



第 4 章 基本抽样问题

本章重点介绍的内容:

- 掌握抽样概念
- 学习制定抽样计划的步骤
- 明确随机样本与非随机样本的区别
- 明确抽样误差和非抽样的概念
- 复习随机抽样法类型
- 了解非随机抽样方法

4.1 重要术语的定义

4.1.1 总体与全域

在抽样领域,总体和全域是可以相互代替的。但在这里,我们将选用“总体”这个术语。总体或同质总体,是指能提供所需信息的人的全体。分析人员的首要任务是定义同质总体,这又经常涉及到对产品和服务目标市场的界定。

举例来说,一个研究人员正在为一种新型非处方感冒药进行产品创意测试。他也许会认为同质总体包括每个人,因为每个人会有患感冒的可能性。即便如此,并非每个患者都会选择这种非处方药。在这种情况下,调查过程中的重要任务是确定哪些人是目标主体,这就要看感冒时他们是否选购或使用这种或多种品牌的药。只有那些购或使用的人们,才应包括在同质总体内。

为同质总体下定义是抽样调查中关键的一步。为达到研究目的,究竟谁的观点才是我们真正需要采纳的呢?定义同质总体没有一定之规,需要研究人员有良好的辨别和判断能力。通常,对于总体的定义是常常基于已有的和潜在的顾客的特征。

4.1.2 抽样与普查

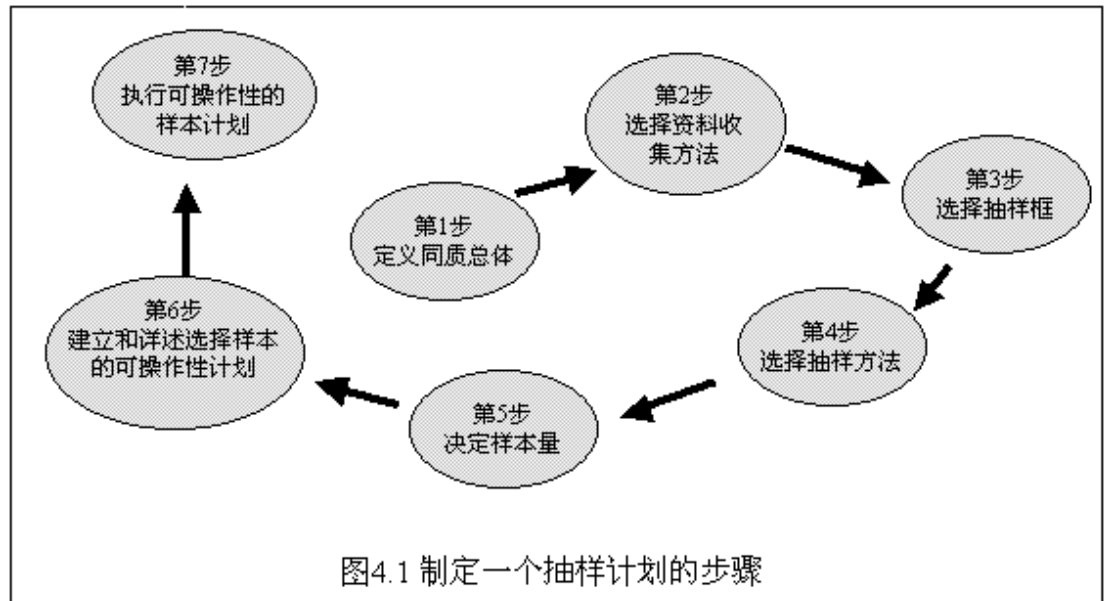
普查这一概念用于描述获取同质总体中每个成员的信息。市场调研中并不经常用到普查,因为其同质总体一般情况下包括成千上万的个体,这样大规模地进行普查在成本和时间上的耗费都是巨大的,以致于在通常情况下是不可行的。

事实已不止一次地证明:一个相对较小、但精心选择的样本能准确地反映出所抽查的总体的特征。一个样本是总体所有成员的一个子集。从总体中子集获得的有关信息,可以用来估测总体的特征。理想的情况是,子集能够代表总体的各个部分。

尽管市场调查中很少用到普查,但是有时它们也适用于某些案例。譬如,在工业产品营销中,一个企业只向少量客户销售极为特殊的产品。这时,普查则是适当和可行的,有可能从整个总体获取信息。尽管普查可能很具有吸引力(因为一般认为普查与抽样相比具有更大的准确性),但有时并非如此。例如,做人口普查时,试图从总体每个成员中获取信息事实上存在很多障碍。我们不能取得总体中每个成员完整、准确的资料,或许总体中某些成员会拒绝提供信息,考虑到这些障碍,即使涉及的总体规模较小,也很难进行普查。你也许读到或曾听到过美国人口普查中存在的问题。

