

高等学校统计学专业规划教材

# 抽样调查

(修订版)

■ 牛成英 庞智强 编著



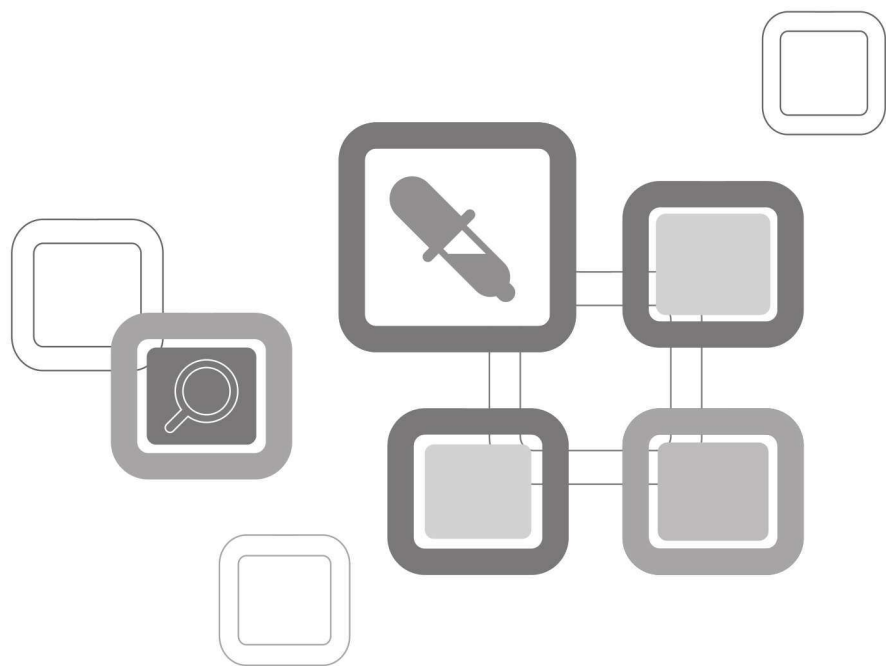
兰州大学出版社  
LANZHOU UNIVERSITY PRESS

高等学校统计学专业规划教材

# 抽样调查

(修订版)

□ 牛成英 庞智强 编著



兰州大学出版社  
LANZHOU UNIVERSITY PRESS

# 第一版前言

抽样调查作为一种科学的统计调查方法，自1895年被第一次正式提出以来，在百余年的发展历程中，经过众多统计理论和实践工作者的不懈努力，现已形成了较为完整的理论和技术体系，并被广泛应用于经济调查、社会调查、科学实验、生产管理等诸多领域。抽样调查也因此而在统计学的学科体系中占据了十分重要的地位，成为统计学中最为活跃、发展也最快的分支之一。

在我国，抽样调查的应用经历了一个缓慢而又曲折的过程，随着社会主义市场经济体制的建立，在政府统计调查方法体系中还是确立了经常性抽样调查的主体地位，并在1995年修订的《统计法》中做出了明确的规定。尽管如此，抽样调查在我国的大面积推广应用还面临着一系列制约因素，其中应用型专业性人才的缺乏即是一个比较突出的问题。因此，加快抽样调查专业人才的培养成了时代赋予我们统计教育工作者的重任。

进行人才培养要有好的教材，但是抽样调查方面的教材却很难编，原因是目前国内外关于这方面的教材版本很多，而且各有特色。本书的编写力求突出实用性，在内容上能涵盖抽样调查的各个不同阶段，既适合于高校经济管理类专业本科学士学习之用，又为广大统计工作者所容易接受。本着这一想法，1997年我们编写了一本“《抽样调查》讲义”，在兰州商学院统计学专业连续试用三年。今天奉献给广大读者的这本《抽样调查》是在上述讲义的基础上，经过较大幅度的修改而成。考虑到随后要单独编写一本《抽样调查学习指南与配套练习》，所以，在本书中没有过多举例。本书第一章、第二章、第三章、第四章、第十一章和第十二章由庞智强编写，第五章、第六章、第七章、第八章、第九章和第十章由马保平编写，全书的内容体系设计和最后的总纂由庞智强完成。

在本书的编写过程中，得到了兰州大学朱杰教授、兰州商学院夏绍萱教授、龚业勤副教授的热情鼓励和帮助，兰州商学院傅德印博士提出了许多好的建议，兰州商学院教材编审委员会和统计学系给予了大力支持，在此一并表示衷心的感谢。此外，我们还参阅了国内外各种版本的同类教材，这里要向其作者们特别表示感谢，因为，如果没有他们的工作成就，则本书的编写就成了无源之水。

