

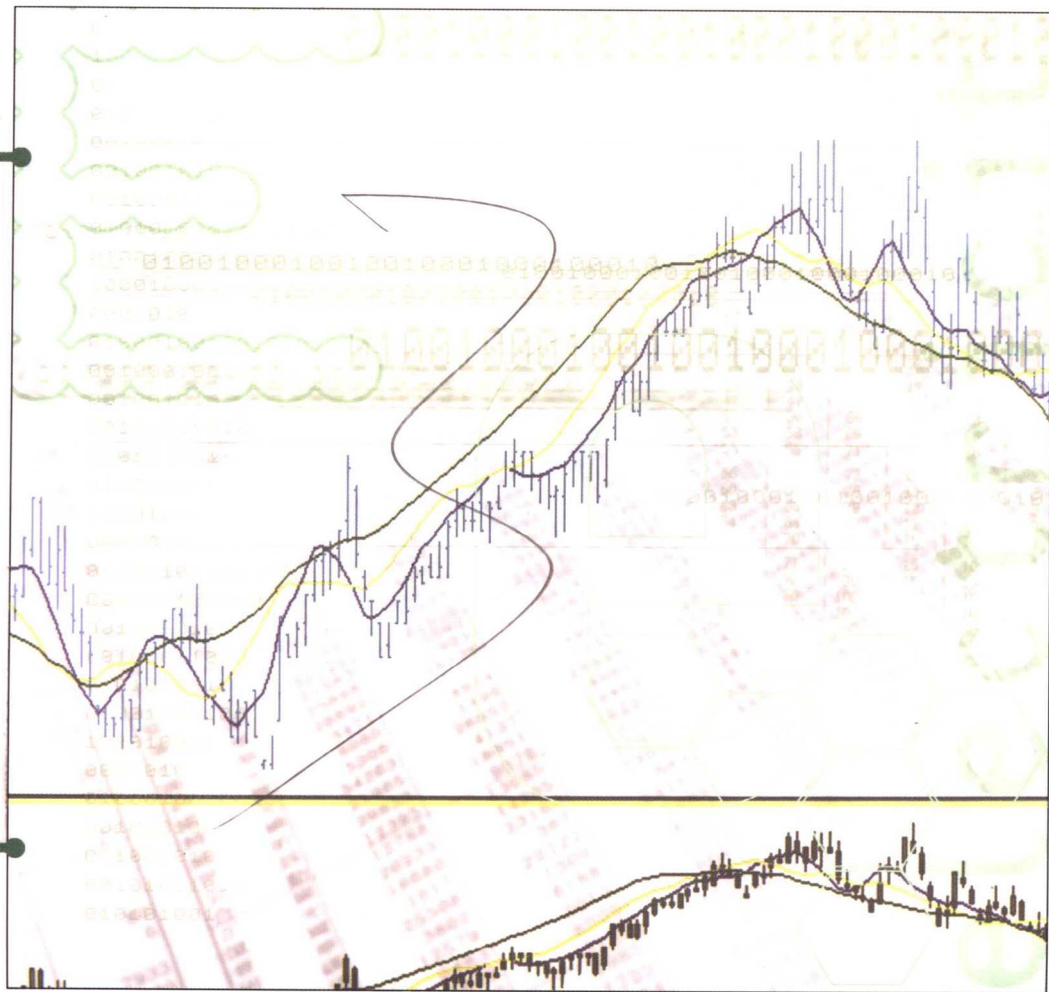


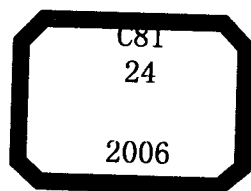
北京林业大学重点建设教材

统计数据分析理论与方法

陈建成 庞新生 主编

中国林业出版社





统 计 数 据 分 析 理 论 与 方 法

陈建成 庞新生 主编

中 国 林 业 出 版 社

图书在版编目 (CIP) 数据

统计数据分析理论与方法/陈建成, 庞新生主编. - 北京: 中国林业出版社, 2006. 10
ISBN 7-5038-4549-X

I. 统… II. ①陈…②庞… III. 统计数据-统计分析 (数学) IV. 0212.1

中国版本图书馆 CIP 数据核字 (2006) 第 093567 号

出版 中国林业出版社 (100009 北京西城区刘海胡同 7 号)