4.2 制定抽样计划的步骤

开发一个具有可操作性的样本计划大致需要 7 个步骤,图 4.1。下面详细讨论这些步骤。



4.2.1 定义同质总体

为了满足研究目的的需要，必须详细说明可提供信息或所需信息有关的个体或实体（譬如公司、商店等）所具有的特性。同质总体可以从以下几方面进行描述：地域特征、人口统计学的特征、产品或服务使用情况、认知程度等。在调查中，从调查表开始部分的过滤性问题，可以看出某个体是否属于同质总体。即使有总体和样本清单，仍有必要使用过滤性问题识别合格的应答者。

另外，为了确定总体包括哪些人，通常情况下，重要的是确定那些应排除在外的人的特征。例如，大部分商业市场调查就因为一些所谓的安全性问题而排除某些个体。通常，问卷调查表上的第一个问题就是询问采访对象或其家庭成员是否从事市场研究、广告或生产与调查内容有关产品的工作。如果采访对象指出他们从事某项工作，那么就不必要去采访他了。这就所说的安全性问题，因为这样的采访对象不保险。他们也许是竞争对手或为竞争对手服务的，所以，我们不能给予他们关于我们打算干什么的暗示。

此外，排除某些个体还有其他原因。譬如，可口可乐公司就宁愿采访一些一周内饮用5瓶或5瓶以上各种包装饮料的人，而不愿采访一些只喝可口可乐的人。因为公司要加深对这些不喝可口可乐而喝软饮料的人的了解。因此，就会排除那些过去一周内经常喝可口可乐的人。

4.2.2 选择资料收集方法

资料收集方式对抽样过程有着重要的影响。例如，电话采访有一种内在优势，购物中心拦截顾客有着自身的劣势。

4.2.3 选择抽样框

在整个过程中的第三步就是确定抽样框。以往，我们把抽样框定义为总体的数据目录或单位的名单，从中可以抽出样本单位。理想的情况是，我们有一个完整和准确的名单。遗憾的是，通常这样的名单是不存在的。例如，在一项调查中，调查总体是那些在近30天内打三轮或三轮以上十八洞高尔夫球的人。但是，根本就没有一种计算方法可以完全提供这份名单。在不存在传统意义上的抽样框的情况下，我们需要依据能够产生具有希望特征的样本个体的程序来反映抽样框。例如，电话号码本就可能是电话调查的框架。这个例子也予明了在抽样框和同质总体间很少有极好的对应关系。在问卷中，同质总体很有可能是城市中的所有



居民。但是，电话号码本就不包括那些没电话的居民和那些没有公布他们号码的居民。

一些潜在的因素证明，公布电话的居民和不公开电话号码的居民在一些重要的特征方面具有很大的区别。很明显地，那些不主动提供电话号码的居民很大可能是房客，居住在城市中心，最近刚搬家，或人口多，孩子小，收入低。在某产品的购买、拥有、使用方面，两种类型的人具有很显著的差别。

在西部、大城市、非白人和 18-34 岁的人，不公布自己电话号码的人一般是比较多的。这项发现已一系列的调查所证实。这个发现的意义很明显，即当样本单位是从电话调查中获得时，还应该包括那些不公布电话号码但是却适合采访的城市居民。在这种情况下，可以使用一种程序来产生样本的名单。随机电话抽取就是随意抽取电话号码来产生样本名单。在抽样领域，形成一个适当的抽样框经常是调查者面临的最有挑战性的问题之一。

4.2.4 选择抽样方法

制定抽样计划的第四部是选择抽样方法。选择哪种抽样方法取决于研究目的、经济实力、时间限制、欲调查问题的性质。可供选择的重要抽样方法可以分为两大类：概率抽样与非概率抽样。每大类中又有许多可供选择的具体方法。

(1) 概率抽样

概率抽样指在总体中的每个单位都具有同等可能性被抽中。简单随机抽样是一种众所周知并广为使用的概率抽样。在概率抽样法中，调查员必须严格遵守正确的选择程序，即要求避免武断或有偏见地选择抽样单位，当严格执行这些程序时，概率论中的法则都是有效的。这就是说，对于预测的范围来说，样本中的数据可以认为是不同于总体数据的，其差异称为抽样误差。