限于编者水平，书中错误和疏漏之处在所难免，敬请广大读者和专家学者批评指正。

编者

2000年春节

## 再版前言

进入新世纪的二十年来，伴随统计学学科与专业建设的快速发展，抽样调查的理论、方法与实践应用也取得了长足的进展。与此同时，大数据时代的到来，也使抽样调查学科发展面临着一系列新的机遇和挑战。

《抽样调查》第一版出版以来，我们在教学应用中也发现了一些该书存在的疏漏和需要进一步完善的内容。比如，原本计划要出版的配套练习等也没有能够及时跟进。伴随统计专业调整分设、专业培养方案修订、课程设置体系改革、教学大纲修改等一系列变革，《抽样调查》的部分章节内容已不适应新的需要。另外，历经《抽样调查》省级精品课程的建设，我们对课程内容框架体系也有了一些新的理解和思考。这些，是我们修订再版《抽样调查》这本教材的背景。

本次修订中，我们将原来的第一章“抽样调查概述”和第二章“抽样调查基本原理”整合优化为新的第一章“抽样调查的基本问题”；删除了原来的第十章“问卷设计与调查”；每一章补充了练习与思考题；修正了第一版中的部分错误；完善了一些概念表述。

修订工作由庞智强教授与牛成英副教授共同完成。其中，庞智强负责整合后第一章内容的修订，牛成英负责其余章节的修订并补充所有章节的练习与思考题，同时对书中的公式和图表进行校对。

本教材的再版，得到了兰州财经大学校长蔡文浩教授的积极鼓励，得到了统计学院同仁的支持，也得到了兰州大学出版社的大力支持，在修订过程中我们也参考了各种新版同类教材，在此一并表示感谢。

限于编者水平，修订后的教材错误和疏漏依然在所难免，敬请各位同行、读者批评指正。

编者  
2020年12月

# 目 录

第一章 抽样调查的基本问题 .....	1
第一节 抽样调查概念与种类 .....	1
第二节 抽样调查的特点与应用 .....	13
第三节 抽样分布 .....	18
第四节 抽样误差 .....	24
第五节 抽样估计 .....	27
思考与练习题 .....	30
第二章 简单随机抽样 .....	32
第一节 简单随机抽样概述 .....	32
第二节 总体参数的简单估计 .....	34
第三节 样本容量的确定 .....	42
思考与练习题 .....	49
第三章 分层抽样 .....	51
第一节 分层抽样概述 .....	51
第二节 总体参数的估计 .....	53
第三节 总样本量在各层的分配 .....	59
第四节 总样本量的确定 .....	68
第五节 分层与提高精度 .....	72
第六节 划分层的问题 .....	76
思考与练习题 .....	80
第四章 比率估计与回归估计 .....	83
第一节 简单随机抽样下的比率估计 .....	83
第二节 分层抽样下的比率估计 .....	89
第三节 简单随机抽样下的回归估计 .....	93
第四节 分层抽样下的回归估计 .....	98
思考与练习题 .....	108

第五章 不等概率抽样 .....	111
第一节 不等概率抽样概述 .....	111
第二节 放回不等概率抽样 .....	113
第三节 不放回不等概率抽样 .....	119
思考与练习题 .....	124
第六章 整群抽样 .....	127
第一节 整群抽样概述 .....	127
第二节 等概率整群抽样 .....	129
第三节 不等概率整群抽样 .....	139
第四节 设计效应和样本容量的确定 .....	142
思考与练习题 .....	145
第七章 等距抽样 .....	147
第一节 等距抽样概述 .....	147
第二节 等距抽样的实施方法 .....	150
第三节 总体参数的估计 .....	152
第四节 具有线性趋势总体的等距抽样方法改进 .....	160
第五节 其他形式的等距抽样 .....	165
思考与练习题 .....	170
第八章 多阶段抽样 .....	172
第一节 多阶段抽样概述 .....	172
第二节 初级单元大小相等的多阶段抽样 .....	175
第三节 初级单元大小不等的二阶段抽样 .....	186
思考与练习题 .....	194
第九章 二重抽样 .....	198
第一节 二重抽样概述 .....	198
第二节 二重分层抽样 .....	200
第三节 二重比率估计 .....	205
第四节 二重回归估计 .....	208
思考与练习题 .....	211
第十章 非抽样误差及其控制 .....	213
第一节 抽样框误差及其控制 .....	213

