学校的理想装备

电子图书·学校专集

校园网上的最佳资源

应用统计学

BOOK

内部资料 非卖品

编 者 的 话

统计学是研究不确定性现象数量规律性的方法论科学，也是对客观现象进行定量分析的重要工具。在工商业管理中经常会遇到原材料的供应、产品质量状况和市场的销售前景等不确定性现象，要进行科学的管理和决策，就要应用统计方法。

统计学可以分为理论统计学和应用统计学两类。前者侧重于统计方法的数学理论，后者侧重于统计方法在各个领域的应用。本书的重点是统计方法的应用，特别侧重于在工商企业经济管理方面的应用，但在内容上也兼顾到统计方法在宏观经济管理中的应用，即国民经济核算方法。由于本书侧重于统计方法的应用和案例分析，因而本书不仅可作为经济与管理类大学 MBA、本科和专科学生的教材，也可作为广大实际管理工作者和研究工作者的参考书。

本书第 1、8、9 章由倪加勋执笔，第 2、3、4、10、11 章由袁卫执笔，第 5、6 章和第 7 章的前 3 节由易丹辉执笔，第 7 章的第 4 节和第 12 章由蔡志洲执笔。全书由袁卫主编，倪加勋主审。

在本书的编写和出版过程中，中国人民大学工商管理学院副院长施礼明教授付出了大量劳动；中国人民大学出版社的徐晓梅同志对本书提出了许多宝贵意见；书中若干例题选自所列参考文献，在此一并表示感谢。由于我们的水平有限，书中错误在所难免，欢迎读者批评指正。

编者

1992 年 1 月

加方序言

本书囊括了应用统计学领域中的最新研究成果。目前，在管理学与管理科学领域里，应用统计学的模型得到了广泛的运用。例如，市场营销人员运用统计模型进行市场研究，生产管理人员运用统计模型探讨机器设备更新的时间，仓库管理人员则运用统计模型确定库存量的问题。因此，管理人员充分地掌握了应用统计学的知识后，会使得企业日常管理富于效率，生产趋于完善，从而增强了企业在世界市场上的竞争能力。

本书是中国人民大学与麦吉尔大学合作出版管理学丛书中的一部。这套丛书是加中大学管理教育项目第Ⅱ周期（CCMEPⅡ）中麦吉尔大学与中国人民大学交流项目里一项活动内容，由中国人民大学编写，共计 15 本。这套丛书的重要意义就在于它结合了中国的实际，符合管理教育中十分强调的理论联系实际的要求。

在本套丛书的研究、写作与编辑的过程中，中国的吉林大学、兰州大学与加拿大的卡尔顿大学和舍尔布鲁克大学也做出了重要的贡献。在此之机，谨以这六所相互合作的院校的名义，我愿向加拿大国际开发总署（CIDA）和中国国家教育委员会的鼎力支持表示衷心地感谢。在他们的大力帮助之下，我们的校际交流项目经历了由 1983 年开始的加中大学管理教育项目第Ⅰ周期的活动，经历了由 1988 年开始的加中大学管理教育项目第Ⅱ周期的活动，最后在 1991 年 9 月建立了中国的工商管理硕士学位（M—BA），以及本套丛书的出版。所有这些成就都表明中国的管理教育在不断发生变化。

中国人民大学强调教学与科研并重，因此，在这套丛书里反映出了作者们的研究成果，使广大读者开卷有益。实际上，中加大学管理教育项目中的一个主要目标就是，在科研与培训中形成这种乘数效应。

最后，我真诚地希望所有的教授与学生们对此书提出批评与建议。这将对开拓管理学与管理教育极为有益。

CCMPⅡ国家项目协调员
麦吉尔大学管理学院院长
华莱士 B.克劳斯顿博士

中 方 序 言

为了适应社会主义市场经济条件下企业管理的需要，培养德、智、体全面发展的务实型高级管理人才，必须改革目前的管理专业设置和课程体系。

管理专业的学生应认真研究中国经济建设与社会发展的方向和特点，跟踪现代管理理论和实践的发展趋势，学生们不仅在理论上要有所建树，而且要有较强的实际工作能力。为此，管理专业的学生在校期间，要系统地学习经济学、管理学、财政金融、会计、生产管理、市场营销管理、信息系统管理、国际工商管理，以及战略管理等相关管理学科的知识。

改革开放的发展，要求在实现管理现代化的过程中，必须大胆吸收和借鉴当今世界各国的一切反映现代社会化生产规律的先进经营方式和管理方法。为此，在中加大学管理教育项目进行第二周期活动的过程中，中国人民大学工商管理学院与加拿大麦吉尔大学管理学院相互合作、共同编审出版这套既适合中国国情、又吸收外国先进经营方式和管理方法的管理学丛书，以期推动并完善中国工商管理硕士课程的建设。

在与加拿大麦吉尔大学管理学院友好合作的过程中，我们衷心地感谢该管理学院院长 W.B.克劳斯顿博士、副院长那格博士，以及其他编委为本丛书的顺利出版所做出的贡献；感谢该院其他朋友们在本丛书的编辑出版过程中给与的诚挚合作；最后，我们还要感谢加拿大国际开发总署通过麦吉尔大学为本丛书的出版所给予的财务资助。我们殷切地期望中加大学管理教育项目会顺利而持久地开展下去，并在更广阔的领域里获得更大的成功。

