高等院校经济学管理学系列教材

Statistics

# 统计学（第三版）

## ——数据的搜集、整理和分析

主　编／孙允午

上海财经大学出版社

高等院校经济学管理学系列教材

# 统计学

## ——数据的搜集、整理和分析

（第三版）

孙允午　主编

上海财经大学出版社

**图书在版编目(CIP)数据**

统计学:数据的搜集、整理和分析/孙允午主编．—3版．—上海:上海财经大学出版社,2013.10

(高等院校经济学管理学系列教材)

ISBN 978-7-5642-1736-5/F·1736

Ⅰ.①统… Ⅱ.①孙… Ⅲ.①统计学-高等学校-教材 Ⅳ.①C8

中国版本图书馆 CIP 数据核字(2013)第 182085 号

□ 责任编辑 刘光本

□ 责编电邮 lgb55@126.com

□ 责编电话 021—65904890

□ 封面设计 张克瑶

TONGJIXUE

**统 计 学**

——数据的搜集、整理和分析

(第三版)

孙允午 主编

上海财经大学出版社出版发行

(上海市武东路 321 号乙 邮编 200434)

网 址:http://www.sufep.com

电子邮箱:webmaster@sufep.com

全国新华书店经销

上海华业装璜印刷厂印刷装订

2013 年 10 月第 3 版 2013 年 10 月第 1 次印刷

787mm×1092 1/16 19.25 印张 492 千字

(习题集:8.5 印张 217 千字)

印数:44 801—51 000 定价:42.00 元

# 前言

无论你从事何种工作、学习何种学科，甚至在每天的生活中，都要接触大量的数据。如何对这些数据进行处理并从中提取有用的信息，为你的学习、工作或生活做出更为正确的决策提供帮助，这就是统计学所研究的内容。

《统计学》是高等院校许多专业的基础课程。随着社会经济和科学技术的发展，上海财经大学《统计学》课程一直在不断地进行改革和完善。改革的目标就是要把原来以描述统计为主要内容的“社会经济统计学原理”和以推断统计为主要内容的“数理统计”两门课程统一，取长补短，以解决这两门课程一些内容的重复和定义的差异等问题，与国际上《统计学》课程的内容逐步接轨。为此，1995 年刘汉良教授主编了《统计学教程》，2001 年徐国祥、刘汉良、孙允午、朱建中编著了《统计学》教材，取得了很好的效果，多次获得国家教育部、财政部的重点推荐和上海市的优秀教材一等奖。经过多年的教学实践和形势发展变化的需要，我们深感《统计学》教材有必要做进一步的修改，以满足在现有的社会经济发展的背景和技术条件下学习本课程的需要，满足高校精品课程建设的需要。与前述的两本教材相比，本教材有如下一些特点：

1. 其体系已与国际上通行的《统计学》教材基本相同，涵盖了现代统计学描述统计与推断统计的全部内容。如在第一章中，除了给出了有关统计的一些重要术语外，重点围绕统计数据和数据的搜集展开。剔除了原有的统计发展史和统计学的研究对象等内容；在第三章中，增加了五数概括和箱线图等内容，使描述统计更趋完整；在第六章中，增加或完善了残差分析、回归模型的建立等内容，为学习“计量经济学”等后续课程打下基础。

2. 根据计算器和电脑已普及的现状，剔除了原有以手工计算为基础的推导和简化公式，给出基本公式后，直接用计算器或电脑进行运算，使统计应用更为便捷。同时，结合电脑软件的应用，大量增加了数据的展示和分析方法。

3. 每章增加了统计引例，力图克服原来举例过少的缺点，方便读者举一反三，更好地掌握统计在实际中的应用。对原有陈旧的、与我国现有的情况不相符的例题也进行了替换。

4. 受高等院校总课时的限制，《统计学》的课时也有减少的趋势。为了更加突出重点，本教材的总体内容有所减少，删除了原来“统计决策”这一章以及分层抽样、整群抽样的参数估计和指数因素分析等内容。对本教材删除内容感兴趣的读者，可查阅前述两本教材。