第二节 调查误差及其控制 .....	220
第三节 无回答及其控制 .....	223
第四节 样本轮换的理论与方法 .....	230
思考与练习题 .....	234
参考文献 .....	236
附表 .....	238

# 第一章

## 抽样调查的基本问题

### 第一节 抽样调查概念与种类

随着社会信息化的不断发展，大到一个国家或地区的经济社会发展战略、发展规划及各项政策法规的制定，小到一个企业生产经营过程的各种管理决策，都离不开对各种信息资源的依赖。信息资源的表现形式多种多样，概括起来大致可分成两大类型：一是定性化的信息，如各种文字性质的消息、新闻、报告等情报材料；二是定量化的信息，如各种数字性质的图、表等指标资料。前者可借助于各种传播媒体而获取，后者则主要通过统计调查整理而形成。抽样调查是统计调查中最常用的基本方法之一，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、管理会计学等学科的方法论基础。

#### 一、抽样调查的内涵

抽样调查的定义可分为广义和狭义两种，广义的抽样调查包括非概率抽样与概率抽样，狭义概念仅指概率抽样，本书取狭义的理解。

抽样调查是指按照一定的程序和方法，从所要研究现象的总体中根据随机原则抽取一部分单元组成样本，通过对样本的调查，获得样本资料，计算出有关的样本指标（统计量），依据对相应的总体指标（参数）作出估计和推算，并有效控制抽样误差的一种统计方法。

随机原则是抽样调查所必须遵循的基本原则。所谓随机原则就是在抽选调查单元的过程中，完全排除人为主观因素的干扰，以保证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，那些单元能够被选作调查单元纯属偶然因素的影响所致。在这里，“可能性”就是常说的机会，对“可能性”大小所作的数值度量在数理统计学中被称作概率，它常常是用统计的频率来近似度量的。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

## 抽样调查

多时，样本就会遵从大数定律而呈正态分布，样本单元的标志值才具有代表性，样本均值才会接近总体均值；按随机原则抽样，才可能实现计算和控制抽样误差的目的。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。

由抽样调查的概念出发，可以将抽样调查工作的全过程划分成三个阶段。

第一阶段为抽样设计阶段，即从现象总体中抽选样本的阶段。在这一过程中，要在对现象总体进行初步分析的基础上编制出尽可能完善的抽样框（抽样单元的名录），选择最适合的抽样方式方法，确定出最佳的样本量，以充分保证所选样本对整个总体的代表性，从而增强抽样推断的效果。同时，还要做好问卷设计和试调查等工作。

第二阶段为调查阶段，即对所选中的样本进行调查，搜集样本资料的阶段。在这一过程中，要解决好具体调查方法的选择，即确定是用走访调查、电话调查、邮寄调查还是其他有关方法，以搜集所需资料。还要设计出对拒访和无回答（包括样品无回答和项目无回答）的预防和处理措施，以及对数据质量的检查、控制措施与评价方法等内容。其基本目标是获取准确的资料，为抽样推断的可靠性打下基础。

第三阶段为数据处理和估计推断阶段，即依据样本资料对所需要的总体资料进行科学估计推算的阶段，是抽样调查工作的最后一个、同时也是最重要的一个阶段。在这一阶段，涉及对推断总体资料的各种方法（估计量）进行比较、选择，估计误差的计算与控制，缺失数据的处理，以及进行数据分析，得出分析结果。其核心是不断提高推断结果的可靠性，圆满完成抽样调查的各项任务。

上述阶段划分可进一步用示意图（如图1-1）来表示：



（图中A表示现象总体，B表示样本）

图1-1 抽样调查阶段划分

这表明，抽样调查工作的出发点是现象的总体，其目的也是要了解总体的有关方面特征，但在具体认识过程中是以样本为核心，绝大部分工作也是围绕样本展开。但其工作的结果绝不仅仅停留在对样本的认识上，而是最终上升到对整个总体的更高层次的认识和把握，也就是以总体作为全部工作的归宿点。

## 二、抽样调查的基本术语

### (一) 总体

在统计研究中，将由在一些基本属性方面性质相同的大量个别事物所组成的整体称之为统计总体，简称总体。通常，将构成总体的、无法再细分的个体称之为总体单元，简称单元。总体具有同质性、大量性、差异性的特征。总体根据所含单元数量的多少可以分成有限总体和无限总体。如果总体中所包含单元的数目为有限个，则该总体就是有限总体，反之是无限总体。