中方编辑委员会
1992 年 6 月 11 日

应用统计学

第 1 章　绪　论

§1.1　统计与统计学

§1.1.1　“统计”一词的含义

在介绍什么是统计学之前，很自然地涉及到“统计”一词的含义。在日常生活中“统计”有着多种含义。例如，开会时主持人要统计一下出席会议的人数；篮球比赛中教练员要统计每个队员的投篮命中率、犯规的次数；农户在农作物收获后统计其产量等等。这时“统计”是一个动词，它有记数的含义。在这个意义上统计的起源是很早的，从历史上看，早在古代奴隶主统治的国家，由于赋税、徭役、征兵等需要，就要掌握人口和土地等数字，据记载，公元前 3050 年，埃及建造金字塔，为征集建筑费，就有对全国的人口与财产调查。罗马皇帝凯撒·奥古斯都曾下过一道命令，要全世界向他纳税，于是要每个人都向就近的收税人登记；英国的威廉大帝下令测量英国的土地，其目的是为了征税和征兵役。我国春秋时期齐桓公任用管仲为相使齐国大治。在反映管仲思想的重要著作《管子》一书中就有这样的论述：不明于计数，而欲举大事，犹无舟楫而欲经于水险也。这就是说不善于利用计数而进行宏伟事业，犹如没有船和桨而想渡过激流险滩一样。可见在这个意义上，“统计”的应用十分广泛，而且是历来治理国家必不可少的一项重要工作。

统计工作的结果形成一系列的数字资料，也称统计资料或统计数据，这是“统计”的另一个含义。它和前面讲的统计工作是紧密相联的，是统计工作的结果，因此也是很早就有的。根据历史记载我国夏禹时代就开始有人口统计数字。春秋时期，《商君书》中指出：强国知十三数，这十三数包括粮食储备、人口及其各项分类数、农业生产资料以及自然资源等，不过当时还没有明确叫做统计资料罢了。随着社会的发展，需要的统计数字也就越来越多，现在只要打开报纸就可以看到各种各样的统计数字。国家统计局每年出版统计年鉴，反映国家的经济、文化教育以及科技发展等情况，这些都是在这个意义上的统计。

除了上面所讲的两个含义之外，“统计”一词还有另外的含义，即作为一门科学的“统计学”，它是本书将要探讨的主要内容，作为一门科学的“统计学”，它的出现要比统计工作和统计资料晚得多。

§1.1.2　统计学的产主

任何一门科学的产生都是与一定的社会背景和其他科学的相互影响分不开的。统计学这一名词最早来自欧洲，当时资本主义还处于萌芽时期，随着商业和手工业的发展，欧洲一些主要国家由于海外市场和殖民地的相继开拓，深感有系统调查国内外情况的必要。在德国和意大利就出现了“国势论”的著作，这些著作主要记载一个国家的地理、历史、政治组织、社会制度、商业和军事力量等。到 18 世纪中叶德国哥廷根大学教授阿亨瓦尔在他的著作《近代欧洲各国国势学纲要》一书的绪论中首次提出了统计学这一名词，他把统计学定义为国家显著事项之结晶体，并解释显著事项是由此可知国家理乱兴亡之迹。由于他最早提出统计学的这一名词，所以有人称他为统计学之父。但是他所称的统计学其内容只是文字的记载，而且限定统计学只研究现实问题，与今天统计学的内容相差甚远，只是在名称上沿用而已。统计学的另一个来源是英国的政治算术学派，英国是资本主义发展较早的一个国家，

也较早地利用数字对人口和经济方面进行记载和推断。其代表人物有格朗特(Graunt)。早在1661年，他在《对死亡表的自然观察和政治观察》一书中对当时英国情况的分析揭示出一系列的数量关系。如男婴出生多于女婴(14∶13)[①]，男性死亡多于女性，一切疾病和事故在全部死亡原因中占有稳定的百分比，等等。他在该书中指出，为了提出一个要在多年内形成的规律需要进行多次观察。因此格朗特虽然没有提到统计学这一名词，尽管他的方法也不完善，但是他在实践中却已应用了现代统计中的大量观察方法去发现规律性的问题，其独特和新颖的方法给人以启示，接着英国的威廉·配第曾写了著名的《政治算术》一书，书中用大量的数字对英、法、荷三国的经济实力进行比较，采用了与过去不同的传统方法，用数字、重量和尺度来表达他自己想说的问题。马克思对威廉·配第的评价很高，认为他“在某种程度上也可以说是统计学的创始人”[②]。但是在政治算术学派的著作中只是应用了数据，而并没有专门论述如何搜集数据，以及如何利用数据进行统计分析等，因此它并不是现代意义上的统计学。当今的统计学是继承了国势学派统计学的名称，内容上除了对国家重要事项的调查外又扩大了研究的范围，同时又吸取了政治算术学派对客观现象进行数字计量和大量观察的方法。但统计学的发展则是与数学的一个分支——概率论的产生与发展分不开的。16世纪以后欧洲赌博盛行，随着殖民事业的开拓，航海运输逐渐频繁，保险事业也随之兴起。赌博的输赢、航海及保险的赔赚都具有偶然的性质。这些偶然现象从个别来看似乎没有规律，但通过对大量现象的观察又可以发现他们具有一定的规律性。这样促使许多数学家从大量的偶然现象中寻找其规律性，逐步形成了概率论。比利时统计学家凯特勒(1796—1874)综合了国势学派和政治算术学派的成果并把概率论的原理和大量观察法引进了统计的研究领域，使统计学的发展进入了一个新的阶且。他认为：“统计学不仅仅是国势的记述，还应当把它作为学术问题来研究，因为统计学是对性质相同的事物进行大量观察，从而探索出社会现象相继不绝的一门学问。”[①]他还借助于一个生动的例子来说明大量观察法的作用。他说，人者分而观之，人心之不同各如其面，几无规律可循，然合人人而观之，则相去不远，其间自有不变的规律在。这说明他已意识到从大量的现象中寻找统计规律性。这正是现代统计学的核心。在他以后又有许多统计学家把统计方法应用于自然科学，出现了一些有重大影响的学者，如卡尔·皮尔逊、费雪、奈曼、伊·皮尔逊和瓦尔德等。他们在实践中进一步发展了统计理论，使统计学逐步形成一门比较成熟的科学。