本书可供非统计专业的本科生使用，也可以供非统计专业的研究生或统计专业的本科生作为参考。设计的课时为每周 4 课时，一学期 17 周共 68 课时为课堂教学时间。若每周 3 课时，则可根据具体需要选择其中的大部分内容。

本教材自 2006 年出版以来，得到了广大读者和有关方面的认可，属上海市精品课程建设项目。经过多年的教学实践，我们发现原教材中存在一些问题，而且社会也在发展，因而有必要对本教材进行修订，以更适应教学之需。主要的修改内容是对原教材中的结构进行了调整，

增加和改变了部分例题以及更新了有关的数据等。此外，还增加了三套模拟试卷并附送统计学习题集，以供读者练习。

本书在编写过程中，得到了上海财经大学统计与管理学院许多同仁的热情相助，本人在此表示衷心的感谢。同时，还要感谢使用过本教材的各位老师和同学以及上海财经大学出版社的编辑，他们提出的宝贵意见和建议使本书更为完善。

**孙允午**
2013 年 9 月

# 目 录

# 第一章

# 统计和统计数据的搜集

## 统计引例

### 永美公司的顾客满意度调查

永美是一家通过电视进行商品直销的公司，追求给其顾客提供优质的服务和高质量的商品。为了加深对顾客的了解，永美要求其顾客填写一份满意度调查表，并寄回公司。调查表中有以下的一些问题：

- 从你订购商品起到收到这些商品的天数。
- 在未来的 12 个月内，你准备花多少钱去购买直销商品？
- 根据你最近在永美的购买情况，你对永美所提供的服务总体评价如何？
  1. 比预想的要好得多
  2. 比预想的要好
  3. 和预想的差不多
  4. 比预想的要差
  5. 比预想的要差得多
- 你如何评价最近在永美购买的商品的质量？
  1. 比预想的要好得多
  2. 比预想的要好
  3. 和预想的差不多
  4. 比预想的要差
  5. 比预想的要差得多
- 在未来的 12 个月内，你是否打算在永美购买其他商品？
  1. 是
  2. 否

请你对此次调查进行审核。此次调查的结果能提供何种类型的信息？永美将如何使用这种信息去提高现有的服务和商品的质量？你认为这个调查中还应包括哪些问题？

## 第一节 什么是统计

“统计”一词从字面上可理解为统而计之，说明了其应用的广泛性。它是根据英语 Statistics 意译而来的，有三个含义：统计工作、统计数据和统计学。

统计工作就是对统计数据进行搜集、整理和分析的过程。最早的统计可追溯到人类社会发展初期的计数活动。随着社会经济的发展，统计工作越来越频繁，也越来越重要了。

统计数据是统计工作所产生的成果，用以描述我们所研究现象的属性和特征，如统计图表、统计分析报告、统计资料汇编、统计年鉴等。统计工作的好坏取决于统计数据资料的数量和质量。

统计学是一门研究总体数量特征的方法论科学。它来源于统计工作，是统计工作及其成果的理论概括和总结；反过来它又指导统计工作，能帮助和促进统计工作日益完善和提高。

在统计学中，我们经常要用到总体、样本、参数和统计量这四个重要的术语，其定义如下：

总体是所研究的具有某些相同性质的全部单位或事件的整体。

样本亦可称为抽样总体，是从总体中抽取的部分单位所组成的整体，用以分析总体。

参数亦可称为总体指标，是综合测量整个总体的某个数量特征。

统计量亦可称为样本指标，是根据样本数据计算的综合测量值，可用以反映或估计、推断总体的某个数量特征。

例如，为了解我国人口的年龄、性别、民族等分布特征，2005 年在我国的所有人口中抽取 1%进行抽样调查。这时，所有具有中国国籍、居住在我国境内的人构成了总体，从中抽取的 1%的人口构成了样本，根据这个人口样本所测量到的人口年龄、性别、民族等分布特征即为统计量，据此推断的我国全部人口的年龄、性别、民族等分布特征即为总体指标。

根据总体所包含的单位是否有限，可分为有限总体和无限总体两类。如果一个总体所含的单位数为无限多个，它就是无限总体；否则，就是有限总体。社会、经济和管理方面的统计总体主要是有限总体，虽然有时这种有限总体所含的单位数目很大，如我国全部人口所构成的总体。