抽样调查中的总体又有目标总体与抽样总体。目标总体也可简称为总体，是指所要研究观测对象的全体，或者说是希望从中获取信息的全体，它是具有同一性质的许多单元的集合。在研究中，定义目标总体是重要的，通常也是困难的。比如在政治选举投票中，目标总体应该是所有有权投票的成年人？还是登记的投票者？还是在最近一次选举中所有投过票的人？目标总体的选择将会对统计结果产生深远的影响。

抽样总体是有可能被抽取到的所有观测单元的集合，即从中抽取样本的总体。在理想的调查中，抽样总体与目标总体应该是一致的，但这种完全一致很少能够实现。对人的调查中，抽样总体经常小于目标总体，这是因为并不是所有目标总体中的人都包括在抽样框里（关于抽样单元的名录），而且有一些被抽中的人并不接受调查。例如，电话调查中，并非所有家庭都有电话，因此可能的被调查者组成的目标总体中有一些人无法与抽样框中的电话号码相关联，并且在一些有电话的家庭中，居民没有包括在调查范围内，因此不符合调查的条件。而且抽样框中一些符合调查条件的人，因为联系不到或者拒绝接受调查，或者因为生病无法接受调查，因此最终也没有给予回答（如图1-2）。



图1-2 目标总体与抽样总体的关系

### (二) 抽样单元

在抽样以前，必须根据实际情况把总体划分成若干个互不重叠并且能组合成总体的部分，每个部分称为一个抽样单元。一个总体无论其所含单位数目是否有限，但总可以划分为有限个抽样单元。因此，从总体所含抽样单元数目有限的角度看，抽样调查中的所有总

## 抽样调查

体均为有限总体。抽样单元又有大小之分，一个大的抽样单元可以分成若干个小的抽样单元，最小的抽样单元就是构成总体的每一个总体单元，也可以称之为基本单元。在一项全国性的调查中，如果把省作为一级单元，则可以把县作为二级单元，乡作为三级单元，村作为四级单元等等；在流动人口抽样中，可以以居委会作为抽样单元，而在家计调查中，则以户为抽样单元。

### (三) 抽样框

抽样框是指抽样前为便于抽样工作的组织，在可能条件下编制的用来进行抽样、记录或表明总体所有抽样单元的框架。通常，抽样框是关于抽样单元的名录，给每一个抽样单元编上一个号码，就可以按一定的随机化程序进行抽样。抽样框可以是一份清单（名单抽样框）、一张地图（区域抽样框）。在与时间有关的调查中，也可以按时间先后顺序排列总体中的单元，这样得到的抽样框称为时序抽样框。例如，对于电话调查，抽样框可能是调查城市中所有居民电话号码的名单；对于入户调查，抽样框可能是所有街道地址的名录；对于农业调查，抽样框则可能是所有农田的名录或者是农田地区的地图。

抽样框是设计实施一个抽样方案所必备的基础资料，一旦某个单元被抽中，也需要依据抽样框在实际中找到这个单元，从而实施调查。编制抽样框是一个实际而重要的问题，必须要认真对待。一般而言，如果总体中的每个元素在清单上分别只出现一次，且清单上又没有总体以外的其他元素出现，则该清单就是一个完备的抽样框。在完备的抽样框中，每个元素必须且只能与一个号码对应。但是，在实际中，完备的抽样框是很少见的，可能不得不使用一些有严重缺陷的抽样框，而又必须研究这些缺陷并加以补救，在这一过程中，可以充分体现抽样的艺术性。

常见的抽样框问题可以概括为四种基本类型：一是缺失一些元素，即抽样框涵盖不完全；二是多个元素对应一个号码；三是空白（一些号码没有与之对应的元素）或存在异类元素；四是重复号码，即一个元素对应多个号码。

当抽样框存在缺陷时，应尽量设法避免这些缺陷：如果已知由这些问题引起的误差比其他原因产生的误差小，并且纠正起来又花费太大的话，可以忽略不管，但在描述样本时应对此加以说明；重新定义总体以适应抽样框；改正整个总体清单，即找出全部缺失元素、清除所有的空白和异类元素、删掉重复号码。当上述方法不能有效利用时，就应该采取其他一些补救措施来抵消抽样框中存在的缺陷。

### (四) 样本

样本是由从总体中所抽选出来的若干个抽样单元组成的集合体，是总体的一个子集。抽样前，样本是一个 $n$ 维随机变量，是样本空间；抽样后，样本是一个 $n$ 元数组，是样本空间的一个点。