(2) 非概率抽样

非概率抽样指从总体中非随机地选择特定的要素（单位）。根据简便易行、减少开支的原则选择总体中的某些要素（单位），会导致非随机性。有目的的非随机抽样可能会系统地排除或过分强调总体的某些部份。例如，一项要调查所有 18 岁以上女性意见的调查若在周一至周五的白天通过电话进行，显然会系统地排除了所有女性。

概率抽样包括以下几个优点：

- 调查者可获得被抽取的不同年龄、不同层次的人们信息。
- 能估算出抽样误差。
- 调查结果可以用来推断总体。例如，在一和概率抽样法的调查中，如果有 5% 的被访者给出了某种特定回答，那么，调查者就可以以此百分比再结合抽样误差，推及总体情况。另一方面，概率抽样也有一些弊病：
- 在大多数案例中，同样规模的概率抽样的费用要比非概率抽样高。通常精挑细选的做法不仅增加调研费用而且还要专门的时间对样本进行复核修改。
- 概率抽样比非概率抽样需要更多时间策划和实施。必须遵守的抽样计划执行程序会大量增加收集资料的时间。

非概率抽样的弊病恰好是概率抽样的优势：

- 不能估计出抽样误差。
- 调查者不知道抽中的单位所具有代表性的程度。
- 非概率抽样的结果不能也不应该推及总体。

看到非概率抽样的这些弊病后，你会奇怪：为什么还使用电脑呢？其实他们经常被市场调研人员使用。使用非概率抽样的原因与其本身固有的优势有关：

- 非概率抽样比概率抽样费用低。非概率抽样的这一特点对那些精确性要求不严格的调查有相当大的吸引力。试探性调查就是其中的一例。
- 一般来计，非概率抽样实施起来要比概率抽样用的时间少。原因在前面已讨论过了。

如果合理运用非概率抽样，它能产生极具代表性的合理的抽样结果。

除了是否选择概率抽样或非概率抽样之外，调研人员还须在众多抽样程序中选择其一。这些程序在图 4.2 中做了概括。

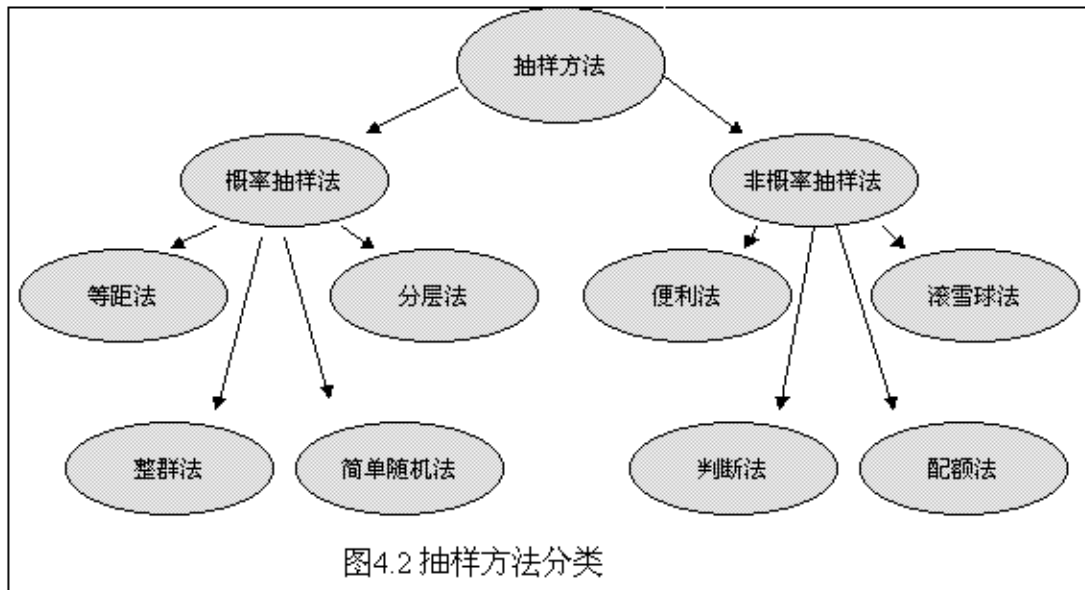


图4.2 抽样方法分类

4.2.5 确定样本量

一旦选定抽样方法，下一步就要确定合适的样本量。关于这个问题，下一章会有详细阐述。在下章中，我们讨论预算的重要性、各种各样的抽选规则、子集量的分析和传统的统计样本量的计算。对非概率抽样，通常我们仅依靠可能的预算、抽选规则、子集量分析来决定样本量。然而，就概率抽样而言，需要在允许误差的目标水平（抽样结果和总体指标的差异）和置信水平（置信区间的概率，置信区间是样本结果加减允许误差形成的一个涵盖总体真值的范围）下，计算样本量。如前所述，基于样本指标推测总体指标是概率抽样的主要优势。