需注意的是，传统的统计学定义总体中所包含的单位具有各种不同的属性和特征，并称其为标志。本教材把原有的标志和单位的概念合并在一起，通称为单位。如指我国全部人口所构成的总体中每一个人，也可以指每个人的年龄，此时所有的年龄构成总体。

根据分析方法不同，统计学可分为描述统计学和推断统计学。描述统计学是关于搜集、展示一批数据并反映这批数据特征的各种方法，其目的是为了正确地反映总体的数量特点。推断总体学是根据样本统计量估计和推断总体参数的技术和方法。

## 第二节 为何需要数据

统计学要研究各种随机变量，对这些随机变量的观察所获取的数据，单位与单位之间或者人与人之间实际上总是存在着不同，而这些不同的数据中包含了我们所需的信息，这能帮助我们在许多场合中做出更为正确的决策。例如：

- 市场研究者需要对产品的特性进行评估，以区分不同的产品。
- 药品制造厂商需要判别一种新药是否比现在正使用着的药更有效。

- 生产部门的经理按惯例要检查生产过程，以检验其生产的产品质量是否符合公司的标准。
- 审计人员想通过查看某家公司的财务报表，以确认这家公司是否是依据了通行的会计准则做报表。
- 财务金融分析人员想判断在未来的五年中，哪些行业中的哪些公司最具有成长性。
- 经济学家想估计我国国内生产总值今年的增长速度。

在本章开始时的永美公司顾客满意度调查中，就表现了以上几个方面的应用。例如，永美公司的调查结果就是搜集到了数据，然后才能对数据进行分析，以根据需要评价标准、测量服务和产品的质量好坏以及协助制定出可供选择的行动方案。

在开始进行统计分析时，最重要的是识别出数据的来源是否合适，数据是否准确。若数据存在偏差、模糊不清或其他不准确的缺陷，即使采用很好的统计分析方法也不可能弥补这些缺陷。

## 第三节　数据的类型

统计数据是对我们所研究现象的属性和特征的具体描述，这种描述包括定性的文字描述和定量描述。为方便统计汇总分析，通常可对定性的文字描述进行数量化，用数字代码替代原来的文字表现形式。为了以后用不同的统计分析方法进行研究，必须搞清楚数据的不同类型。而从不同的角度看问题，数据就有不同的分类方法，形成不同的类型。同时，进行数据分类必须遵循两个重要的方法原则：(1)互斥原则，即每一个数据只能划归到某一类型中，而不能既是这一类，又是那一类；(2)穷尽原则，即所有被观察的数据都可被归属到适当的类型中，没有一个数据无从归属。

### 一、定性数据和定量数据

传统的统计学把数据划分为用文字描述的定性数据和用数字描述的定量数据。如一家企业的所有制形式可以是国有、私营、股份制和外资等，在本章的“统计引例”中消费者对永美所提供服务的总体评价等都属于文字描述的定性数据；而企业的净资产额、净利润额等，在本章的“统计引例”中消费者在永美从订购至收到这些商品的天数以及消费者准备在未来的 12 个月内花多少钱去购买家用电器等，都属于数字描述的定量数据。

### 二、离散型数据和连续型数据

若我们所研究现象的属性和特征的具体表现是相对固定的，则可称这种数据为常量；若其具体表现在不同时间、不同空间或不同单位之间可取不同的数值，则可称这种数据为变量。变量有离散型和连续型之分。离散型变量的数据是可列的，如一家公司的职工人数、某地区的企业数等。连续型变量的数据可以取介于两个数值之间的任意数值，如销售额、经济增长率等。定性数据只能是离散型的，例如，对问题：“你最近持有股票吗?”的回答，就限于简单的是或否。再如，对永美公司调查中的问题：“在未来的 12 个月内，你是否打算在永美购买其他商品?”的回答也是如此。定量数据既可以是离散型的，也可以是连续型的。如对“你现在订阅了几份杂志?”的回答是离散型的，对你的身高是多少米的回答，以及对在永美公司的顾客满意度调查中所问的问题：“在未来的 12 个月内，你准备花多少钱去购买直销商品?”的回答都是连续型的。