样本是总体的缩影，是总体的代表。抽样效果的优劣，依赖于样本对总体是否有充分

的代表性。一个完美的样本就应该是总体的一个规模缩小的版本，但反应的是整个总体的所有特征。当然，对于复杂的总体并不存在这样一个完美的样本（即使这样的样本存在，如果对整个总体进行测量，我们也不知道它是否完美），但是，一个优良的样本应该尽可能地再现总体中我们感兴趣的特征，每个人样单元都能表现已知数量总体单元的特征，在这种意义上样本才是有代表性的。样本的代表性越强，用样本指标对总体全面特征的推断就越精确，推断的误差就越小；反之，如果样本的代表性越弱，推断的误差就越大，推断结果就越不可靠。要增强样本的代表性，使其能达到估计或推断的预期效果，就必须分析影响样本代表性的因素，以便加强控制。一般情况下，影响样本代表性的因素有以下几个方面：

(1) 总体内部的差异程度。若总体内部各单元的变量值水平比较集中，即平均离散程度（标准差）很小，从中任意抽部分单元做样本，样本特征很近似于总体特征，样本的代表性就强；反之，如果变量值的分布很分散，即平均离散程度很大，从中抽取样本单元的随机波动也很大，必将影响样本的代表性。

(2) 抽样单元数的多少（或称样本容量的大小）。一般说来，样本容量以大为好，但要根据实际情况，以掌握适度为宜，要在保证一定可靠程度的情况下，尽可能满足及时性和经济性的要求，取得好的效益。

(3) 抽样方法。抽样方法一般分为放回抽样和不放回抽样。放回抽样也叫重置抽样或重复抽样，它是在总体 $N$ 个单元中随机抽取 $n$ 个单元时，每次抽取一个单元进行记录后又放回原来的总体，参加下一个单元的抽取，即下一个单元仍然在原来的全部抽样单元中抽取，以此类推，直到抽足所需单元数为止，因而同一个抽样单元有被重复抽中的可能。不放回抽样也称为不重置抽样或不重复抽样，它是在每次抽取一个新的单元之前，将已抽中的单元不再放回原来的总体，下一个单元的抽取在剩余的抽样单元中进行，以此类推，直到抽足所需单元数为止，因而每个抽样单元最多只能被抽中一次，不可能重复被抽中。

放回抽样与不放回抽样相比，不放回抽样的样本代表性优于放回抽样。因为放回抽样中，有些单元有被重复抽取的可能，从而使样本单元数在总体中的散布面缩小，样本的代表性减弱，故在实际工作中常采用不放回抽样。有鉴于此，在本书以后内容中，如没有特别的声明，则一般只涉及不放回抽样。理解了不放回抽样的方法及有关内容，也就容易理解和掌握放回抽样的方法。

以上三种影响因素中，差异程度的大小是由事物内部和外部联系决定的，是客观性的因素，人们只能认识了解，不能改变。抽样单元数的多少和抽样方法是人们可以选择和控制的，为主观因素，只要掌握和控制了这两个因素，在一定程度上，人们也就能控制样本的代表性，以期达到抽样数目尽可能小，使估计和推断结论达到预定的精确程度和可靠程度的要求。另外，等概率抽样与不等概率抽样相比，以不等概率抽样的样本代表性较等概率抽样为好。

## 抽样调查

### (五) 样本容量与样本可能数目

样本中包含的抽样单元个数称为样本容量，又称样本含量或样本大小，简称样本量。总体中所含抽样单元个数称为总体容量，样本容量与总体容量之比为抽样比，用 $f$ 表示，即 $f = n/N$ 。

样本可能数目则是在容量为 $N$ 的总体中抽取容量为 $n$ 的样本时，所有可能被抽中的不同样本的个数。用 $Q$ 表示。

当 $N$ 和 $n$ 一定时， $Q$ 的多少与抽样方法有关，其计算方法如表1-1所示。

表1-1 不同抽样方法和条件下样本可能数目

抽样方法	放回抽样	不放回抽样
考虑顺序	$Q = N^n$	$Q = P_N^n = \frac{N!}{(N-n)!}$
不考虑顺序	$Q = C_{N+n-1}^n$	$Q = C_N^n = \frac{N!}{n!(N-n)!}$

正确理解样本可能数目的概念，对于准确理解和把握抽样误差的计算、样本统计量的抽样分布、抽样估计的优良标准等一系列理论和方法问题都有十分重要的帮助。

### (六) 参数与统计量

在抽样调查中，通常将反映总体数量特征的综合指标称为总体参数。设总体单元数为 $N$ ，各单元的标志值为 $Y_1, Y_2, \dots, Y_N$ 。常见的总体参数主要有：