§1.1.3　统计学是怎样一门科学

统计学是研究有关收集、整理和分析数据从而对研究的对象加深了认识并作出一定结论的方法和理论。下面对统计学的性质和特点作进一步解释。

1.统计学研究的对象是客观现象的数量方面。前面曾经谈到，早期统计所研究的问题有人口调查、出生与死亡的登记、保险业中赔款额和保险金的确定等，后来又扩大到社会经济和生物实验等方面。随着人类活动各种实践

[①] 格朗特发现施洗礼的男婴为139782人，女婴为130866人，为方便起见，简化为14万人和13万人，因此其比为14∶13（见参考文献[2]。）

[②] 《马克思恩格斯全集》第23卷，第302页。

[①] 《概率论书简》第4编《关于统计学》.见参考文献[2]。

的需要，各个领域都要研究事物的数量方面，以及密切联系数量方面来研究事物的本质。因此统计的应用越来越广泛。目前不论社会的、自然的或实验的，凡是有大量数据出现的地方，都要用到统计学。凡能以数量来表现的均可作为统计学的研究对象。统计方法已渗透到其他科学领域，成为当前最活跃的学科之一。

2.统计学研究的是群体现象的数量特征与规律性。客观世界是十分复杂的，但根据其不同的性质加以分类就形成各种群体，在统计学中把所研究的某类客观现象的群体称作总体。所以统计学所研究的是总体的数量特征及其分布的规律性。总体是由许多个体组成的，各个个体在数量特征上受必然和偶然二种因素的支配，必然因素反映了该总体的特征，但由于受偶然因素的影响又是有差异的，如何通过这些个体的差异来描述或推断总体的特征就产生了统计学。举一个例子来说，假如我们仅仅是要知道某一物件有多重，只要把它放在天平上一称就可知道，那么不需要学习统计学也能解决。进一步说，要了解一批物件的重量，而这批物件中每件的重量都是相同的，不存在差异，那么只要称其中的一件就可知道其他物件的重量，也不是统计学所要解决的问题。只有当要了解这批物件的重量，而其中每一件的重量又有差异时，就需要描述这批物件(总体)的重量分布以及概括其特征如平均重量等，这时才产生统计问题。另外如果对某一物件称重时有误差，每次观察的误差又是随机的，这时要了解这一物件的重量通常只能根据有限次的观察加以推断。这就产生了统计推断的问题。当然，要了解总体的特征和分布的规律离不开搜集个体的数据，但这仅是研究总体的一种手段，统计研究的最终目的是研究总体的数量特征及其规律性。

3.统计学是一门方法论的科学。在统计学界对统计学的性质有实质性科学和方法论科学之争。我们认为统计学是实用性很强的科学，就统计工作来说，它总是研究实际问题的，统计的方法也是从现实问题中产生的。然而统计学的发展有一个过程，早期的国势学派和政治算术学派虽然也利用一些统计方法来记述和分析现实问题，但这时还没有形成独立的统计学。随着统计方怯的应用日益广泛，其内容也不断发展和充实，尤其是概率论的发展为统计方法提供了理论基础，使统计的方法相对独立地形成了自己的科学体系，即统计学。其内容包括如何去搜集资料，如何对搜集的资料加以整理、概括和表征，以及如何对取得的数据进行分析和推断等一系列方法。这些方法和原理构成了统计学的基本内容。这将在本书的以后各章分别介绍。目前统计方法已成为科学研究和各种管理的重要工具，它是一门年轻而引人入胜的科学，并且还在不断地发展。

§1.2　统计学的分科以及和其他学科的关系

§1.2.1　统计学的分科

任何一门科学，随着人们对它的研究逐步深入，总是在不断地发展与进步。由于研究人员观察的角度不同，研究的重点不同，必然会出现各个相互联系而又有区别的分支，统计学也不例外，大致有以下两种分类：

1.描述统计学和推论统计学。统计学分为描述统计学和推论统计学，一方面是反映统计发展的两个阶段，同时也各有不同的侧重。描述统计学是研究如何对客观现象进行数学的计量、概括和表示的方法。有些客观现象的数字计量是比较简单的，如人的身高、体重等；而在某些领域如社会经济方面就比较复杂，比如要计量居民生活费的变动或不同国家人民生活水平的比较，就要涉及到很多方面。因此需要确定一些反映现象数量特征的范畴，即统计指标。研究某一个问题，要选择恰当的统计指标，然而一个指标往往只能说明某一方面的问题，要想用一个指标来全面评价社会经济现象的发展状况常常是困难的。为了比较全面系统地认识社会经济现象，就需要用若干个相互联系的指标来反映所研究问题的各个侧面，这些相互联系的统计指标就构成了统计指标体系。有了统计指标体系后就必须搜集数据，搜集数据的方法也要根据不同的研究对象和研究目的来确定。在一些自然科学中通常是根据实验观察来获得数据，而在社会科学中往往通过调查和访问来取得。对于搜集来的数据，开始总是许许多多杂乱无章的原始资料，难以直接看出什么问题，于是需要对数据简缩、加工，按各种项目分门别类加以整理，综合成一些图、表醒目地表达出来。这一系列的内容都是描述统计所要研究的。

推论统计学也称归纳统计学。在 20 世纪之前统计学基本上处于描述阶段，戈赛特、费希尔、奈曼等人对统计学的发展，使推论统计学的理论不断丰富并成为统计学研究的主流。它是研究如何根据部分数据去推论总体的情况，因为对客观现象搜集数据时，有时不可能对全部单位去作调查，即使有时可以调查但工作量也太大。例如要研究我国人口的年龄结构、婚姻状况、文化水平以及死亡率等情况，我国是 11 亿人口的大国，这些情况都要用普查的方法工作量极大，如果没有足够数量且训练有素的工作人员，难免出现误差，其结果就不见得理想，故通常采取抽取部分样本进行研究，从而对总体作出推断。当然，样本只是总体的一部分，包含的信息不完备，而且抽样时是随机抽取的，必然会出现误差，对于推断的结论是否可靠要冒一定风险。但是推论统计根据概率论的原理可以使归纳推断所产生的不确定性得到度量。为了使样本对总体作出更可靠的推断，推论统计也要研究如何对抽样进行设计等内容。

