\sqrt{n}}$ 的分布函数 $F_n(y)$ 对于任意 y，就有：

$$\lim_{n\to\infty} F_n(y) = \lim_{n\to\infty} P\left\{\frac{\overline{Y}_i - \mu}{\sigma/\sqrt{n}}\right\} = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t$$

通过这个定理可以知道，不论总体服从什么分布，当 n 很大时，样本的平均数 $\overline{Y}$ 近似于具有参数 μ 和 $\sigma/\sqrt{n}$ 的正态分布（即极限正态分布）。这个定理是大样本统计推断的理论基础。中心极限定理，并非证明正态分布的存在，而是用来说明近似地遵从正态分布的概率变量的现象，说明样本平均值的分布接近于正态分布。中心极限定理表明：样本平均值分布的平均值等于总体平均值，即 $E(\overline{Y}) = \mu$；样本分布的标准差为 $\sigma/\sqrt{n}$。中心极限定理说明，用样本平均值产生的概率来代替从总体中直接抽出样本来计算的抽取样本的概率，为抽样推断奠定了科学的理论基础。

（二）概率抽样调查的误差分布理论

概论抽样调查的目的是把对总体中有限的部分单位的调查结果作为普遍适

用于总体的估计和推断。但是，样本是随机抽出的，不同的随机样本就会得出不同的估计量。在同一总体中往往可以抽出多个样本，可以得到同样多的估计量，基于总体指标都存在或大或小、或正或负的偏误，因此，用样本指标来推断总体指标，就存在抽样误差。承认这一点，不是证明抽样调查不准确，不能用来推断总体，而相反，正是利用可能发生的抽样误差，加上样本指标，来推断在多大的概率度下总体指标在一个怎样的范围之内。

18 世纪末，法国数学家拉普拉斯与德国数学家高斯研究误差分布，建立了误差分布理论。

在一个既定的总体中，抽选一定含量的样本，可能抽选到的样本有多个，因此可以取得多个可能的样本指标（主要指平均数和成数）。如果将所有可能的样本指标组成频率分布，可发现样本指标 $\bar{y}$ 愈接近总体指标 $\bar{Y}$ 的可能样本数愈多，频率愈大；偏离 $\bar{Y}$ 愈远的可能样本个数愈少，频率愈小，即形成两端小中间大的 $\bar{y}$ 可能值的分布，同时也是抽样误差的分布。按正态分布的基本条件，可能样本指标的分布从理论上说是遵循正态分布的。样本指标的分布，通常又叫抽样分布。数理统计已证明，可能样本指标是否严格遵循正态分布，由两个主要条件所决定：一是抽样总体的分布形态，二是抽样数目的大小。如果样本抽自正态总体，无论抽样数目是大是小，可能样本指标都是遵循正态分布的；如果样本抽自非正态总体，只要抽样数目较大（$n>30$），可能样本指标也是接近或遵循正态分布的。

认识抽样误差及其分布的目的，就是希望所设计的抽样方案所取得的绝大部分的估计量能较好地集中在总体指标的附近，通过计算抽样误差的界限，使抽样误差处于被控制的状态。

（三）概率抽样调查理论的广泛应用

由于发展和论证了大数定律、中心极限定理和误差分布理论，抽样方法才有了科学的依据。同样，在抽样调查中也广泛应用了概率理论。

概率也称或然率，是指某一事件可能发生的机会，也就是某个事件可能发生的次数与所有可能发生事件总次数之比。等概率就是机会均等，不等概率就是机会不均等。概率通常有古典和统计两个意义：

（1）古典意义，就是事物有有限个均等的可能结果。如掷一粒骰子，有六个有限的均等可能的结果，如预定可能结果为任意一个点数，则实现任一点数的机会均为 $\frac{1}{6}$。古典概率由于受“结果有限”和“均等可能”的限制，在实践

中有很大的局限性。

（2）统计意义，就是用统计的频率作概率的近似度量。如在某产品总体中，合格品 m 占被抽检品 n 的频率为 $\frac{m}{n}$，于是可用频率 $\frac{m}{n}$ 作为任意抽检一件产品结果为合格品的概率的近似值，是统计意义上的概率。