有些连续型变量在具体整理分析时，可进行离散化处理。例如严格地讲，人的年龄是一个连续型变量，因为从人们的出生时点到统计的时点是一个连续变量，但在实际统计工作中，往往是按实足年份进行离散化的处理。

对连续型变量的量度还受到测量工具的影响。例如，人们的身高是一个连续型随机变量，但由于测量工具的精确程度不同，某人的身高可能是 1.70 米、1.701 米、1.700 9米或者是 1.700 87米。从理论上讲，不可能出现两个人的身高是完全相同的这种情况。因为测量的工具越精确，就越有可能区分出他们身高的不同。但是，大部分测量工具都不是十分精确的，无法测出细小的差别。因而，即使随机变量确实是连续的，也经常会在试验或调查的数据中发现取值相同的观测值。

## 三、数据的四个层次

社会统计学往往把数据划分为四个层次，即把定性数据再细分为定类数据和定序数据，把定量数据再细分为定距数据和定比数据。

1. 定类数据，也称定名数据，这种数据只对事物的某种属性和类别进行具体的定性描述。例如，对人口按性别划分为男性和女性两类，数量化后可分别用 0 和 1 表示；对企业按所有制性质划分为国有企业、集体所有制企业、股份制企业、合资企业、私营企业、外资企业等，可分别用 1、2、3、4、5、6 等表示；在永美公司的调查中，消费者对未来的 12 个月内是否打算在永美购买其他商品的回答结果，可分别用 1 表示是、用 2 表示否。这种数码只是代号而无顺序和多少大小之分，不能区分大小或进行任何数学运算。定类数据形成各种类型，它们的排序是无关紧要的，哪一类在前、哪一类在后对所研究的问题并无实质性影响。而且，它们能够进行的唯一运算是计数，即计算每一个类型的频数或频率(即比重)。

2. 定序数据，也称序列数据，是对事物所具有的属性顺序进行描述。定序数据不仅具有定类数据的特点，将所有的数据按照互斥和穷尽的原则加以分类，而且还使各类型之间具有某种意义的等级差异，从而形成一种确定的排序。这种序列测定在社会经济管理工作中应用很广泛。例如，对企业按经营管理的水平和取得的效益划分为一级企业、二级企业等；对青年职工按所受正规教育划分为大学毕业、中学毕业、小学毕业等；在永美公司的调查中，消费者对永美所提供服务的总体评价等都属于定序数据。这种排序是确定的，对所研究的问题有特定的意义。但是，它并不能具体测定各等级之间的间距大小，例如不能计算一级企业和二级企业的有实质意义的量的差距。类似地，也不能计算服务质量比预想的要好与差不多之间的差距。

3. 定距数据，也称间距数据，是比定序数据的描述功能更好一些的定量数据。它不仅能将事物区分为不同类型并进行排序，而且可以测定其间距大小，标明其强弱程度。温度是典型的定距数据，如 10℃、20℃等。它不仅有明确的高低之分，而且可以计算差距，如 20℃比 10℃高 10℃，比 5℃高 15℃等。定距测定的量可以进行加或减的运算，但却不能进行乘或除的运算，其原因是在定距数据的数值之间虽有确定的间距，但是没有自然确定的原点，即它的零点是人为指定的，所以不能得出某天的最高温度 20℃比最低温度 10℃高出 1 倍的结论。

4. 定比数据，也称比率数据，是比定距数据更高一级的定量数据，它不仅可以进行加减运算，而且还可以作乘除运算。定比数据与定距数据的显著区别是它有一个自然确定的、非任意的零点，也即在数值序列中，零值是有实质意义的。例如，人的年龄、体重都没有负值，以零为

绝对界限，一个人的年龄不能比零岁更年轻，体重也不能比零更轻。因此，我们既可以说甲某人 60 岁，比乙某人 30 岁年长 30 岁，也可以说甲的年龄是乙的 2 倍。几乎所有的物理量都可以进行定比测定；绝大多数的经济变量也可以进行定比测定，如产量、产值、固定资产投资额、居民货币收入和支出、银行存款余额等。