网址 www.cfph.com.cn

E-mail forestbook@163.com 电话: (010) 66162880

发行 中国林业出版社

印刷 北京林业大学印刷厂

版次 2006 年 10 月第 1 版

印次 2006 年 10 月第 1 次

开本 787mm × 960mm 1/16

印张 12

字数 215 千字

印数 1 ~ 1000 册

定价 26.00 元

前 言

统计数据与人人密切相关，我们每天都会和大量的数据打交道，都会用到相关的数据做出必要的决策。数据，是统计的语言；指标，是统计的手段；准确，是统计的生命；应用，是统计的灵魂。对统计数据若用之不慎，将比该死的谎言还糟糕。对真实的统计数据采用适当的方法进行科学分析，可以帮助我们正确的理解事物的发生根由、运行特征和发展态势，是人们对客观现实进行量化认识的前提。对统计数据的分析不仅要准确、及时、方便，而且更重要在于科学，因此对统计数据的分析方法不仅要正确，而且对数据的来源、质量的评估和系统的完整都要全面把握，再正确的方法，没有真实、及时、系统、科学的数据，那都是徒劳的。市场经济的理论和实践对统计数据的需求急剧增加，对统计数据的科学分析提出了更高的要求，统计量化分析研究已成为现代经济科学研究的重要手段。

统计学在分析数据、探索数据规律性、研究现实问题中已形成许多各具特色的思想方法，除一些基本统计方法与技术外，本书主要介绍一些比较实用、常用的统计数据分析理论与方法。现代统计分析方法的功能从研究问题的角度大致可归纳为：分类分析、简化结构、相关分析、预测决策分析。对于每种功能的分析都有其适应的方法，见下表所示。

具体统计分析方法与四大功能关系

方法 功能	回归分析	判别分析	聚类分析	主成分 分析	因子分析	典型相关 分析	对应分析
分类分析	—	+	+	—	—	—	—
结构简化	—	—	+	+	+	—	+
相关分析	+	—	—	+	+	+	+
预测决策分析	+	+	+	—	—	—	—

在实际应用中,对于解决什么样的问题,应该使用什么样的方法,或学习什么样的方法,能够解决什么样的问题,往往很难对应,经常使我们陷入困境。在本书编写过程中,笔者试图以案例研究为导向,以统计思想和方法应用为宗旨来解决这一困境。本书内容精练,方法较全,可读性强,是当前同类教材中比较系统的一本。本书主要针对已具备一些基本统计知识的高年级本科生和非统计类的研究生,为他们学习掌握比较高级的统计数据分析的方法提供参考。

本书的编写与出版得到北京林业大学教务处的鼎力支持,在此深表感谢!

本书由主编组织北京林业大学、中央财经大学和国家林业局有关专家共同编写,由于时间仓促,书中错误在所难免,请读者批评指正。

陈建成

2006. 8. 18

目 录

前 言

第一章 统计数据的获取	(1)
第一节 统计数据的来源	(1)
第二节 常用的抽样方法	(3)
第三节 抽样调查的基本程序	(6)
第二章 统计数据的描述	(9)
第一节 数据的计量和类型	(9)
第二节 品质数据的描述	(12)
第三节 数值数据的描述	(17)
第三章 统计数据的质量评估	(28)
第一节 数据质量的涵义与特征	(28)
第二节 统计数据质量评估	(29)
第四章 相关和回归分析	(44)
第一节 相关分析	(44)
第二节 回归分析	(47)
第五章 时间序列分析	(74)
第一节 时间序列分析简介	(74)
第二节 确定型时间序列分析	(77)
第三节 随机时间序列分析	(88)
第六章 聚类分析	(101)
第一节 相近程度的度量方法	(101)
第二节 系统聚类	(104)
第三节 k -均值聚类	(108)
第四节 聚类分析需要注意的问题	(108)
第七章 判别分析	(110)
第一节 距离判别法	(110)

第二节 费歇判别分析法	(113)
第三节 贝叶斯判别法	(116)
第四节 逐步判别分析法	(122)
第八章 主成分分析和因子分析	(129)
第一节 主成分分析	(129)
第二节 因子分析	(135)
第九章 对应分析	(140)
第一节 列联表和列联表分析	(140)
第二节 对应分析	(147)
第十章 指数分析	(151)
第一节 指数的作用及分类	(151)
第二节 指数的编制	(153)
第三节 常用指数	(157)
第四节 其他指数	(163)
附表1 正态分布的密度函数表	(172)
附表2 正态分布表	(173)
附表3 正态分布的双侧分位数 (u_α) 表	(175)
附表4 χ^2 分布表	(176)
附表5 χ^2 分布的上侧分位数表	(178)
附表6 t 分布表	(179)
附表7 F 分布上侧临界值表 ($p \{F > F_{0.05}\} = 0.05$)	(180)
附表8 F 分布上侧临界值表 ($p \{F > F_{0.01}\} = 0.01$)	(181)
参考文献	(182)