二、非概率抽样调查的理论基础

（一）非概率抽样与模型抽样

1. 模型抽样

模型抽样（model sampling）是基于对调查总体中变量分布的广泛假设而采取的一种抽样方法。对变量分布的这种广泛假设与概率抽样中依据的严格假定形成鲜明对照，通常称为超总体（superpopulation）假设。

2. 非概率抽样

非概率抽样在抽样时不遵循随机原则，而是按照研究人员的主观经验或其他条件来抽取样本的一种抽样方法。

非概率抽样并没有严格的定义，它也有许多不同的抽样方法，它们的一个共同特点就是用一种主观的（非随机的）方法从总体中抽选单元。由于不需要完整的抽样框，非概率抽样是一种快速、简单且节省费用来获取数据的方法。使用非概率抽样的问题是，我们不清楚能否通过样本对总体进行推断，原因是用非概率抽样从总体中抽选单元的方式可能会导致较大的偏差。例如，在非概率抽样中调查人员经常主观地决定哪些单元入样。由于调查人员往往倾向于选择总体中那些最容易接触到的和最友好的单元，使总体中很大一部分单元完全没有被抽中的机会，而这些单元与被抽中的单元之间可能有系统差异。非概率抽样不仅会使调查结果出现偏差；另一方面调查人员也可能有意选择具有平均特征的那些单元，从而有排除极端值的倾向，这将减少总体中明显的变异性。相反，概率抽样能通过随机抽选单元而避免这种偏差。

由于非概率抽样不是按照随机原则的方式抽取样本，也就是说在抽样时，总体单元的入样概率事先未知，入样与否与研究人员的经验和主观意志有很大关系，因此，非概率抽样在应用时更需研究人员具备深厚的背景知识与相关经验。

3. 非概率抽样的产生及发展

从整个抽样调查的历史来看，非概率抽样先于概率抽样而广泛应用。20 世纪初，挪威统计学家凯尔（A. N. Kiar）竭力提倡并推广抽样调查，他认为普查项目不可能很仔细，提出了所谓的代表性（representative）样本调查，并在退休金和疾病保险金的调查中取得了较好的效果。结果在随后的一段时间内，目的性抽样（当时代表了非概率抽样）始终占据着主导地位，这种情况直到 1934 年内曼（J. S. Neyman）发表那篇被称为奠定了抽样理论基础的论文后才告结束。在这篇划时代的论文中内曼解释了概率抽样相对于目的性抽样的诸多优势，一个经典的例子就是，1921 年意大利的人口普查按照他建议的对大约 1 400 个行政区进行分层选样而不是按目的判断选样选出 29 个大区的办法，同样的，抽样比例（15%）也更有力地证明了这一点。此后 20 多年，概率抽样理论获得了广泛的发展，世界主要的统计机构都青睐于概率抽样，首部抽样教科书也在 1950 年前后面世。

概率抽样的结果明显优于非概率抽样。然而，由于实际中的调查没有一个能严格匹配于经典教科书的概率抽样方法，因此非概率抽样（被抽象化地称为模型抽样，在此“模型”是“随机”的对称）能够很好地辅助概率抽样调查。比如，概率抽样中的无回答问题要通过满足一定假设的模型才能获得解决，而模型体现的是一种主观的假定，一种非概率化的操作手段，因此从某种意义上来说，现在大部分的调查均是随机加模型的混合模式。

非概率抽样的应用越来越受到重视，通常是出于下述几个原因：

（1）严格的概率抽样几乎无法进行（在很多情况下如此）。例如，调查对象的总体边界不清而无法制作抽样框。此外有些研究为了更切合研究目的，不得不按照需要从总体中抽取少数有代表性的个体作为样本。

（2）为了保证随机的原则，对抽样的操作过程要求严格，实施起来比较麻烦，费时费力，因此如果调查的目的仅是对问题的初步探索，或是为了获得今后研究的线索，或是为了提出假设，而不是由样本推论总体，采用概率抽样就不一定是必需的。