上述统计数据四个层次的描述功能是依次增大的，因而它们的运算功能也是依次增大的，可概括为表 1－1。

**表 1－1　　四个层次统计数据的比较**

| 数据的层次 | 运　算 | 特　征 | 举　例 |
|---|---|---|---|
| 1. 定类数据 | 计数 | 分类 | 产业分类 |
| 2. 定序数据 | 计数 | 分类 | 企业等级 |
| | 排序 | 排序 | |
| 3. 定距数据 | 计数 | 分类 | 温度 |
| | 排序 | 排序 | |
| | 加、减 | 有基本的测量单位 | |
| 4. 定比数据 | 计数 | 分类 | 商品销售额 |
| | 排序 | 排序 | |
| | 加、减 | 有基本的测量单位 | |
| | 乘、除 | 有绝对零点 | |

统计数据的四个不同层次表明对不同研究对象定量分析的条件和形式是不同的，必须根据具体对象和问题加以区别。例如，对企业职工可以计算他们的平均工资和平均收入，但却不能计算他们的平均道德水平和平均政治信仰。掌握统计数据的不同层次，对于正确地分析数据和选择检验方法（参数检验和非参数检验）是十分必要的。

必须指出，统计数据四个层次的高低之分只是就客观事物量化程度和运算功能来说的，而不是指统计研究本身的高低之分。如果从客观对象量化分析的难易程度来看，定比数据和定距数据是对定量数据的测量，比较直接和容易，而定类数据和定序数据则是对属性的测量，量化过程就困难得多，特别是对多维的复杂现象和过程的测量就更加困难。例如，对科技创新和文化活动的测量比对生产活动的测量要困难；对经常困扰人们的各种原因引发的通货膨胀和国民经济运行的周期性波动的测量，显然比对产品产量和产值的测量要困难得多；对诸如贫困与富裕、生活质量、社会公平与进步、综合国力等社会和政治问题的定量分析，无疑比经济问题又要困难得多。

在实际问题中，定距数据使用的机会较少，而且在许多场合可以采用与定比数据同样的处理方法，通常可把两者合并在一起。如在统计软件 SPSS 中，分别用“Nominal”、“Ordinal”和“Scale”表示数据的三种类别，最后一种就是定距数据与定比数据的合并。本书此后就把这种合并后的数据统称为定量数据，只在特别需要时才加以区别。

## 四、截面数据和时间序列数据

为了对数据采用不同的分析方法，可根据数据所反映的时间特点分为截面数据和时间序

列数据。截面数据是所搜集的不同单位在同一时间的数据，时间序列数据是所搜集的同一总体或单位在不同时间的数据。例如，所有上市公司公布的2012年的年度净利润就是截面数据，而某公司公布的2002～2012年的年度净利润就是时间序列数据。

### 五、原始数据和次级数据

原始数据是指直接从各个调查单位搜集的、尚未经过整理的统计数据资料，也称一手数据。次级数据是指那些已经加工整理过的，往往是公开发表的数据，如从报纸杂志、统计年鉴、会计报表上取得的数据，也称二手数据。

## 第四节 数据的来源

数据的主要来源有以下几个方面：

1. 从政府机构、各种行业组织、公司和企业所公布的数据中获取。
2. 设计一次试验以获取必要的数据。
3. 从观察研究中获取。
4. 进行一次调查。

第一种搜集数据的方法就是把政府机构、各种组织和公司所公布的数据作为来源，这种数据往往是次级数据，而其他三种方法获取的数据往往是原始数据。各级政府的统计部门是数据的主要搜集和汇编者，而且要公布一些重要的数据，如国内生产总值和消费价格指数等。各个行业的数据由相关的部门或组织予以公布，如中国人民银行会及时公布金融统计数据。按照有关政策，上市公司须及时公布本公司的财务数据和其他重要的信息。此外，每天的媒体会发布关于股价、气候和各种文体活动的大量统计数字信息。