按照数据是否可以控制有关因素分为试验数据和观测数据；也可以根据数据获得的途径分为原始数据和二手数据。数据可以从总体中获取，但更多时候是从样本中抽取出来的，样本是总体的一个代表，样本有简单随机样本，也有方便样本，依抽样时采取的方法而定。

第一节 统计数据的来源

从统计数据本身的来源看，最初都是来源于直接的调查或试验。但是从使用者的角度看，数据则主要来源于两种渠道：一是来源于直接的调查和科学试验，这是数据的直接来源，我们称之为第一手数据或直接数据；二是来源于其他人的调查或试验的数据，这是数据的间接来源，我们称之为第二手数据或间接数据。本章主要从使用者的角度来讲述统计数据的搜集方法。

一、统计数据的直接来源

数据的直接来源主要有两个渠道：一是调查或观察，二是试验。调查是取得社会经济数据的重要手段，其中包括政府统计部门进行的调查，如人口普查，也有其他部门或机构为特定目的而进行的调查，如市场调查等；试验则是取得自然科学数据的主要手段。这里我们将集中讨论取得社会经济数据的主要方法。

1. 普查(Census)

普查是为某一特定目的而专门组织的一次性全面调查，如我国定期进行的人口普查、农业普查、工业普查、第三产业普查、统计基本单位普查等。世界各国一般都定期进行各种普查，以便掌握有关国情、国力的基本统计数据。普查是适用于特定目的、特定对象的一种调查方式，它主要用于搜集处于某一时点状态上的、不能够或者不适宜定期用全面统计报表搜集的社会经济现象的数据，目的是掌握特定社会经济现象的基本全貌，为国家制定有关

政策或措施提供依据。普查作为一种特殊的数据收集方式，具有以下几个特点：①普查通常是一次性或者周期性的；②普查一般要规定统一的标准调查时间，以避免调查数据的重复或遗漏，保证普查结果的准确性。③普查的数据一般比较准确，规范化程度也较高，因此它可以为抽样调查或其他调查提供基本的依据，特别是与抽样调查的数据资料相互验证，以提高调查质量，保证调查精度处于合理的范围之内。④普查适用的范围比较狭窄，只能调查一些最基本的、最一般的现象。

2. 抽样调查(Sampling Survey)

抽样调查是从调查对象的总体(Population)中抽取一部分单位作为样本(Sample)进行调查，并根据样本调查结果来推断总体数量特征的一种非全面调查方式。根据是否按照随机性原则进行抽样，又可以分为概率抽样调查与非概率抽样调查。抽样调查是实际中应用最广泛的一种调查方式和方法。它有以下几个特点：①经济性。这是抽样调查的一个最显著的优点。由于调查的样本单位通常是总体单位中的很小一部分，调查的工作量小，因而可以节省大量的人力、物力、财力和时间。②时效性强。抽样调查可以迅速、及时的获得所需要的信息。由于工作量小，调查的准备时间、调查时间、数据处理时间等都可以大大缩减，从而提高数据的时效性。③适应面广。抽样调查可以获得更广泛的信息，它适用于对各个领域、各种问题的调查。④准确性高。概率抽样调查的数据质量有时比全面调查更高，因为全面调查的工作量大、环节多、登记性(或调查)误差往往很大，而抽样调查则相反。当然，用样本数据去推断总体时，不可避免地会有推断误差，但这种误差的大小可以计算并加以控制，因此推断的结果通常是可靠的。

3. 统计报表(Statistical Report Forms)

统计报表是按照国家有关法规的规定，自上而下地统一布置、自下而上地逐级提供基本统计数据的调查方式。统计报表是搜集统计数据的一种重要方式，在我国几十年的政府统计工作中，已形成了一套比较完备的统计报表制度，它已成为国家和地方政府部门统计数据的主要来源。