(1) 总体总值 $Y$ ：也称总体总量。如某地区粮食总产量。其数学表达式为

$$Y = \sum_{i=1}^N Y_i = Y_1 + Y_2 + \dots + Y_N \quad (1.1)$$

(2) 总体均值 $\bar{Y}$ ：也称总体平均数。如某地区粮食平均亩产。其数学表达式为

$$\bar{Y} = \frac{Y}{N} = \frac{1}{N} \sum_{i=1}^N Y_i \quad (1.2)$$

(3) 总体比例 $P$ ：是总体中具有某种特性的单元数目所占比重。如产品的合格率。

$$P = \frac{\sum_{i=1}^N Y_i}{N} \quad (1.3)$$

其中， $Y_i$ 是示性变量，当第 $i$ 个单元具有某种特征时， $Y_i = 1$ ，否则 $Y_i = 0$ 。

(4) 总体比率 $R$ ：是总体中两个不同指标的总值或均值的比值。如总收入与总支出之比。

$$R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}} \quad (1.4)$$

其中,  $Y, X$  表示总体中两个指标的总值;  $\bar{Y}, \bar{X}$  表示总体中两个指标的均值。

当总体范围确定后, 总体参数就是客观存在的, 但是未知, 需要通过抽样调查, 根据抽样调查结果对总体参数进行推断。

在抽样调查中, 利用调查所得的样本数据构造统计量, 对总体目标进行估计。设样本单元数为  $n$ , 各单元的标志值为  $y_1, y_2, \dots, y_n$ 。简单随机抽样中与总体参数相对应的统计量有:

(1) 样本总值

$$y = \sum_{i=1}^n y_i \quad (1.5)$$

(2) 样本均值

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (1.6)$$

(3) 样本比例

$$p = \frac{\sum_{i=1}^n y_i}{n} \quad (1.7)$$

其中,  $y_i$  是示性变量, 当样本第  $i$  个单元具有某种特征时,  $y_i = 1$ , 否则  $y_i = 0$ 。

(4) 样本比率

$$r = \frac{y}{x} = \frac{\bar{y}}{\bar{x}} \quad (1.8)$$

其中,  $y, x$  表示样本中两个指标的总值;  $\bar{y}, \bar{x}$  表示样本中两个指标的均值。

统计量是样本的函数, 是随机变量, 其结果取决于抽样设计和被选入样本的总体基本单元的特定组合。统计量的真正价值并不在于自身的结果数值, 而是在于提供有关总体参数的信息。

### 三、抽样调查的种类

按抽取样本的具体原则不同, 抽样调查可以分为概率抽样和非概率抽样。

#### (一) 概率抽样

概率抽样也称随机抽样, 即在抽选样本的过程中, 完全按随机原则进行, 使抽样过程不受主观干扰, 每个总体单元被选中与否与其入样概率大小有关。因此, 概率抽样所得到的样本对所研究总体具有一般代表性, 利用概率抽样结果可以推断总体特征, 也可以计算并控制抽样误差。所以, 概率抽样是一种科学的抽样方法。使用概率抽样所得到的样本称为随机样本。

概率抽样有两条基本准则: 其一, 单元是随机抽取的; 其二, 调查总体中的每个单元

## 抽样调查

都有一个非零的入样概率，并且能计算出这些概率。在概率抽样中，如果总体中每个单元被抽中的概率都相等，则为等概率抽样；如果每个单元被抽中的概率不完全相等，则为不等概率抽样。例如：在一个盒中放入红、黄、蓝色球各一个，随机从中抽取一个，则各色球被取到的概率均为三分之一，这就是等概率抽样；而如果在盒中放入红、黄色球各一个、蓝色球两个，从中随机抽一个，则抽到红、黄色球的概率各为四分之一，而抽到蓝色球的概率为四分之二，这就是不等概率抽样。

概率抽样按其组织方式不同，可分为简单随机抽样、分层抽样、等距抽样、整群抽样、多阶段抽样等不同类型。

(1) 简单随机抽样。简单随机抽样也称为纯随机抽样，是从总体的 $N$ 个抽样单元中，每次抽取一个单元时，使每一个单元都有相等的概率被抽中，连续抽取 $n$ 次，以抽中的 $n$ 个单元组成简单随机样本。