第二种搜集数据的方法就是实验。在一次试验中，对整个过程都要进行严格的控制。例如，在检验洗衣机洗净程度的研究中，研究人员不是去询问顾客他们所认为的哪种牌子的洗衣机洗衣效果最好，而是通过实际洗涤脏衣服，来研究哪种牌子的洗衣机效果最佳。正确的试验设计需在相应的后续课程中予以讨论，因为这要涉及到复杂的统计运算。然而，为了对检验和试验有所了解，本教材的第五到第七章中给出了试验设计的基本思想。

第三种搜集数据的方法就是通过观察研究。研究人员通常是在自然状态下进行直接的观察。如关于动物行为的大部分知识都来自于这种方法，而天文学和地理学因为很难进行实验和调查，只能通过观察进行研究。同样，商务和管理中的观察研究可搜集到大量有关的信息，用以帮助做出正确的决策。例如，观察路口的交通流量、观察顾客在商场的购买行为和观察流水线上的产品质量等。

第四种搜集数据的方法是进行统计调查。它对所调查的人们的行为不进行任何控制，仅提出诸如出生年月、爱好、消费习惯、对某一事件的看法和其他特征方面的问题，然后对他们回答的结果进行整理、编码、列表和分析。

统计调查是搜集社会经济原始数据的主要方法，在调查工作展开之前需编制一份统计调查方案。调查方案的主要内容应包含以下五项：确定调查的目的；确定调查对象和调查单位，调查对象就是想要进行调查的现象的总体；拟订调查的具体内容并设计出调查表，调查表有单一表和一览表两种形式；确定调查时间，即要明确调查的数据资料所属时间和调查工作的起讫时间；编制调查的组织计划。

数据来源和科学技术发展有密切的联系。现在，可以方便地获得大量及时和精确的数据，这要归功于信息技术的广泛使用。当产品在超市、百货商店和其他渠道被销售出去时，条形码自动地记录存货的数量。自动取款机（ATM）和其他网上银行使得交易能被及时记录下来。旅行机构有精确到最近一分钟的关于航班和旅馆的空位数据。十年前要花数小时甚至数天的交易，现在只要在瞬间便可完成。图书馆的使用也有了新的含义。人们不再限于诸如书籍和报刊等印刷出来的媒体。基于电脑的信息系统，如使用 CD 盘上的数据库、在国际互联网上冲浪，或与其他网络用户用电子邮件交换信息等，使你能通过电子技术方便地寻找到数据。

## 第五节　搜集数据的组织方式

### 一、普查、抽样、统计报表制度和重点调查

搜集数据的组织方式主要有普查和抽样。此外，还有统计报表制度和重点调查。

普查是为了某些特定的目的而组织的、对总体中的全部单位都进行的调查。普查通常是为了搜集某些不宜或不能用其他方法取得而准确性要求又比较高的全面统计数据资料，以掌握一个国家或地区的重要的国情国力，作为制定政策和长期发展规划的依据。普查存在的缺陷是需耗费大量的人、财、物力和时间，因此普查只能是间隔一段时间搞一次。我国现阶段组织的普查主要有十年一次的人口普查、五年一次的经济普查和十年一次的农业普查。

抽样是只对总体中的部分单位进行的调查，这部分单位的集合即称为样本。与对总体进行全面普查不同，统计抽样主要是选取一个对总体具有代表性的样本，抽取的样本可以提供用于估计整个总体特征的信息。

统计报表制度是按一定的表式和要求，自上而下统一布置，自下而上提供统计资料的一种统计调查方法。这种搜集统计数据方法是伴随着计划经济而产生的，并曾在我国占主导地位。现在，在社会主义市场经济条件下，仍是我国搜集统计数据的组织方式之一。

重点调查是对总体中的重点单位进行的调查。所谓重点单位，是指在总体中这些单位个数虽然较少，但它们的变量值在总体的变量总值中占有很大比重。通过对这些单位的调查，就能了解总体的基本情况。例如，要了解全国钢铁生产的基本情况，只要调查鞍钢、宝钢、首钢、武钢、包钢等十几家特大型的钢铁企业就可以掌握全国钢铁企业生产的基本情况。重点调查只有在总体中存在重点调查单位的情况下才可以采用，否则就不能进行重点调查。