从以上介绍可以看出，描述统计学是统计学的基础，而推论统计学则是近代统计学的核心。本书基本上照顾到这两方面的内容，第 2 章和第 12 章属于描述统计内容，其余各章都是有关推论统计的内容。

2.理论统计学和应用统计学。理论统计学是指统计学的数学原理，它扎根于纯数学的一个领域——概率论。从广义来说统计理论可以认为是包括概率论的，此外还包括一些并不属于传统概率论的内容，如随机化原则的理论，各种估计的原理，假设检验的原理以及一般决策的原理，这些原理可以看成是概率论公理的扩增。在统计实践中常常会遇到一些新问题，使原有的统计方法不适应，需要统计学家针对新问题去建立一个与实际情况相适合的统计

模型，创造新的统计方法去分析，这就要靠统计理论来指与理论统计学相对应的是应用统计学。前面已经提到统计学是应用性很强的一门学科，因此将统计学的基本原理应用于各个领域就形成各种各样的应用统计学。它包括一整套统计分析方法，有的是一般性的统计方法，它适用于各个领域，如参数估计、假设检验、方差分析、相关与回归等。有的则是某一专业领域中特有的分析方法，例如经济统计学中的指数分析法等。近几十年来，由于统计研究的范围越来越广，一些科学实验也日趋复杂，统计方法也相应地复杂化和专门化，在应用统计方法中必须对因模型和实际情况的不一致而引起的各种误差的性质和大小作出判断，或提出改进的措施。由于统计的工具更加专门化了，它们就缺少通用性，一个统计学家要熟悉所有的专门工具已不可能。适应这种发展的需要，要求既要熟悉统计知识又要熟悉某一领域业务的应用统计学家就应运而生，同时也产生了相应的应用统计学。这类统计学的特点不着重于统计数学原理的推导，而是侧重于阐明统计的思想，并将理论统计学的结论作为工具应用于各个具体领域。本书主要是属于应用统计学，特别是侧重于经济管理方面的应用。

§1.2.2　统计学与其他学科的关系

1.统计学和数学的关系。统计学是研究客观现象数量方面的，它是应用数学的一个分支，因此与数学的关系十分密切，且与其他的应用数学有一定的共性。如和数学中的有关定理一样，统计中的一些分布也是客观现象数量方面的一种抽象。统计学中也要使用很多数学方法，学好统计学，尤其是理论统计学需要有坚实的数学基础。但另一方面，统计学与其他的数学分支相比又有其特殊性。(1)统计方法处理的数据必须是受到偶然性的影响而产生差异的数据。偶然现象在统计学中常称为随机现象，因此统计学及其理论基础概卒论不同于其他数学分支的一个特点在于它是研究随机现象的一门学科。(2)在方法上数学常常是用演绎的方法，即在作结论时，是从一些假设命题、已知的事实出发，按一定的逻辑推理去得到有关的结论。而统计学在本质上是用归纳的方法，它是根据观察到的大量个别情况，“归纳”起来去推断总体的情况，这一点与概率论的方法也有区别。因此目前国际上也有一种趋势，把统计学看成为与数学独立的学科。

2.统计学与其他专门学科的关系。统计方法有根广泛的实用性，其一般的数据分析方法适用于其他任何科学中的偶然现象，因此它与很多专门学科都有关系。但是统计方法只是从事物的外在数量表现去推断该事物可能的规律性，它本身不能说明何以会有这个规律性，这是各专门学科的任务。例如，用统计方法分析一些资料得出，吸烟与某些消化道疾病有关，这是从吸烟和不吸烟者的发病串的对比得出的结论，它不能解释吸烟何以会增加患这类疾病的危险性，这是医学这一专门学科的任务。所以统计方法只是一种工具，应用它进行定量分析时必须和定性分析结合起来。尤其是将统计方法应用于社会经济领域更应如此，因为社会经济现象比自然现象更为错综复杂，而又不可能象自然科学那样在实验室内排除其他因素进行试验。统计分析方法虽然是一个强有力的工具，但必须慎重使用，任意解释数据而不了解其背景是危险的做法。例如在苏联沙俄时代曾有统计资料表明，医生接生的婴儿死亡率高于接生婆接生的死亡率。能否由此说明医生的医务水平以及卫生条件不如接生婆呢？实际情况是当时医生还不普及，人们只有遇到难产时才去找医生接生，因此二者有不可比的因素。同时也需要指出，定性和定量分析相结

合也不等于先有了概念再去搜集数据，因为人们总是能精心选择数据来支持某项假设。虽然统计方法中也常用假设检验的方法，但必须遵循科学的方法，不能是主观有意识地挑选数据。而且统计的结论往往不能证明一项假设，只是表明这项假设与取得的数据结论并不矛盾。统计分析中还应注意，若因两个现象恰好一起变动就作出它们之间有关系的结论，这样做也常常是有危险的。例如近年来我国电视机的拥有量和刑事犯罪案件都在增加，能否由此说明电视机的普及鼓励了犯罪呢？显然，这样的结论是危险的，刑事犯罪案件的增加是受多种因素影响的结果，因此有必要在应用统计方法时全面周密地考虑全部有关材料，要把统计学知识和其他有关专业知识紧密结合起来，才能更好地发挥统计这一工具的作用。