4.2.6 制定选择样本单位的操作程序

无论使用概率或非概率抽样，在一个项目的资料收集阶段必须指定和明确选择样本单位的操作程序。对于概率抽样的成功来说，这个程序更为重要，必须详细、清晰，不受访问员的干扰。若不能制定合适选择样本单位的操作程序，则整个抽样程序会陷入困境。表 4.1 提供了一个适于操作的抽样计划。

表 4.1 适用于操作的抽样计划

以下指南是有关你在某个街区访问时应走的路径。在城市中，这可能是一个城市街区；在农村，街区可能是一块被道路包围的土地。
1. 如果在你的路线中遇到死胡同，继续沿这条路或街道的另一面向反方向走。在可能的地方右拐，每隔两户住家访问一户。
2. 如果你沿街区走了一圈，又回到了出发点而没有完成列出的电话簿上家庭的 4 个访问，那么可以试着访问起点的那一家。（很少用到这一方法）
3. 如果你调查了整个街区，还是没有完成所要求的访问，则继续从街（或乡间小路）对面最近的第一家住户开始，并访问这一家。如果不是，就访问这家左边的一家。永远遵守右手法则。
4. 如果这一地区街对面从第一号开始都没有住户，在第一号对面的街区转一圈，并遵循右手法则。（这意味着你将按顺时针方向在街区转一圈。）然后沿路线每隔两户访问一户。



5. 在起始门牌号对面邻近的街区绕公平上圈后，如果你没有完成所需的访问，就按顺时针方向，在下一个街区访问。
6. 如果第三个街区的住户数不够完成你的任务，就继续几个街区直到要求的户数完成为止；这些街区要按顺时针方向绕原来的街区来找。

4.2.7 抽样计划的实施

在实施适于操作的抽样计划前，应先对其进行讨论研究。这一步很重要，它包括检查、明确是否要根据拟好的详细程序来实施计划。

4.3 抽样误差和非抽样误差

请假想一种情景：我们的目标是想确定一个特殊人群的平均年龄。如果可以获得总体中每个人的确切信息，我们就可以计算出这些具有同一特征的人的平均年龄。总体的参数是总体真实特征定义的数值。假设（总体参数或平均年龄）为 36.3 岁。正如已讨论的那样，要调查整个总体中每一个人是不可能的，但是，研究人员会抽取样本，并根据样本的调查结果对总体的特征（参数）进行推测。如在计算平均年龄的问题中，研究人员从 25 万人的总体中抽取 400 个样本单位。估计的总体平均年龄要从本数据中计算出来，假设样本年龄为 35.8 岁。同样可以在总体中抽出第二个 400 人为样本，再计算其平均数，结果是 36.8 岁。另外，可能有很多适当的方法从总体中抽出若干个样本。研究人员将对不同的样本计算出相当接近但不完全同于总体参数的样本指标。

样本结果的准确性受到两种误差的影响：一种是抽样误差，另一种是非抽样误差（测量误差）。下列公式描述了在估计总体平均数时这二种误差的影响。

$$\bar{X} = \mu + \varepsilon_s + \varepsilon_{ns}$$

其中， \bar{X} 为样本平均数； μ 为真正的总体平均数； ε_s 为抽样误差； ε_{ns} 为非抽样误差或测量误差。

抽样误差是指所选样本的结果不能全代表总体而导致的误差。有两类抽样误差：随机的和管理上的。管理上的抽样误差涉及抽样执行中的问题，即样本的设计和和执行中有缺陷而不是样本不能代表总体。这类误差能在样本设计和执行中通过小心谨慎而避免或使之极小化。随机抽样误差是由于偶然事件引起的，是无法被避免的。这类误差只能依靠增加样本量使之缩小，但不能完全消除。非抽样或测量误差是在研究过程中由于计算的不准确和偏见等原因产生的不同于抽样误差的各种误差。