（3）调查对象不确定或者根本无法确定。例如，对某一突发（偶然）事件进行现场抽样调查等。

（4）总体各单位间离散程度不大，且调查有关人员具有丰富的抽样调查经验。这样即便是非概率样本，仍然可以从经验判断意义上进行推论，经验丰富老道的专家进行的经验判断意义上的推论往往不亚于统计学意义上的推论。

非概率抽样一般具有操作方便、省钱省力的特点，统计分析方面通常也比

概率抽样简单，而且若能对调查总体和调查对象有较好的了解，非概率抽样照样可获得相当的成功。但是需要注意的是，非概率抽样由于每个总体单元进入样本的概率是未知的，而且由于排除不了调查主体的主观影响，因而无法说明样本对总体的代表性误差究竟有多大。毫无疑问，这种误差有时相当大，又无法估计，将给整个研究带来很大困扰。因此，采用非概率抽样方法获得的数据一般不能计算抽样误差，也不能从概率的意义上控制误差并以此来保证推断的准确性。但是在操作上，如果非概率抽样的具体抽样方法与某些概率抽样方法的抽样过程差异不大（或者说只存在理论上的差异），那么非概率抽样得到的样本就可以以“概率化”的方法进行研究与推断。以下我们将尝试讨论几种具体的非概率抽样方法。

4. 非概率抽样的理论基础

非概率抽样的各种具体方法就是基于这种广泛假设的，所以非概率抽样应归入模型抽样的范畴。关于便利抽样，人们假设对于所关注的变量特征而言，各个达到一定规模的样本之间差异很小，因此随意抽选的单元和按随机原则抽选的单元一样都能代表总体。而在判断抽样中，研究人员含糊和含蓄地认为挑选出来的是“典型的项目”，均具有充分的代表性。关于配额抽样，要研究的特征变量在总体结构中的分布是均匀的或是随机的，因此配额设计能够很好地体现总体的结构特征。

这些不同类型的模型抽样都同样依赖于超总体假设的有效性。在严格的概率抽样中，对总体的推断也完全可以通过统计方法实现，而不需要对有关总体做出服从某种随机分布的假设。事实上，在概率抽样中，总体的每一个单元都有一个已知的非零中选概率。

在非概率抽样中，是人为控制了总体的一些变量，而假定未控或不可控的变量在总体中则是随机分布的（自动随机化）。这种总体自动随机化现在仍然是一种不言而喻的当然情况，通常似乎没有理由抱有怀疑。尽管对于样本的抽选是不一致的，也可察觉不受控制的差异性确实存在，但人们总是希望这种差异在总体中是随机分布的，这种愿望可以用下述说法来表达：“没有任何理由来怀疑随机性。”但原因的不明必将导致理由的缺乏，却不能证明理由不存在。正如非参数统计中无先验信息的情形，就是假定参数的分布是均匀的一样。恰像基什（Kish）所指出的那样：“很多卓有成就的学科（例如物理、化学和生物等）的巨大进步都没有用到概率抽样，在这些领域里，推断是根据对总体有着适当的、自动的随机化这一判断而做出的。这样不仅未导致产生偏倚，而且提供了

模型抽样的成功案例。自然科学研究里尤其具有很多这种基于总体天然随机化假定而获得成功的例子。”

（二）非概率抽样的效度与信度

信度与效度的概念来源于测量理论。所谓测量就是对我们所观察的现象赋予一定的数值。测量理论假设，分析人员所关注的表达某一现象的概念通常都是不能直接进行测量的，总是需要通过操作化定义所说明的指标来进行间接测量。操作化是指一种陈述，告诉分析人员一个概念如何来测量，指标是指观测值的一个集合，是应用操作化定义得到的结果。指标与概念的联系可以通过式（1.1）表示：