二、数据的间接来源

对于大多数使用者来说，亲自去做调查往往是不可能或不必要的。他们大多使用其他人调查或者试验得到的数据，这就是我们前面提到的数据的间接来源或者称为二手数据。

二手数据主要是公开出版或公开通报的数据，当然也有些是尚未公开出版的数据。在我国，公开出版或者报道的社会经济统计数据主要有来自国家

和地方的统计部门以及各种报刊媒体,如《中国统计年鉴》、《中国社会统计年鉴》、《中国人口统计年鉴》、《世界经济年鉴》及《国外经济统计资料》等。

除了公开出版的统计数据外,还可以通过其他渠道使用一些尚未公开的统计数据;各种报刊、杂志、图书、广播、电视媒体也是我们搜集数据的主要途径。现在,随着互联网的发展,大量的电子版数据公布在各国政府部门和企业的网站上,充分利用搜索引擎,例如百度、google等,我们可以获取需要的各种数据,不仅可以节省大量的时间,而且还省去了录入的麻烦。但是,在使用从网络获取的数据时,一定要格外注意考察数据的真实性和完整性,以免造成分析结果的偏差和错误。

第二节 常用的抽样方法

抽样调查是社会经济生活中应用最广泛的获取数据的方法,样本抽取是否具有代表性直接关系到统计推断的优劣,如何抽取样本,大有讲究。

一、与抽样调查有关的基本概念

1. 总体与个体

我们将所研究的所有个体的集合称为总体。总体有大小之分,我们常常用总体容量去衡量总体的大小,该指标是指一个总体所包含的个体数目,总体容量越大,这个总体就越大。

2. 样 本

样本是按照某种原则(如随机原则、方便原则等)从总体中抽取的一部分个体所构成的集合,随机原则可理解为概率原则,即样本中的每一个个体都以一定的概率入样,当然有等概率和不等概率之分,等概率是指总体中每个个体被抽中的概率都相等,例如,从我们研究的儿童总体中随机地抽取30名,测量他们的体重、身高等;不等概率是指总体中每个个体被抽中的概率都不相等,例如在研究全国钢铁生产情况时,由于钢铁企业的规模差异很大,因此,应该按照不等概率原则抽选样本,大企业入样的概率比小企业要大。除此之外,还有方便原则,例如我们在研究一列货车运输的焦炭的质量时,只能通过容易取到的部分去判断整车的质量。

3. 统计量

在进行统计分析过程中,我们的最终目的是要把握总体在某一个或几个变量上的数字特征,例如,某地区全体儿童的平均体重、平均身高,或者在某个体重或身高范围内的人数比重等,上述这种就给定的变量由总体所计算

得来的数，称作总体参数。显然，为获得总体参数的具体值，我们应该做的第一件事就是面向总体所包含的个体搜集有关方面的数据，接下来再将搜集得到的数据进行加工计算得到想要得结果。

理想的情况是，就总体所包含的全部个体来搜集数据，进而经过加工计算得到反映总体数字特征的精确的参数值。但由于数据搜集成本的限制，通常情况下很难甚至无法做到这一点。现实的做法是，就总体中的一个样本来进行数据搜集，得到样本数据，由样本数据计算与总体参数对应的统计量，再由统计量的值来推断对应的总体参数。统计量是关于样本的函数，由于样本本身是随机的，从同一个总体抽出来的不同样本也不一样。因此，对于不同数据或样本，统计量的取值也不一样；也就是说样本的随机性决定了统计量的随机性。

二、常用的抽样方法

抽样方法很多，归根结底可以分成两类：随机抽样和非随机抽样。

(一)随机抽样

随机抽样又称为概率抽样，是根据一个已知的概率选取被调查者，无需在抽取样本时进行判断或抽选。

随机抽样是一种科学的抽样方法，抽样误差可以估计，因此能够控制抽样误差。只要可能，就应该尽量采用。随机抽样的基本方法有：

1. 简单随机抽样

简单随机抽样是最基本的形式，它是所有其他抽样方法的基础。在理论上最易于处理，但在实际实施时通常会有很大的困难。主要原因是它需要一个对全部基本单元的完整抽样框且所得的样本单元相当分散，调查不便。因此在大规模抽样调查中很少单独采用简单随机抽样。

2. 分层抽样

将总体小的单元按某种原则进行划分成为若干个子总体，每个子总体称为层。在每层中独立进行简单随机抽样或其他抽样，这样的抽样就称为分层抽样。分层抽样的估计先对各层进行，然后再综合对总体参数进行估计。