4.4 概率抽样方法

4.4.1 简单随机抽样

简单随机抽样是最完全的概率抽样。如前面提到的，随机抽样就是总体中每个单位在抽选时有相等的被抽中机会。在简单随机抽样条件下，抽样概率公式为：

$$\text{抽样概率} = \frac{\text{样本单位数}}{\text{总体单位数}}$$

例如，如果总体单位数为 10000，样本数为 400，那么抽样概率为 4%，计算过程为：

$$0.04 = \frac{400}{10000}$$

如果一个抽样框（列出所有的总体单位）是可以得到的，那么调查人员可以选择简单随机抽样方式，步骤如下：

- (1) 对总体的每个单位进行编号，总体单位数为 10000 的总体可编号为 1-10000；
- (2) 在随机数表中，从任意的一个编号数开始向上数或向下数或跳跃数选编号，在



0001 和 10000 之间选出 400 个（样本单位数）；

(3) 在有明确总体单位的数字表中选出的数字将包括在样本中。

简单随机抽样的优越性在于，它看起来简单，并且满足概率抽样的一切必要的要求，保证每个总体单位在抽选时都有相等的被抽中机会。简单随机抽样以一个完整的总体单位表为依据，在现实中编制这样一个完整的表是极其困难的，也是不可能做到的。简单随机抽样可以通过电话随机拨号功能完成这个步骤。最后，简单随机抽样可以从电脑档案中挑选调查对象。电脑资料是可以得到的或从电脑档案中抽取随机样本的记录，如顾客名单。

4.4.2 等距抽样

等距抽样经常作为简单随机抽样的代替物使用。由于其简单，所以应用相当普遍。等距抽样得到的样本几乎与随机抽样得到的样本相同。

使用这种方式，必须获得一份总体单位表，这一点与简单随机抽样方式一样。调查人员必须决定一个间隔，并在此间隔基础上选择单位，样本距离可通过下面公式确定：

$$\text{样本距离} = \frac{\text{总体单位数}}{\text{样本单位数}}$$

例如，假设你使用本地电话号码本并确定样本距离为 100，那么 100 个中取 1 个组成样本。这个公式保证了整个列表的完整性。

等距抽样方式随意用一个起点，例如你正使用一本电话本，必须随意取出一个号码决定从该页开始翻阅。假设从第 53 页开始，在该页上再另选一个数决定从该行开始。假设选择从第 3 行开始，最后在该行任选一个数，这就决定了实际开始的位置。假定从第 17 个数开始，那么以此为起点，样本距离就确定下来了。

等距抽样方式相对于简单随机抽样方式最主要的优势就是经济性。等距抽样方式比简单随机抽样更为简单，花的时间更少，并且花费也少。使用等距抽样方式最大的缺陷在于总体单位的排列上。一些总体单位数可能包含隐蔽的形态，调查者可能疏忽，把它们抽选为样本。然而，这种缺陷在使用字母表时可以消除。

4.4.3 分层抽样

分层抽样是一种卓越的概率抽样。具体程序是：

- 把总体各单位分成两个或两个以上的相互独立的完全的组（如男性和女性）。
- 从两个或两个以上的组中简单随机抽样，样本相互独立。

总体各单位按主要标志加以分组。尽管一些作者指出，分层抽样的要求没有指明分层标志，但是根据常识的判断，分组的标志与我们关心的总体特征相关。例如，如果你正在进行一次政治性民意调查，要预测选举结果。结果证明，男性和女性投票的方式大不相同，那么性别是划分层次的适当标志。如果不以这种方式进行分层抽样，分层抽样就得不到什么效果，花再多时间、精力和物资也是白费。前面的例子中，将性别作为分层抽样的标志，我们得到男、女两组，各组都相互完全独立。在一组中（男性组或女性组）保证每个总体单位都有被选的机会，没有哪些总体单位是不能被抽中的。第二步就是进行简单随机抽样，在每个组中独立进行。

分层抽样与简单抽样相比，人们往往选择分层抽样，因为它有显著的潜在统计效果。也就是说，如果我们从相同的总体中抽取两个样本，一个是分层样本，另一个是简单随机抽样样本，那么相对来说，分层样本的误差更小些。另一方面，如果目标是获得一个确定的抽样误差水平，那么更小的分层样本将达到这一目标。由于排除了一种变差的来源，所以分层抽样在统计上更为有效率。为此，后面的部分将做详细的解释。