$$\text{指标} = \text{概念} + \text{误差} \tag{1.1}$$

一个好的指标只有很小的误差，而一个不好的指标与对应的概念之间只存在微弱的联系。通常使用效度和信度这两个技术指标对测量手段或测量工具进行评估。

在抽样调查中，调查结果的效度问题与问卷设计联系紧密，来自效度较高的问卷设计的研究结果其效度一般也较高。而调查结果的信度则主要取决于计量方法和抽样程序的效度。好的计量方法是内部有效性的关键因素，而依据一定抽样程序得到有代表性的样本则是外部有效性的关键。就非概率抽样调查结果而言，其信度的主要威胁，一是在于抽样程序中存在的主观性，二是在于非定量化的计量方法的低估计精度。前者注定非概率样本必然存在选择偏倚（selection bias），从而并不似概率抽样可以精确推断，后者则导致非概率抽样不能提供一个估计推断精度的量化指标，如标准误差、方差等。

对样本选择偏倚和估计精度这两个方面的相关影响，至今的认识有：

1．样本选择偏倚

首先好的样本应尽可能地远离选择偏倚。当目标总体与抽样总体不一致，即二者出现交叉现象时，选择偏倚就会出现。如果调查研究居民家庭收入时忽略了暂住人口，那么调查估计的结果（不管是均值还是中位数）就有可能偏大。便利抽样通常也是有偏的，因为易于选择、可能回答的有时候并不能代表那些难于选择、不愿回答的总体单元。下面是通常会发生选择偏倚的一些情景：

（1）使用便利抽样获取的样本。

调查者以便利抽样取得青少年的一个样本，研究他们与父母、老师谈论对AIDS看法的频繁程度。但是那些愿意与调查者分享观点的青少年也必然更可能

与他们的父母、老师谈论此事，如果仅仅把他们谈论 AIDS 的时间平均起来就代表他们谈论 AIDS 的频繁程度，显然会高估真实值。

（2）判断取样获得的样本。

调查者会默认一套与他们感兴趣的一些特征相关的样本选择程序，有目的、“故意地”选择有代表性的样本。如果要估计购物商场中的顾客的花费，通常你会选择那些看起来花钱中等的人作为调查对象，而不敢去调查那些买了很多东西的人，这是你的心理在作怪!

（3）抽样框的不完善。

抽样框中没有完全包括目标总体，或者包括了非目标总体的单位，都会造成抽选偏倚。在我国的城镇住户调查中，经常会出现抽样总体与目标总体不一致的地方。

【例 1.1】考虑在重庆某地区进行民意调查时，如何通过电话调查了解居民的倾向性意见和建议。

解 有很多因素导致了目标总体和抽样总体的不一致，从而造成抽样调查产生抽选偏倚。并不是所有的住户都拥有电话，因此目标总体中可能会有倾向性意见和建议的一部分人口在抽样框中并没有体现；即便在有电话的住户中，没有常住居民的注册登记的居民也没有资格接受调查；抽样框总体中符合条件的人口也可能因为无法触及、拒绝接受调查或者不能回答而不能回应调查。如图 1.1 所示。

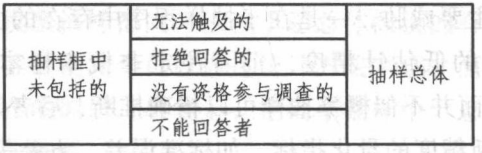

图 1.1 抽选偏倚的产生

（4）在抽样过程中，用方便的单位代替指定的但不易调查的单位。

概率抽样的设计没有过多考虑现实的环境，很可能选定的抽选单位是不易接近的，或是费用极其昂贵的。比如说，在调查当中指定的回答者不在家，调查员就可能会尝试用一下户来代替，严格来说这是概率抽样所不允许的，但无奈在实地调查中经常发生。补救的措施只有制定更细的抽样方案再加上对调查员进行更严格的培训。