分层抽样适用于调查本身既需要对总体进行估计也需要对局部(层)进行估计的情况。分层抽样实施和组织都比较方便。当层内单元指标差异较小，而层间单元指标差异较大时采用分层抽样可以大大提高估计的精度。例如在家庭调查中，将住户家庭按城市、农村以及不同职业分层，由于不同层家庭的收支水平和生活习惯相差较大，因而这样的分层抽样精度较高。

3. 整群抽样

若总体中的每个抽样单元可以分成若干次级单元, 抽样仅对初级单元抽, 若某个初级单元被抽中, 则调查这个单元中所有次级单元, 这种抽样方法称为整群抽样。这里的群就是指初级单元。整群抽样的缺点是效率不够高。由于一个群内的(次级)单元多少有点相似, 故对每个次级单元都进行调查会造成浪费, 故若按总样本量(按小单元)计算, 整群抽样的精度比直接对总体中所有次级单元进行简单随机抽样低, 但这可以通过适当地多抽样来得到弥补, 从而完全有可能在总费用相同的条件下获得更高的精度。

4. 多阶抽样

多阶抽样也称多级抽样。若初级单元内的次级单元相似程度较大, 如前面所说的那样, 调查所有次级单元会造成很大的浪费。此时一个自然的想法是在被抽中的初级单元中再对次级单元进行抽样, 这就是二阶抽样。类似的, 可用三阶抽样、四阶抽样等。例如在全国抽省、省中抽市、县, 市、县中抽区、乡或镇等。多阶抽样既保持了样本相对集中, 又避免了不必要的浪费, 而且实施也比较方便。它也不需要每个初级(或二级)单元都有一个完整的抽样框, 但多阶抽样的估计比较困难。

5. 系统抽样

若总体中的抽样单元按某种次序排列, 在规定的范围内随机抽取一个(或一组)初始单元然后按一套事先确定的规则确定其他样本单元的抽样方法称为系统抽样。与其他几种抽样不同的是: 这里只有初始单元是随机抽取的, 其他样本单元都随着初始单元的确定而确定。最简单的系统抽样是在取得一个初始单元后, 按相等的间隔抽取其他样本单元, 这就是所谓的等距抽样。系统抽样的主要优点是实施方便, 不需对所有样本单元进行随机抽取, 也不一定需要一个完整的抽样框。如果对总体单元的指标按其排列次序的变化规律有所了解, 并加以合理利用的话, 系统抽样的效果也很好。它的主要缺点是在多数情形得不到估计量的简单的精度估计。事实上, 许多实用而行之有效的系统抽样并不属于严格的概率抽样。

上面对几种常用的抽样方法作了简单的介绍, 在实际运用中会有许多变化。例如在某些方法中, 入样单元既可不放回也可放回; 可以进行等概率抽样也可进行不等概率抽样。在具体设计抽样方案时, 还要考虑多种复杂的因素。

(二)非随机抽样

非随机抽样不是完全按随机原则选取样本的, 常用的非随机抽样方法主要有以下几种:

1. 方便抽样

方便抽样就是按抽样者的方便, 随意地抽取样本。例如, 调查员在大街上, 商店里随意地寻找行人或顾客进行调查。这种抽样方法的优点是简便易行, 但是, 它受抽样者主观意志支配, 很难保证样本的代表性。

2. 定额抽样

定额抽样就是依据一定的标志将总体分成若干层(若干部分), 确定各层在总体中所占的比例。并按这些比例分配样本总量在各层的数额, 让调查员抽到每一层所需的“定额”为止。这种方法在民意测验和市场调查中用得较多, 例如, 盖洛普在美国总统竞选的预测中采用的就是定额抽样法。

3. 经验抽样

经验抽样就是抽样者根据自己的经验抽取具有平均水平的、典型的单元作为样本。这种方法可以利用抽样者所掌握的信息, 避免发生极端的偏误。如果总体的变异程度较大, 又不希望样本量太大, 而抽样者是富有经验的, 经验抽样可能获得较高的精度。但是, 这种方法受主管因素的影响, 不但由于评价标准不同, 意见难以统一, 而且无法避免因抽样者的主观爱好和认识水平所造成的严重差错。