你也许会问：“如果分层抽样统计效果更好，为什么不一直使用这种方式呢？”原因有两个：

(1) 将样本适当划分层次所需的信息常常是得不到的。例如，几乎没有人知道某种特殊产品的消费者的人口统计特征。注意，我们强调的是：“适当”分层。要适当分层并得到分



层的好处，必须选择个层次间存在明显差异的因素为分层基础。

(2) 即使必要的信息是可以得到的，但是从所得信息的价值看，分层所需的时间和费用不划算。

对于简单随机抽样，调查者完全依照随机原则抽取总体中有代表性的样本；对于分层抽样，调查者按各组总体单位数占全部总体单位数的一定比例来抽取样本，某种程度上更具有代表性。例如，调查者可能了解到，尽管男性和女性同样有可能是某种特殊产品的使用者，但女性更有可能成为产品的主要使用者。在设计一个方案分析产品消费情况时，抽取的样本中如果女性不具更大的代表性，将会导致片面的消费设计方案。假定女性占总体的 60%，男性占总体的 40%。即使每件事都做得完全正确，但是简单随机抽样的程序可能会抽取到这样的样本，女性占 55%，男性占 45%，这种结果是由于样本的波动。当我们将一枚硬币掷 10 次时也会有这种误差。正确的结果将是 5 次正面和 5 次背面，但是大部分时间我们得到的结果不是这样。同样，即便设计正确和操作简单的随机抽样，也不可能从女性占 60% 和男性占 40% 的总体中恰好抽取到一个含 60% 的女性和 40% 的男性的样本。然而，对于分层抽样，调查者可以控制样本含 60% 的女性、40% 的男性。

如上所述，提高分层样本的精确度要付出一些代价。实现正确的分层抽样有三个步骤：

首先，辨明突出的（重要的）人口统计特征和分类特征，这些特征与所研究的行为相关。例如，研究某种产品的消费比率时，按常理认为男性和女性有不同的平均消费比率。为了把性别作为有意义的分层标志，调查者肯定能够拿出资料证明男性与女性的消费水平明显不同。用这方式可识别出各种不同的显著特征。调查表明，一般来说，识别出 6 种重要的显著特征后，再增加显著特征的辨别对于提高样本代表性就没有多大帮助了。

第二：确定在每个层次上总体的比例（如性别已被确定为一个显著的特征，那么总体中男性占多少比例，女性占多少比例呢？）。利用这个比例，可计算出样本中每组（层）应调查的人数。当然，在做最终决定之前，还要确定是否按总体比例或不按总体比例分配样本各组单位数，以达到最佳样本。

- 按比例分配，即样本各组比例与总体各层比例等同。

用公式表示：

$$\frac{n_i}{n} = \frac{N_i}{N}$$

式中 n—样本量；

N—总体单位数；

i—各组（层）单位数。

- 非比例分配或最佳分配，它能产生最有效的样本，并为样本提供最精确的和最可靠的估计。这种方法需要一个双重方案，其一是考虑样本量中各组比例与总体各层比例相关；其二是，样本量中各组单位数的样本还要考虑总体各层之中共有特征的离差（变差）。前者固然很重要，它涉及对总样本数的估计；后者对总体某个变异较大的层，通过扩大其在样本量中的比例，可提高样本的代表性，相应也提高了估计的准确性。如果总体各层之中变异（标准差）相同，不成比例抽样则没有太大意义。

最后，调查者必须从每层中抽取独立简单随机样本。现实中，这个步骤可以以不同的方式完成。假设按分层抽样计划要求访问 240 位女性和 160 位男性。样本从包括男性和女性的总体中抽取，在访问过程中对被访的男性和女性的人数进行跟踪。在访谈过程中的某个时点，可能访问了 240 位女性和 12 位男性。自此以后，只访问男性，直到达到 160 人的目标。按这种方式，样本中男性和女性的比例将与第二步中的要求一致。

分层抽样在市场研究中用到并不很频繁。因为，我们往往预先得不到给样本分层的必要信息。分层不能建立在猜测或预感的基础上，而应建立在对总体特征及其内部变量关系和所



调查行为的充分认识的基础上。分层抽样经常被用在政治选举和大众媒体的研究，在这些领域，研究者更情愿也更容易掌握分层的必要信息。

4.4.4 整体抽样

迄今为止，我们讨论的抽样类型全部是按单位抽取，即按样本单位数，分别一个单位一个单位地抽取。在整群抽样中，样本是一组单位一组单位地抽取。这里有两个步骤：

- 同质总体被分为相互独立的完全的较小子集。
- 随机抽选子集构成样本。

如果调查者在抽中的子集中观察全部单位，我们就有了一级整群样本。如果在抽中的子集中再以概率方式抽取部分单位观察，我们就有了二级整群样本。分层和整群抽样都要将总体分为相互独立的和完全子集。它们的区别是，分层抽样的样本是从每个子集中抽取，而整群则是抽取部分子集。全部观察一级样本或先抽取部分子集再在抽中的子集中抽取部分单位，取得二级样本。