（5）无回答。

即便是调查设计能够精准控制其他选择偏倚并使之达到最小，无回答还是歪曲了很多调查的结果。无回答者之间的差异性是很大的，而且差异程度是未

知的，除非你事后能得到无回答者的有关信息。报纸或新闻期刊中的大多数调查的回答率都很低，有的甚至只有 10%，把这样的结果推广到总体当中是很难有说服力的。

从上面可以看到，抽样设计、抽样程序、实地工作都会成为抽选偏倚的来源，要完全控制它，几乎是不可能的，而且通常的调查会同时遇到上述中多个问题的组合。

【例 1.2】《文学摘要》（*Literary Digest*）（1932, 1936a, b, c）自 1912 年开始进行民意测验预测美国总统选举，其民意测验以精确性著称，因为它成功预测了美国 1912 年与 1932 年之间的历次总统选举。1932 年，其民意测验预测罗斯福（Roosevelt）会赢得 56%的普选（popular vote）选票和 474 张选举团（electoral college）的选票，最终结果是 58%和 472。正是有这样一个预测精确的记录，《文学摘要》的编辑们对预测 1936 年的总统选举充满自信。总统选举临近，他们说，他们的民意测验包含着 30 年的不断革新和不断完善，基于各种出版物使用近百年的商业抽样调查方法，现在的邮寄名单来自美国住户电话簿、俱乐部和社团的花名册、城市姓名地址录、已注册选民名单，还有一些分类的邮寄单和职业数据。在 10 月 31 日，民意测验显示，共和党候选人兰登（Alfred Landon）会赢得 55%的选票，而罗斯福只会赢得 41%，《文学摘要》登出文章，“兰登，1 293 669；罗斯福，972 897：来自 1 000 万选民的结果”。但是最终的结果令他们十分诧异，罗斯福得到了 61%，兰登只有 37%，罗斯福以压倒性的优势赢得了 1936 年的选举。

《文学摘要》的预测误差幅度之大令人吃惊，这是重要民意测验曾做出过的最大误差。这么大的误差是怎么来的呢?是什么地方出了重大差错呢？第一个问题是其抽样框与目标总体的差异，《文学摘要》使用的抽样框严重依赖电话簿和汽车登记名册。1936 年，拥有电话和汽车的家庭一般比其他家庭富裕得多，而罗斯福的经济政策主张多与社会底层相关。但是样本框偏倚并不能解释全部的问题。在《文学摘要》垮掉以后，由 Squire（1988）和 Calahan（1989）所做的分析表明，拥有电话和车的家庭尽管不如一般家庭支持罗斯福的程度深，但是还是倾向于赞成的。

误差如此大的另外一个原因，很可能是此次民意调查的回答率偏低。上面提到，近 1 000 万的问卷邮寄出去之后，只有近 230 万收回，如此庞大的样本只有不到 25%的回答率。例如，在宾夕法尼亚州（Pennsylvania）的一个镇，每个已登记选民都收到了一份邮寄问卷，但只有 $\frac{1}{3}$ 的样本返回。据 Squire（1988）

的研究报告，支持兰登的选民似乎更倾向返回问卷，很多罗斯福的支持者尽管也在邮寄名单上，却甚至不曾记得收到过问卷。

《文学摘要》的民意测验的教训是，纯粹的追求样本的数量并不能保证精确性。《文学摘要》的编辑们开始的时候非常满意，因为他们向所有选民的 1/4 邮寄了问卷而且收到了样本量为 230 万的巨大样本，但巨大的没有代表性的样本和小的没有代表性的样本一样无助于最后的推断，甚至会造成更大的危害！调查的抽样设计远比样本的绝对数量重要得多。

为了考察样本选择偏倚的影响，假定估计量 $\hat{\theta}$ 是正态分布的，即 $\hat{\theta} \sim N[E(\hat{\theta},\sigma_{\hat{\theta}}^2)]$，$E(\hat{\theta})$ 与总体真值 θ 之间有一段距离 B，偏倚大小是 $B = E(\hat{\theta}) - \theta$。假如我们并不知道有任何偏倚存在，计算这个估计值的标准差时，当然算的是关于均值 $E(\hat{\theta})$ 的分布的标准差，而不是对其值 θ 的标准差。我们以 σ 来代替 $\sigma_{\hat{\mu}}$，同时假定置信水平为 0.05，来考虑偏倚的存在如何影响估计的结果。我们来求估计值的误差大于 1.96σ 的真正概率，这个误差是与真值 σ 相比较而计量得出的。即

$$P\left\{\left|\hat{\theta}-\theta\right| \geqslant 1.96\sigma\right\} \tag{1.2}$$