简单随机抽样是一种一步抽样法，是所有概率抽样的基础，也是评估其他抽样策略效率的基准。与其他抽样技术相比，简单随机抽样的特点在于：这种抽样方法是最简单的抽样技术；抽样框不需要其他（辅助）信息，唯一需要的只是一个关于调查总体所有单元的一个完全的清单和与其如何联系的信息；关于样本量的确定、总体参数估计与方差估计都有现成的标准公式可以利用，技术发展已经成熟。但实际应用中也存在一定的局限性：抽样框中即使有现成的辅助信息也不加利用，这使得与其他利用辅助信息的样本相比，估计效率较低；由于样本在总体中的地理分布范围比较广，如果采用面访，费用较高；有可能抽到一个“差的”样本；如果不用计算机，而用随机数表抽一个大样本将十分单调劳神。

(2) 分层抽样。分层抽样也称类型抽样，是实际工作中最常用的抽样技术之一。它是在抽样之前，先将总体按一定标志划分为若干个层（类），然后在各层内分别独立地进行抽样。由此所抽得的样本称之为分层样本，各层所抽得的样本也是互相独立的。如果每层中的抽样都是简单随机的，则这种抽样就称为分层随机抽样，由此所得到的样本称作分层随机样本。

采用分层抽样时，总体被分为同质的、互不重叠的若干层（子总体），然后在每一个层中独立地抽取样本，对每个层进行抽样可以使用任何一种抽样方法。为充分利用分层抽样的效率，层内必须有高度的同质性，即同一层内的单元的调查指标应是相近的，而不同层之间的差异应尽可能大。

分层抽样的优点是能提高总体参数估计值的精度；能保证样本对被定义为层的那些子总体的代表性，既可以对总体参数进行估计，也可以对各层的目标量进行估计；操作与管理方便；能避免得到一个“差的”样本；在不同的层中可以使用不同的抽样框和不同的抽样方法。分层抽样的缺点是要求抽样框中的所有单元都必须有能用于分层的辅助信息；与简单随机抽样相比抽样框的创建需要更多的费用，也更为复杂。

(3) 整群抽样。整群抽样是先将总体各单元划分成若干群（组），然后以群为抽样单元，从中随机抽取一部分群，对中选群内的所有基本单元进行全面调查。确切地说，这种

抽样组织形式应称为单级整群抽样。之所以采用整群抽样，是因为抽选群能大大降低数据收集的费用，当总体的分布比较广且调查采用面访时更是如此；同时，从总体中直接抽选个体在实际调查中并不总是可行的（没有关于个体的抽样框）。

整群抽样的优点：由于样本不如简单随机样本那样分散，能大大降低数据收集的费用；当总体单元自然聚集成群（例如：住户、学校）时，整群抽样比简单随机抽样更容易；如果对于调查变量而言，群内单元差异较大，而不同群的差异较小，整群抽样策略比简单随机抽样的统计效率更高（例如为估计性别比采用按户的整群抽样）。整群抽样的缺点：若群内单元有趋同性，则整群抽样的统计效率比简单随机抽样低，但对这类效率的损失可通过增加群的抽取个数来弥补；在进行调查前通常不知道一个群内具体有多少个单元，所以无法提前知道调查总样本量；方差估计可能比简单随机抽样更为复杂。

(4) 等距抽样。等距抽样也称机械抽样，是将总体各抽样单元按一定的标志和顺序排列以后，每隔一定的距离（间隔）抽取一个单元组成样本进行调查的抽样方法。具体方法是：假设总体由 $N$ 个单元组成，并按某种顺序编上1到 $N$ 的号码，要在其中抽取容量为 $n$ 的样本，先在前 $K$ 个单元中随机抽取一个单元，以后每隔 $K$ 个单元抽取一个单元，由所有抽中的单元共同组成样本，这种抽样方法就是等距抽样，这里 $K$ 称为抽样间隔。可见，抽出了第一个单元就等于决定了整个样本。

等距抽样的优点：在事先没有总体单元名录的情况下也可以使用这种抽样方法，此时可以使用并构造一个概念抽样框（只需要单元的排列顺序），每间隔 $K$ 个单元抽一个单元，直到总体的末尾；与简单随机抽样相比，等距抽样样本的分布较好（这还取决于抽样间隔及名录是如何排列的）；等距抽样比简单随机抽样简单。等距抽样的缺点：如果抽样间距正好碰上总体变化的某种未知的周期性，就会得到一个“差的”等距样本，从而影响抽样精度；在使用概念框时，不能预先知道最终样本量；抽样方差没有一个无偏的估计量。