所有概率抽样法 需要列举或提供一些已组织好的关于目标总体所有单位的统计表。在整群抽样中，研究人员开发出了不需列举所有单位而只需列举子集的抽样。抽取子集后，再列举其内部单位统计表，最后取得样本。

地理区域抽样是整群抽样的典型方式。挨门挨户去调查一个特定城市的调查者也许会随机抽选一些区，较集中地访查一些群体，大量减少访问时间和经费。整群抽样被认为是概率抽样技术，因为它随机抽出群和随机抽出单位。

在整群抽样下，我们假定群中单位与全及总体一样存在异质性。如果一群中单位的特征非常相似，那么就违反了假设。在上面的例子中，由于共同环境使群内差异小而群与群之间差异大。一般来说，要解决这个问题可以扩大群数，然后从各群中抽取少量单位数，以保证样本的代表性。

这种整体抽样的类型是一个两级抽样。第一级是群的抽取。第二级是群内单位的抽取。多级区域抽样或多级概率抽样也许有着更多的步骤。这种类型被应用于覆盖广阔区域的全国性调查。在这种抽样方法下，调查者随机抽取相对较小的区域。例如，在一个州内的挨门挨户的调查也许包括以下步骤：

- (1) 选取这个州内的县以表明不同区域都有代表性，县也许是通过与县中家庭数成正比的方式抽出，家庭数目较多的县比家庭数目较少的县抽中的机会要大一点。
- (2) 在样本县中抽住宅区域。
- (3) 在中选住宅区域中抽取家庭。

从统计效率的立场看，一般说来整体抽样的效率没有其他类型高。换言之，一组一定大小的整体样本，将比一组简单随机样本或一组同样大小的分层样本有更大的抽样误差。我们用下面的例子来说明这种方法的高成本绩效及其低统计效率。

比如，我们需要在某个特定的城市选取一个由 200 个家庭组成的样本来进行家访。如果这 200 个家庭是通过简单随机抽样抽取的，它们会散布到整个城市之中。但通过整群抽样可以抽取城市中 20 个住宅区并从每区中抽取 10 户人家来采访。容易看出，整群抽样可以大幅度降低费用，采访者将不必花费大量时间穿梭于各个被采访的对象之间。然而，简单随机抽样的抽样误差却小一些。散布在整个城市的 200 户人家，提高了得到一组涉及各个方面被访者代表的可能性。相反，如果采访仅仅在城市中选定的住宅区中进行，那么，就有可能错过、夸大或降低某些宗教，社会或经济团体的代表意义。

如前所述，整体抽样没有简单随机抽样的统计效率高。简单随机抽样也可视为一种特殊的整群抽样，其中样本量与整群相当，从每群中选一个样本单位。在这一点上，整体抽样的统计效率和简单随机抽样相同。从这一点开始，我们只选择一个群并从中抽所有的样本单位。例如，可以在城市中选一个相对小的地区，在那访问 200 人。如果这种方式能抽出代表整个



城市居民的样本，那么，工作也太容易了！

4.5 非概率抽样

一般而言，任何不满足概率抽样要求的抽样都被归为非概率样。非概率抽样的缺点是，不能计算其抽样误差，这意味着评估非概率抽样的总体质量有很大的困难。我们知道它们不满足概率抽样所需的标准，但问题是它们脱离标准有多远？资料的使用者必须对非概率抽样进行评估。评估应该建立在对非概率抽样方法论仔细评价的基础上。那么，使用的方法是否能够覆盖目标总体的各个部份？或者样本是否无目的地倾向于一些特殊方面？这些是仔细评估时必须考虑的。经常使用的非概率抽样有四类：便利抽样、判断抽样、配额抽样和滚雪球抽样。

4.5.1 便利抽样

顾名思义，便利抽样被应用是因为便利。如 Frito-Lay 公司的 R&D 部门经常让员工对其开发的新产品进行初步测试。固然，这个方法看上去有很大的偏差，然而他们不要求雇员评估现有的产品或竞争对手的产品进行比较。他们要求雇员提供总的感觉（如成色、脆度、油腻度）。在类似情况下，便利抽样是获得必要信息的有效而实用的方法。在进行试探性调研时，即缺乏经验而急需真实数据的近似值时，这种方法很实用。