§1.3　计量水准

在对客观现象进行计量时，对于客观现象的特征凡取值在一个以上的广义上可统称为变量，但这些变量的取值有的可以用数字来表示，如反映人口特征的年龄，反映物体特征的长度、重量，反映家庭特征的收入水平等。有的变量则不能用数字来计量，如反映人口特征的性别、宗教信仰、文化程度，反映物体特征的颜色、规格型号等，这些变量也称为属性变量。在利用统计方法处理时对数字变量可以进行较精密的处理，如计算平均数、方差等，而属性变量则不能，因此我们把能进行较精密处理的计量称作有较高的计量水准。一般说来较高的计量水准可以变为较低的计量水准，如把具体的年龄可以变为老年、中年、青年、儿童，把具体的成绩变为优、良中、及格、不及格等，而较低的计量水准不能变为较高的计量水准。计量的水准有四种：列名的，顺序的，间隔尺度的以及比率的，现分别予以说明。

§1.3.1　列名水准

这种计量水准在统计中是一种最原始、最低的或最有限制性的一种计量。例如，前面列举的人口按性别、民族、宗教信仰分类就是一种列名的计量。列名的分类没有严格的顺序，也就是说分类时在顺序上是可以改变的，如统计男女性别，将哪一类放在前面是一样的，但在各类之间是相互排斥的，这就是说一个人只能归属于其中的一类。这里需要注意的是在现代统计中为了便于应用计算机进行整理汇总经常把这些类别用数字代号来表示，如用“1”代表汉族，“2”代表满族，“3”代表蒙古族等等，但这些数字仅仅是一个符号，这些数字不能进行代数运算，例如 1＋2 不等于 3。

§1.3.2　顺序水准

顺序的水准较列名的水准要高一级，它的排列是有一定顺序的，如学生考试的成绩划分为优、良、中、及格、不及格就是一种顺序的计量。即优比良高一级，良比中又高一级，如此等等。然而它们之间的差别不能运算。例如获奖分为一等奖、二等奖和三等奖，虽然一等奖明显地高于二等奖，二等奖又明显地高于三等奖，然而一等奖、二等奖与三等奖之间的距离不能确切计量。因此顺序计量水准与上面列名水准的区别只是各类之间有“大于”(或“优于”)的概念，此外二者是一样的。

§1.3.3　间隔水准和比率水准

这两种计量水准是可以用数字计量的最高的计量水准，即这些计量水准之间不仅有顺序的关系而且可以计算间隔，例如老张的工龄是 20 年，大李的工龄是 10 年，小王的工龄是 5 年，可以计算老张的工龄比大李长 10 年，而大李的工龄比小王长 5 年。这些数字不仅可以计算间隔，而且可以进行加减乘除的数学运算，仍然是有意义的，例如可以说目前老张的工龄是小王工龄的 4 倍。

间隔水准和比率水准之间没有明显的差别，唯一的差别是间隔计量水准是使用任意一个零作为起点，而比率水准则使用一个实际的零作为起点。因此只有比率水准时，两个变量的比率才是有效的。一个典型的间隔计量水准的例子是以摄氏来计量温度，当 0℃时并不是完全没有温度，它只是海平面上水结冰的温度，因此这里只是以 0℃作为一个起点，根据物理上的绝对零度约是-273℃。如果说 40℃比 20℃高 20℃是可以的，如果说 40℃与 20℃之比相当于 20℃与 10℃之比就不恰当了。因为 40℃实际上是代表 273＋40＝313

°C，20°C是代表 273＋20=293°C，10°C是代表 283°C，所以 313°C：293°C与293°C1283°C之间是不相等的。形象地说，间隔计量水准好像是从桌面开始测量高度，比率计量水准是从地面开始测量高度。除此以外二者没有什么差别，可以合为一种计量水准来处理以后的变量。

第 2 章　统计数据的描述

§2.1　统计数据的搜集

数据资料是经济管理和工商企业管理决策的基础。为宏观和微观经济管理所需的统计数据一方面来自国家和地方各级统计及管理部门已经公布的数据资料，另一方面来自为该管理问题的研究所专门进行的调查。

§2.1.1　来源子出版物的数据资料

为保证国家宏观经济管理与决策的科学性，根据我国统计法规定，国家建立集中统一的统计系统，实行统一领导分级负责的统计管理体制。国务院设国家统计局负责组织领导和协调全国统计工作，并由国家统计局和省、自治区、直辖市的人民政府的统计机构，依照国家规定定期公布国民经济及社会发展的各种统计资料，供国民经济各管理部门和全社会使用。这些由国家统计局和各部委提供的公开的及未公开的出版物构成了各级政府、企业管理部门管理决策的重要资料来源。其中重要的统计出版物见表 2.1。

表 2.1 中所列的统计出版物中，有的提供我国国民经济宏观的各部门数据资料(如中国统计年鉴、中国信息报)，有的提供国民经济某方面的数据资料(如中国物价统计年鉴、中国劳动工资统计资料等)，有的提供世界各国国民经济数据资料(如国外经济统计资料、世界经济年鉴等)。与此同时，我国1982 年进行的全国人口普查,1985 年进行的全国工业普查等专项调查也都出版了资料专辑，服务于全社会。除此之外，各省市都出版有各省市的统计年鉴，各部委都有内部的统计资料，这些都是我们进行管理决策的资料来源。我们要善于充分地利用各种出版物所提供的经济统计信息，在此基础上做出科学的管理决策。