4. 滚雪球抽样

抽样者先找出少数个体, 通过这些个体了解其它个体。再由已了解到的个体去发现更多的个体, 以次类推, 这就像滚雪球一样, 了解到的个体越来越多, 越接近总体。例如, 对吸毒者的调查就可以采用这种方法。容易看出, 只有当个体之间能够并且愿意联系时, 这种方法才能实行。

与随机抽样方法相比, 非随机抽样方法的缺点在于无法估计抽样误差, 因此不能控制抽样误差。确定了抽样方法后, 我们就涉及到调查方式的选择问题, 常用的调查方式主要有访问调查、邮寄调查、(计算机辅助)电话调查、群组座谈会和个别深度访问, 以及直接观察法和试验法等方法; 其中访问调查、邮寄调查、(计算机辅助)电话调查都需要设计调查问卷, 以问卷为基础来进行调查, 因此, 这几种方法又可以统称为问卷法。关于调查方式的介绍, 国内外相关著作都作了大量介绍, 这里就不再赘述。

第三节 抽样调查的基本程序

对抽样调查, 不同的项目所包含的步骤也不尽相同, 但大致上都包含以下几个重要的步骤:

(1) 明确调查的目的, 确定调查方式和所需估计的目标量。通过一次调

查要达到什么目的?调查哪些指标?需要估计哪些目标量?都是首先需要明确的。因为调查的具体形式和组织,抽样方案的制定以及调查数据的处理都取决于调查的目的和调查的目标量。抽样方案确定后,调查目标量的任何改动,往往会使已制定的方案不再适用。因此在此阶段,主持单位必须会同有关专家进行反复讨论和审定。在这一步骤中,首先要确定总体范围及抽样单元。这个问题有时并不简单,例如对残疾人的调查,首先要明确残疾人的划分标准。在确定需要估计的目标量也即总体参数时,要注意防止列入过多的调查项目(指标)。项目过多,不仅会增加调查和以后数据处理的费用和时间,还可能使不回答率增加,并影响原始数据的质量,因而是得不偿失的。在这一阶段还必须同时确定调查的方式是采用当面询问还是通过调查表或者是两者结合?对于需要技术测试的调查,还要确定测试或分析的方法。对调查的目标量应提出具体的精度要求,作出调查的经费预算,确定调查的标准时刻等。对调查表,应在认真设计的基础上征求有关专家的意见反复修改力求完善。许多情况下,在正式调查前往往还需进行一次试调查。

(2)抽样设计,给出相应的数据处理公式。这是一个抽样调查中总体设计的最重要部分,包括选择抽样方案的类型,确定样本的抽取方法及样本单元。在制定具体的抽样方案时,既要考虑方法的科学性又要照顾到实际的可行性。例如设计一个全国性抽样调查,需要考虑一个多阶抽样,此时前一、两阶抽样是关键的,必须采用一些效率高的抽样方法,复杂一些也无妨,因为前一两阶的抽样可以由设计者自己来实施。与此相对最后一、两阶抽样则由于涉及基层就必须采用尽可能简单的抽样方法。在制定抽样方案时,还必须同时考虑到调查以后的数据处理问题,给出与抽样设计相匹配的总体参数的估计公式以及估计量的精度公式。

在抽样方案确定以后,就要实施抽样,即确定需要调查的样本单元。为此,事先要准备好相应的抽样框及其他有关资料。

(3)调查的实施,即取得样本数据的过程。为保证调查的质量,确保原始数据尽可能正确,应建立相应的职能机构,事先进行调查员的培训,制定并采取各种质量控制措施等。

(4)数据处理。现代大规模的抽样调查所获得的数据,一般都在电脑上进行处理。首先是编码并录入数据,建立数据库。在正式处理前,要对已录入的数据进行编辑加工按一定规则检出并处理原始数据中存在的或在录入过程中混入的异常数据,检查方法应同时采用统计及逻辑检查。经过反复检查的数据即可做进一步的处理,其中最重要的是按在抽样设计时给出的总体参数的估计公式与估计量精度公式,计算每个目标量的估计值及其相应精度,

特别是方差与变异系数的估计值，有时还需要结合分析目的进行其他统计处理，例如列联表分析与多元分析等。

(5)分析调查结果，提出最终报告。根据数据处理的结果以及调查目的，对调查结果进行综合分析，提出最终的调查报告。

除上述的步骤外，实行一项抽样调查还有相当多的行政组织与部门的协调工作。例如，根据我国的统计法，任何一项统计调查如果在本系统内进行，要在相应的统计部门备案；如果调查对象涉及本系统以外的部门或人员，要向相应的统计部门申请，在得到批准后，方可以进行，否则，调查对象有权不予回答。因此，在进行抽样调查之前要履行必要的手续，使调查合法并使调查结果受到法律保护。