对于上端，

$$\begin{aligned} P\left\{\left|\hat{\theta}-\theta\right| \geqslant 1.96\sigma\right\} &= P\left\{\hat{\theta}-E(\hat{\theta})+B \geqslant 1.96\sigma\right\} \\ &= P\left\{\frac{\hat{\theta}-E(\hat{\theta})}{\sigma} \geqslant \frac{1.96\sigma-B}{\sigma}\right\} = \frac{1}{\sqrt{2\pi}}\int_{1.96-\frac{B}{\sigma}}^{\infty} \mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t \end{aligned}$$

同样对于下端，也有相类似的结果：

$$P\left\{\left|\hat{\theta}-\theta\right| \geqslant 1.96\sigma\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-1.96-\frac{B}{\sigma}} \mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t$$

从积分的形式可以清楚地看出，干扰的量完全取决于偏倚对标准差的比率。将这个结果列入表 1.2 中。

表 1.2 显示，对误差大于 1.96σ 的全部概率，如果偏倚小于 $\frac{1}{10}$ 的标准差，那么偏倚的影响就很小；当偏倚等于标准差的 $\frac{1}{10}$ 时，则总的概率是 0.0511，而不是我们原以为的 0.05；当偏倚进一步增大时，干扰就变得更严重；当 $B = \sigma$ 时，总的误差大于 1.96σ 的概率是 0.17，大于预设的 0.05 的 3 倍多。

表 1.2　偏倚 B 对于误差大于 1.96σ 的概率的影响

$\frac{B}{\sigma}$	误差为下列数值时的概率		总和
	$<-1.96\sigma$	$>1.96\sigma$	
0.02	0.023 8	0.026 2	0.050 0
0.04	0.022 8	0.027 4	0.050 2
0.06	0.021 7	0.028 7	0.050 4
0.08	0.020 7	0.030 1	0.050 8
0.10	0.019 7	0.031 4	0.051 1
0.20	0.015 4	0.039 2	0.054 6
0.40	0.091	0.059 4	0.068 5
0.60	0.005 2	0.086 9	0.092 1
0.80	0.002 9	0.123 0	0.125 9
1.00	0.001 5	0.168 5	0.170 0
1.50	0.000 3	0.322 8	0.323 1

科克伦认为，作为一条公认的工作原则，若偏倚小于估计值标准差的 $\frac{1}{10}$，偏倚对估计值准确度的影响是可以忽略不计的；甚至当它等于 $\frac{1}{5}$ 时，这个干扰对总的误差的概率也只是中等的影响，超过 $\frac{1}{5}$ 时就要谨慎使用估计的结果。试想如果我们能找到 $\frac{B}{\sigma}$ 的上限，再看这个上限是在哪个范围，从而对结果有一个好的评估。但在实践中，要确定 $\frac{B}{\sigma}$ 的一个可靠的上限，通常不那么容易，当样本量足够大时，我们可以相信 $\frac{B}{\sigma}$ 不会大于 0.1。

2. 估计精度

通常，我们希望得到一个没有选择偏倚的样本作为总体的一个缩影，但现实的复杂性使得这样的希望只能作罢，而只能找出“偏倚”影响因素加以控制改进。上面提到的对非概率样本推断信度的最大威胁是抽样程序中存在主观性，选取样本当中存在主观的成分使得推断具有内在的不可靠性，因为最终的结果一定会有判断的成分在内。通过经验积累，研究者发现下述几种方法可以提高信度：

（1）对参与调查的人员进行严格的培训可以改进信度。

（2）让不同的专家共同参与非概率选样工作。

对同一问题同时进行分析，不同专家的意见看法会有所不同，选择其中具