### 二、抽样的优点

1. 适用的范围广。对于有限总体，从理论上讲，既可以进行普查也可以进行抽样；对于无限总体，就只能进行抽样。若理论上可以而实际上很难采用全面普查的情况，也只能采用抽样，如产品质量的破坏性检验、居民住户调查等。

2. 与全面普查相比，抽样最大的优点是节省人、财、物力和时间。

3. 随机抽样可以比普查更为精确。

### 三、抽样的类型

如图 1－1 所示，抽样有两种类型：非随机抽样和随机抽样。

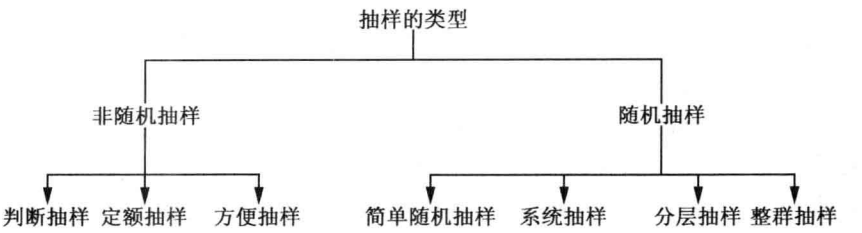


图 1—1 抽样的类型

非随机抽样是不按照随机原则来抽取样本中的单位或个体。随机原则是根据概率的基本原理，确保总体中每一个单位都有同等被选中的机会。因此，随机抽样又称为概率抽样，而非随机抽样又称为非概率抽样。因为非随机抽样在抽取单位时不考虑抽选的概率，发展到现在为止的概率抽样理论尚不能应用于这种情况。许多公司进行的调查是通过给那些浏览互联网的人们提供调查表。这种调查的结果能及时提供大量的数据。但是，这种样本包含的是那些愿意填写调查表的国际互联网用户。常见的非随机抽样有判断抽样、定额抽样和方便抽样。在许多研究中，只有判断抽样这类非随机抽样可以采用。非随机抽样具有方便、快速和低成本这样一些优点。另一方面，也存在着两个主要的缺点：因抽样的偏差而导致的精确性差，以及其结论缺乏普遍性，而这两个缺点已远远抵消了其优点。因而，应该尽量少用非随机抽样，除非你想以较低的成本获得大致的结果，或者想通过小规模的初步研究为更为复杂的调查做准备。

判断抽样又称为典型调查，是从事有关工作的专家按照一定的标准有意识地在总体中选择若干有代表性的单位组成样本进行调查，代表单位的选取标准应根据统计研究的目的而定。例如在编制物价指数时，对商品的调查既不是普查也不是随机抽样，而以选择代表性商品为宜。通常是在划分类别的基础上选取交易额大的若干种商品作为代表性商品。此外，也要考虑到商品在各年度交易中的连续性。判断抽样的调查结果可以用来说明总体，或作为总体的代表，但是其代表性如何，则取决于代表单位的选取是否合适。由于抽样过程没有依据随机原则，判断抽样的抽样误差不能准确地计算出来，而随机抽样的误差不仅可以计算，还可以把其控制在一定的范围之内。

定额抽样是根据已定的单位数抽取样本，往往是对总体了解甚少时采用。如想获取某地区化妆品的销售情况，对该地区的 5 家商厦进行调查。

方便抽样是为了取样方便，随意地抽取样本单位。街头偶遇式调查就是一种最为常见的方便抽样。

## 四、随机抽样

随机抽样是根据随机原则来抽取样本单位。实际问题中应该尽量采用随机抽样，因为到目前为止，这是根据样本对总体进行统计推断的唯一方法。随机抽样过程中首先要定义一个抽样框，抽样框中应含有总体中的全部单位，它可以是总体各单位的编号、名录或一份地图这样的数据源。样本是从抽样框中抽取的，若总体中的某些部分或单位没有包括在抽样框中，则抽样将是不准确和有偏的，并会导致错误的结论。

随机抽样有四种最为常见的组织方式：简单随机抽样、系统抽样、分层抽样和整群抽样。这些抽样方式的成本费用、精确程度和复杂性都各不相同，以下将对这些方式展开讨论。