完成了对统计数据的收集，接下来的任务就是对这些数据进行加工整理，使其条理化、系统化；进一步，还可以为这些数据绘制图表。通过这样的描述就可以大大简化数据，使我们可以比较容易地发现数据中所包含的信息，为统计分析提供思路。

第一节 数据的计量和类型

一、数据的计量尺度

统计数据是对客观现象进行计量的结果，因此，在搜集数据时必然要对现象进行计量或测度。根据计量学的一般分类方法，按照对事物计量的精确程度，可将所采用的计量尺度从低级到高级、从粗略到精确划分为定类尺度、定序尺度、定距尺度和定比尺度四个层次。采用不同的计量尺度可以得到不同类型的统计数据，而不同类型的统计数据又适用于不同的统计分析方法。

1. 定类尺度(Nominal Scale)

定类尺度也称类别尺度或者列名尺度，它是按照事物的某种属性对其进行平行分类或分组的一种测度，主要用于非参数的统计推断。生活中，用这样的计量尺度来测度的事物很多，例如性别、民族、国籍、商品的品牌等；该计量尺度是最粗略、计量层次最低的计量尺度，是其他计量尺度的基础，主要特征体现为：

(1) 只能区分事物的类别，但无法比较类别间的大小，也就是说定类尺度只具有“=”或“ $\neq$ ”的数学特征。

(2) 对事物的区分必须符合穷尽和互斥的要求。类别穷尽是指在所做的全部分类中，必须保证每一个元素或者个体都能归属于某一类别，不能有所遗漏；类别互斥则是指每一个元素或者个体只能归属于一个类别，而不能在

其他类别中重复出现。

(3)对定类尺度数据进行分析的统计量主要是频数或者频率。虽然定类尺度计量的结果只能表现为某种类别,但是为了在统计处理中的方便,特别是便于计算机识别,通常都用数字对不同的类别进行编码。

2. 定序尺度(Ordinal Scale)

定序尺度又称顺序尺度或有序水平,是对事物之间等级差或顺序差别的一种测度。受教育水平、职称、商品的质量等级、人们对某一事物的态度等都是用这种计量尺度来测度的;其计量精度要优于定类尺度,主要特征体现为:

(1)不仅能区分事物的类型,而且能够比较各类型间的优劣和顺序,不仅具有“=”或者“ $\neq$ ”的数学特征,而且有了“ $>$ ”和“ $<$ ”的数学特征。

(2)对事物的区分同样要求穷尽和互斥。

(3)对定序尺度的数据进行分析的统计量主要是频数和累积频数(Cumulated frequency),或频率和累积频率。同样,我们需要对定序尺度计量的结果进行编码以方便计算机的读取和统计运算。但与定类尺度不同的是,在分析中,我们不仅可以计算得到不同受教育水平的人数,而且还可以计算在某一水平之下或者之上的人数。

3. 定距尺度(Interval Scale)

定距尺度也称为间隔尺度或间隔水平,是对事物类别或者次序之间间距进行的一种测度。常见的用定距尺度来测度的如考试成绩、各种心理测试的得分、某个地区的温度等;它是一种较定类尺度和定序尺度更为高级、更为精确的计量尺度,其主要特征体现为:

(1)不仅能区分事物的类别、进行排序、比较大小,而且还可以精确地计量大小的差异。定距尺度的计量结果为数值,并且可以计算差值,因而它的计算也就超越了只能比较相等或者大小的运算范畴,进而可以进行加、减运算了

(2)没有绝对零点,即可以以任意一个0为起点。这里的“0”表示一个数值,即“0”水平,而不表示“没有”和“不存在”。例如,一个被访者的统计学考试成绩为0分,并不表示说他没有考试或者没有任何统计学知识;同样,一个被访者的IQ得分为0,也不能说他是没有智商的。

4. 定比尺度(Ratio Scale)

定比尺度又称比例尺度或者比较水平,是对事物之间比值的一种测度,可用于参数和非参数统计推断。我们在日常生活中,大多数情况下使用的都是定比尺度。例如,年龄、收入、某地区每年的失业人数、犯罪人数等。它