表 2.1　　我国几种重要的统计出版物

出　版　物	出 版 单 位
中国统计年鉴	中国统计出版社
中国信息报	中国信息报社
中国城市统计年鉴	新世界出版社
中国物价统计年鉴	中国统计出版社
中国工业经济统计年鉴	中国统计出版社
中国社会统计资料	中国统计出版社
中国农村统计年鉴	中国统计出版社
中国劳动工资统计资料	中国统计出版社
中国固定资产投资统计资料	中国统计出版社
钱国城镇居民家庭收支调查资料	中国统计出版社
国民收入统计资料汇编	中国统计出版社
世界工业统计汇编	中国统计出版社
国外经济统计资料	中国财政经济出版社
世界经济年鉴	中国社会科学出版社
海关统计	中华人民共和国海关总署

§2.1.2　来源于统计调查的数据资料

统计调查的目的是适应管理的需要去搜集公开或内部出版物中没有的统计资料。统计调查在组织形式上主要有普查、报表制度和抽样调查。普查这种形式是属于对所研究总体的全面调查，抽样调查按照随机原则只从研究总体中抽取一部分，因而属于非全面调查。

1.普查。普查是一种专门组织的、一次性的全面调查。它主要用来搜集某些不能够或不适于用定期的全面统计报表搜集的统计资料，以搞清重要的国情、国力和一定范围上的社会经济现象的总量。从宏观上讲，普查是了解国情、国力的重要调查方法，如我国于 1977 年进行了全国职工人数普查，1978 年进行了全国科学技术人员和基本建设在建项目普查，1982 年进行了第三次全国人口普查，1985 年进行了全国工业普查，1990 年又进行了第四次全国人口普查。

2.统计报表。统计报表是我国搜集统计资料的一种主要方式方法，同时也是我国政府统计工作的一项基本内容。统计报表的主要内容是按照我国宏观经济管理和国际对比的需要，并按照国家有关法规的规定，自上而下地统一布置，然后自下而上地逐级提供基本统计资料的一种统计调查。统计报表要以一定的原始记录为基础，按照统一的表式、统一的指标、统一的报送时间和报送程序进行填报。统计报表，按调查范围不同，分为全面的和非全面的。全面的统计报表要求调查对象的每一个单位都填报，非全面的统计报表只要求调查对象的一部分单位填报；统计报表，按报送周期长短不同，分为日报、旬报、月报、季报和年报等。报送的周期越短，人、财、物力花费的就越多。因而，周期短的日报和旬报只限于填报最主要的指标。凡是年报、季报能满足管理要求的，就不要用日报、旬报，统计报表，按报表内容和服务的范围，又分为国家的、部门的和地方的三种。国家统计报表从整个国民经济的角度出发，按国民经济部门划分，计有农业、工业、基建、物资、国内商业、对外贸易、劳动工资、交通运输、人民生活等内容，为整个国民经济管理服务。部门的和地方的统计报表主要为国民经济各部门和各地方管理服务，按各部门和各地方要求填报。应该统一并尽量减少各类报表。

3.抽样调查。抽样调查是科学研究和现代管理决策的一种重要调查方法，它通过从所研究总体中抽出的样本对总体数量特征作出推断，是一种既节省人、财、物力又能保证一定可靠性的科学方法。国家统计系统没有城调队和农调队，就是利用抽样调查方法搜集城市居民收支和消费、农业产量和农民生活的有关数据。抽样调查方法在企业产品质量控制和检验等方面均有广泛的应用。该方法的介绍请阅第 9 章抽样调查。

§2.1.3　统计调查方案

统计调查方案是指导整个调查过程的计划文件，内容应包括：

1.调查目的。调查目的要明确、具体，使参加调查的所有人员都清楚所进行调查的意义和要求。例如第三次全国人口普查的目的是“查清我国人口数字，查清我国人口的地区分布和社会经济结构情况，为有计划地进行社会主义现代化建设，统筹安排人民的物质和文化生活，制定人口政策和规划提供可靠的资料”。

2.调查对象和调查单位。调查对象(即调查范围)是根据调查目的确定的研究总体。例如第三次人口普查的对象确定为“具有中华人民共和国国籍并在中华人民共和国境内居住的人”。这样就把调查对象的界限划清了，凡是

在外华侨或虽住国内但无中国国籍的外国人都不属调查对象。

调查单位是构成调查对象的每一个单位，是调查内容的承担者。例如人口普查的调查单位是每一个人。

3.调查项目和调查表。调查项目即调查的内容，基本上是说明调查单位的特征的，在统计上称为标志(即总体单位的特征或属性)。调查单位的标志有数量标志和品质标志两类。数量标志是反映调查单位数量特征的，加人口普查中每一个人的年龄。品质标志是反映调查单位质量特征的，不能直接用数字表示，如人口普查中每个人的姓名、性别、民族等。

调查项目通常以表的形式来表示，称为调查表。它是登记调查单位资料(即数量与品质标志)的统一表格。如第三次全国人口普查的调查表见表2.2。为了帮助填表者正确地填写调查表，必须编制填表说明，以解释调查项目的内容和说明有关数字的计算方法。

表 2.2　　第三次全国人口普查登记表

本户　省、市　县　公社　生产大队　生产队　集体户
地址____自治区____市____街道____居委会____居民小组____名称____

每个人都填报						六岁及六岁以上的人填报	十五岁至十五岁以上的人填报				十五岁至六十四岁的妇女填报	十五至四十九岁的育龄妇女填报
一	二	三	四	五	六	七	八	九	十	十一	十二	十三
姓名	与户主的关系	性别	年龄	民族	常住人口的户口登记状况	文化程度	行业	职业	不在业人口状况	婚姻状况	生育子女总数和存活子女总数	1981年生育状况

4.调查时间。对调查项目所反映的时间和调查工作进行的时间都要有明确的规定。前者如第三次全国人口普查规定标准时点为1982年6月30日24时,后者如第三次全国人口普查登记时间规定在7月1日至7月10日之间完成人口普查登记工作。