(5) 多阶段抽样。前面介绍的几种抽样方式，均是从总体中通过一次抽样，就可产生一个完整的样本，这类抽样可称为单阶段抽样。但抽样调查的实践中，常常面对的总体规模庞大，而且分布在广大区域内，很难通过一次抽样产生完整的样本。因此应是根据实际情况将整个抽样程序分成若干个阶段，逐阶段地进行抽样，以完成整个抽样过程。多阶段抽样就是用两个或更多连续的阶段抽取样本的过程。第一阶段抽取的单元称为初级或一级（阶）抽样单元（PSU），第二阶段抽取的单元称为次级或二级（阶）抽样单元（SSU），以此类推可定义更高级（阶）抽样单元。

在实践中如果是先从总体中随机抽取一部分一级单元，然后再从被抽中的一级单元内随机抽取部分二级单元并对它们进行全面调查，则称为是二阶段抽样。如果在被抽中的二级单元中，再抽取部分三级单元组成样本，并对抽中的三级单元进行全面调查，这就是三阶段抽样。类似地，有四阶段抽样或更高阶的抽样，通常将二阶段及以上的抽样称为多阶段抽样。二阶段抽样是最常见的多阶段抽样，例如，第一阶段抽样用地域框抽地理小区（PSU）；第二阶段抽小区内的住户（SSU）。

## 抽样调查

多阶段抽样的优点：当群内单元对于调查变量是同质时，多阶段抽样比单阶整群抽样的效率高；样本分布相对集中，因此采用面访的旅行时间和费用都大为降低；不需要有整个总体的名录框，所要的只是在每个阶段抽样都有一个完整的抽样框。多阶段抽样的缺点：虽然多阶段抽样的效率比单阶整群抽样的高，但它没有简单随机抽样的效率高；通常不能提前知道多阶段抽样的样本量，因为在具体调查前，我们不知道在每个大单元中具体包含多少个下级单元（若固定每个大单元中的抽样数目，则总的样本量也可控制）；调查的组织较为复杂（比单阶整群抽样复杂）；估计值与抽样方差的计算较为复杂。

有些时候，概率抽样在实际中难以实现，如从海水中抽样，通常只限于一部分；从运煤货车中抽样，一般是从顶部进行等。此时，只能采用非概率抽样。

### (二)非概率抽样

非概率抽样也就是非随机抽样，是由抽样者根据个人的知识、经验和研究目的，在主观分析判断的基础上来进行抽样。由于在抽样过程中不完全按随机原则进行，带有一定的主观随意性，所以所抽到的样本对所研究总体不具有一般代表性，不能用来估计总体特征，也无法估计和控制抽样误差。用非概率抽样取得的样本称为非随机样本。

非概率抽样主要包括：

(1) 方便抽样。方便抽样是指在抽样时依据方便原则，以达到最大限度降低调查成本的目的。典型的形式是“拦截式”调查，如街边、公园或居民小区拦住行人实施调查，在商场门口或柜台向消费者进行调查等。方便抽样操作简单，能及时取得所需的信息，节省调查费用。主要局限是样本信息无法说明总体状况，不能根据样本信息对总体进行数量特征的推断，因而样本不适合描述性研究和因果关系研究。比较适合探索性研究，通过调查发现问题的，产生想法和假设，也可以用于正式调查前的预调查。

(2) 配额抽样。配额抽样是依据一定的标志将总体的各单元划分成若干类型，并按各类型在总体中所占的比例分配样本总量在各层的数额，以抽到各类型所需配额为止。从各类型中抽取样本的方法没有严格限制，一般采用方便抽样的方法抽取样本单元。配额抽样是美国的盖洛普（George Gallup）创立的，它是非概率抽样中比较科学的一种抽样方式。

(3) 判断抽样。判断抽样是指在抽取样本时，调查人员根据经验、判断和对研究对象的了解，有目的地选取一些单元作为样本，实施时根据不同的目的有重点抽样、典型抽样、代表性抽样等方式。抽样时不依据随机原则，因而调查结果不能用于对总体有关参数进行估计。

(4) 滚雪球抽样。滚雪球抽样是指调查中先从总体中找出满足调查条件的少数个体，通过这些个体了解其他个体，再由已了解到的个体去发现更多的个体，以此类推，了解到的个体越来越多，最后接近于总体。滚雪球抽样往往用于对稀少群体的调查，主要优点是容易找到那些属于特定群体的被调查者，调查成本比较低，适合对特定群体进行研究的资料搜集。