有人认为，与概率抽样相比，便利抽样应用比率增长很快。其原因是由于在低发生率和难以分类情形下消费者资料库的可获得性。例如，一家公司开发出了一种新型脚癣治疗仪，并且需要在受疾病困扰的人中做一次调查。它发现这些人仅占人口的 4%。这表明在电话调查中，调查员在找到一个忍受脚癣痛苦的人之前，不得不与 25 个人交谈。一个变通的办法是努力取得一个已知目标者（忍受脚癣痛苦者）的名单。这样，调查成本和所需时间就会大大减少。尽管一个名册中也许会包括那些购买产品时使用优惠券的人，公司仍甘愿低成本快速获得低质量的样本。从 Survey Sampling 公司可以获得 3000 多份类似的名单，具体可参见 www.ssisamples.com/lite/lite.gry。

4.5.2 判断抽样

判断抽样适用于调查员基于选择标准抽取典型样本的任何情形。购物中心进行的大部分市场或产品测试调查基本上都属于判断抽样。就市场测试而言，是选择一个还是几个市场需看它们能否代表全及总体。为调查产品品味而选择购物中心取决于调查人员的判断。特殊的购物中心会吸引不同阶层的消费者，而这些人恰好是某种被调查产品的既定调查群。

4.5.3 配额抽样

配额抽样是根据一定标志对总体分层或分类后，从各层或各类中主观地选取一定比例的调查单位的方法。所谓“配额”是指对划分出的总体各类型都分配给一定数量而组成调查样本。因而，配额抽样较之判断抽样加强对样本结构与总体结构在“量”的方面的质量控制，能够保证样本有较高的代表性。配额抽样类似于随机抽样中的分层抽样。不过，有两点重要的区别：配额抽样的被调查者不是按随机原则抽出来的，而分层抽样必须遵守随机原则。其次，在分层抽样中，用于分类的标志，应联系研究目标来选择，而配额抽样无此要求。

4.5.4 滚雪球抽样

滚雪球抽样是指，通过使用初始被调查者的推荐来挑选另外的被调查者的抽样程序。这种方法用于低发生率或罕见的总体中进行抽样。至于低发生率或罕见的总体，是指全及总体中所占比例很少的那一部分。要找到这些少见总体中的个体，代价是很大的，使得调查人员因为费用的原因不得不使用例如滚雪球抽样那样的技巧。例如，某保险公司可能想得到过去 6 个月中从健康保险转入康复组织的全国性个体样本，为了找到符合条件的 1000 个样本。可能需要在全国范围内进行大量的调查。然而，若先取得特征总体中 200 个最初样本单位，平均从每个最初调查者那里得到另外 4 个人的名单，以此来完成这 1000 个样本单位就经济多了。



滚雪球抽样调查的优点是调查费用大大减少,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,因为那些个体的名单来源于那些最初调查过的人,而他们之间可能十分相似。结果,样本可能不能很好地代表整个总体。另外,如果被调查者不愿意提供人员来接受调查,那么这种方法就会受阻。

小结

总体或全域是指能提供所需信息的一群人的总和。普查指从总体的所有个体中搜集令人满意的信息。一个样本仅仅是总体的一个子集。制定抽样调查计划的步骤如下:定义同质总体、选择资料收集的方法、选择抽样框、选择抽样方法、决定样本容量、确定一个可执行的实施计划、然后执行抽样调查计划。抽样框是列出总体中个体名单的方法,样本就是从中抽出来的。

由于概率抽样法是按随机原则抽选样本,以致总体中每个单位者可能按已知非零的概率被抽中。非概率抽样调查法包括按非随机原则从总体中选择特定单位的各种方法。概率抽样法有一些非概率抽样法所没有的优点,包括信息来源于总体中具有代表性的各层次,抽样误差可被计算出来,并且可以用调查结果来估计总体。然而,概率抽样法的费用比非概率抽样法高,并且通常需用更多的时间来设计和实施调查。

抽样调查结果的准确度是由抽样误差和非抽样误差决定的。抽样误差是因为样本不能理想地代表总体特征造成的。有两种形式的抽样调查误差:随机性误差和管理性误差。随机性抽样误差是因为偶然事件产生的,不可避免,只能通过增大样本容量而减少误差。概率抽样法包括简单随机抽样、等距抽样、分层抽样和整群抽样。非概率抽样包括便利抽样、判断抽样、配额抽样和滚雪球抽样。