除以上四个方面以外在调查方案中还应明确调查的具体方式方法(如直接观察法、采访法、报告法等)；规定调查工作的组织领导(如人员培训、调查经费和文件等)；布置有关调查的准备工作(如宣传教育、干部培训、文件印刷、调查经费的预算和开支办法等)。

§2.2　统计数据的整理

§2.2.1　统计数据的分组与次数分配

统计数据的整理是按照管理研究的目的，将搜集到的原始数据资料进行加工，从中提取有用的信息并探索其中的数量规律性。统计数据的分组是数据整理的第一步，即按不同的分组标志将数据划分为几个部分。在第 1 章中我们介绍了四种计量水准，其中列名水准和顺序水准是按照性质或属性标志对数据分类，例如人口数按性别、职业、地区、部门分类等等，这种分类称为按品质标志分组；间隔水准和比例水准是按照数字或数量标志对数据分类，例如工资、粮食产量、销售额的多少等等，这种分类称为按数量标志分组。

在我们国内的统计学教材中，通常不是按上述四种计量水准进行数据的度量和分组，而是将数据分成四种数列，即：空间数列、品质数列、时间数列和变量数列。空间数列是按不同地区标志进行的分组，例如人口按各省、布、自治区分组所形成的数列即空间数列。品质数列是按现象的性质、类别标志进行的分组，例如人口数按性别或民族分组所形成的数列即品质数列。时间数列是按照时间发生的先后顺序所形成的数据数列，例如我国解放后各年的人口数字就是时间数列。变量数列是按某一数量标志大小顺序进行的分组，例如某企业接工资收入多少进行的分组就构成变量数列。在以上介绍的四种数列中，空间数列和品质数列是按品质标志进行的分组，其计量水准属于列名水准和顺序水准。变量数列是按数量标志进行的分组，其计量水准属于间隔水准和比例水准。时间数列是一种较特殊的分组，共计量水准应属于间隔水准或比例水准。

统计整理为了达到探索客观事物内部数量规律性的目的，要提取数据中的有用信息，就需要根据研究目的进行以上所介绍的某种分组。数据观察值在各组中的个数称为次数，各组间的次数称为次数分配，是进行统计分析的重要工具。假定某研究所研究人员的月工资收入如表 2.3。

表 2.3　　某研究所研究人员月工资收入　　(单位：元)

人员编号	月收入	人员编号	月收入	人员编号	月收入
1	106	11	99	21	85
2	84	12	94	22	106
3	110	13	119	23	101
4	91	14	88	24	105
5	109	15	118	25	96
6	91	16	97	26	105
7	111	17	103	27	107
8	107	18	106	28	128
9	121	19	95	29	111
10	105	20	106	30	101

表 2.3 中最大值为 128 元，最小值为 84 元。可选择 10 元作为组距，则该所 30 名研究人员的工资收入可用划记法进行分组，见表 2.4。

表 2.4　　　　　　　　　划记法表

表 2.4 中第 1 种划记法是我国常用的画。“正”字的方法，每 个“正”字表示 5 个单位数值；第 2 和第 3 种划记法为国外常用的方法，第 2 种中画“■”表示 5 个单位，第 3 种先按单位个数画 4 个点，然后将 4 个点相连又有 6 条直线，画一个“■”表示 10 个单位。将分组后各组的次数与分组数值同时列出，就是次数分配表，见表 2.5。

表 2.5　　某研究所研究人员月工资收入次数分配表　　（单位：元）

月工资收入分组	次　数
80 — 90	3
90 — 100	7
100 — 110	13
110 — 120	5
120 — 130	2
合计	30

我们编制次数分配的目的是从数据中提取对我们探索内在数量规律性有用的信息。经整理后，我们大致可以看出该所 30 名研究人员月工资收入的分配规律，即工资收入最高者不超过 130 元，最低者不少于 80 元。在 30 人中，多数人的工资在 100—110 元间，形成两头小中间大的规律。但 80—100 元之间低收入的两组人数比 110—130 元之间高收入的两组人数多,因而是不对称的分配。在对该次数分配进行进一步的计算后，我们还可以得到工资分配的均值、中位数和方差等数量特征值，为反映该研究所人员工资的内在规律性提供概括性的数据。

以上我们编制该研究所月工资收入表是按 10 元一组划分的,这称为等组距的次数分配。根据所研究问题的需要，也可以按不等组距进行分组，如人口按年龄分组，既可以按 0—4、5—9、10—14……这样每 5 岁一组的等组距进行分组，也可以按不同年龄类型进行不等组距分组，如表 2.6。

表 2.6　　　　人口按不同年龄类型分组

	男　性	女　性
婴幼儿	0 — 6 岁	0 — 6 岁
少　年	7 — 17 岁	7 — 17 岁
中青年	18 — 59 岁	18 — 54 岁
老　年	60 岁以上	55 岁以上

在编制次数分配表时，为表明在某一数值之上或之下次数有多少，可以在次数分配的基础上，编制累积次数分配表和累积相对次数分配表。

次数分配也可以用各种图形表示，使得次数分配的数量规律性更直观、更形象。常用的图形有直方图、次数多边形图、次数曲线图和累积次数图。

直方图就是在平面直角坐标系中，用横轴表示分组的标志，用纵轴或直方条形的高度表示各组的次数。用表 2.5 的资料所绘制的直方图见图 2.1。

图 2.1　某研究所研究人员月工资收入分配的直方图

直方图更形象地显示出该研究所人员月工资收入的分配特征，这些数量特征如分配的对称性与偏斜性、均值的大小、分散的程度等就是数据数量规律性的表现。如果我们用一些直线线段连接各组条形顶端中值，就绘出了次数多边形图。如果观察值次数很多，组距很小，所绘成的次数多边形图就会形成一条平滑的曲线，即次数曲线。不同的客观事物有不同的数量规律性，因而也就形成了形状不同的次数曲线。在日常生活和经济管理中，较常见的有四种次数曲线，即正态分布曲线、偏态曲线、J 形曲线和 U 形曲线(见图 2.2)。正态分布曲线也称为钟形曲线(见图 2.2 中的(a)图)，形为左右对称的倒挂的大钟，这是客观事物数量特征表现最多的一种次数曲线，如人的身高、体重、智商；电子管中的热噪声、电流、电压，还有纤维长度、细纱强度；钢的含碳量；粮食作物产量；橡胶的抗张力等等。其所有的试验、测量和观测误差等都服从正态分布。

图 2.2　几种常见的次数曲线

偏态曲线根据长尾拖向哪一方又可分为正偏(或右偏)和负偏(或左偏)两种曲线。例如人均收入分配的曲线就是正偏曲线，即低收入的人数较多，而高收入的人数较少，二者的收入水平差距较大。

J 形曲线中又有正 J 形和倒 J 形曲线两种，应用较广的分别是西方经济学中的供给曲线和需求曲线，供给曲线(正 J 形曲线)表现为随着价格(横轴)的增加，供给量(纵轴)以更快的速度增加；需求曲线(倒 J 形曲线)表现为随着价格(横轴)的增加，需求量(纵轴)以较快的速度减少。供给和需求曲线的交叉点即供求平衡点。

U 形曲线又称为生命曲线或浴盆曲线，人和动物的死亡率往往服从 U 形曲线分布。婴儿和动物的幼仔，由于抵抗力弱，死亡率很高，随着对新环境的适应和年龄的增长死亡率逐渐降低。到了中年时期死亡率最低同时也相对稳定，进入老年期后又逐渐增高，即形成一个浴盆形状的分布曲线。产品的故障和报损情况也有类似的分布规律。

§2.2.2　洛伦茨(Lorenz)曲线

洛伦茨曲线是美国经济学家洛伦茨在本世纪初提出的、应用累积次数分配曲线描述一个国家或一个地区收入分配平均程度的一种图示方法。例如北京市房山县某村 1981 年和 1988 年人口及收入资料如表 2.7。

将表 2.7 的相对累积人数作为横轴，相对累积总收人作为纵轴，并将 1981 年和 1988 年各收入组的数据画在坐标系上，连接各点形成一条光滑曲线即为洛伦茨曲线(见图 2.3)。

图 2.3　1981 年和 1988 年北京市房山县某村收入分配的洛伦茨曲线

表 2.7　北京市房山县某村 1981 年和 1988 年人口及收入情况表

图 2.3 中虚线部分代表 1981 年的数据，实线部分代表 1988 年的数据。由于我们已将人数和收入转换成相对累积数字，因而坐标系的横轴和纵轴都是以 1.0 为单位的相同长度线段。从左下方零点到右上方顶点的对角线表示收入分配的绝对公平线，即某一比例的人数得到相同比例的收入。横轴底线

和纵轴右边线称为绝对不公平线，表明接近 100%的人没有收入，而极少数人获取了全部收入。绝对公平线和绝对不公平线是收入分配的两种极端情况，实际不可能存在。不同的经济制度、不同的地区、不同的发展阶段的收入分配均有差别，这种差别表现在右下半边三角形中曲线的特征上。我们看到,该村 1981 年的曲线在 1988 年曲线之上,表明 1988 年收入分配差距拉大。但要注意，这种差距加大并不一定是坏事情，因为收入差距太小则不能调动积极性，且由于各人能力不同，一定的收入差距也反映了多劳多得的按劳分配制度。故不同地区和不同时期的收入分配差别要具体情况具体分析。

§2.3　集中趋势的测度

一组数据经过分组整理后，形成了次数分配。将次数分配用直方图或次数分配曲线图画出来，我们对该组数据的变化规律就有了直观的了解。然而，进一步的推断、决策等研究不仅要求我们对其分布变化的规律有直观的了解，而且要求我们用几个最简洁又最能充分描述其分布数量特征的统计量将其分布变化的规律性表示出来。这些统计量包括：数据次数分配集中趋势的测度、离散程度的测度、偏态程度的测度及峰度的测度。由于管理所研究的大部分总体分布都是正态的或近似正态的，而对偏态和峰度很少研究，所以我们在本教材中也予以省略。这一节将介绍数据集中趋势测度的几种方法。

§2.3.1　均值

均值又称为算术平均数，用$\overline{X}$表示。均值的计算公式是

$$\overline{X}=\frac{X_1+X_3+\cdots+X_N}{N}=\frac{\sum_{i=1}^{N}X_i}{N} \qquad (2.1)$$

式中：N是总体单位数；Σ是求和符号；$\sum_{i=1}^{N}X_i$是对总体所有单位数值求和。

现假定表2.3的资料是某研究所全部30名研究人员的工资收入资料，则按（2.1）式计算

$$\overline{X}=\frac{106+84+110+\cdots+111+101}{30}$$

$$=\frac{3105}{30}103.5\,(\text{元})$$

结果 103.5 元就是全所 30 名研究人员月平均工资数。该平均工资可用来反映研究所研究人员的一般工资水平，是抽象掉具体人员工资之后的一个代表值。在均值的计算中，个别高工资和个别低工资的工资额互相抵消了，因而具有集中趋势的作用。

如果数据经过了整理，形成了次数分配，均值的计算公式为

$$\overline{X}=\frac{X_1f_1+X_2f_2+\cdots+X_kf_k}{f_1+f_2+\cdots+f_k}=\frac{\sum_{i=1}^{k}X_if_i}{\sum_{i=1}^{k}f_i} \qquad (2.2)$$

式中：X_i是第i组的组中值；f_i是第i组的次数。由表2.5的次数分配数据计算的均值为

$$\overline{X}=\frac{85\times3+95\times7+105\times13+115\times5+125\times2}{3+7+13+5+2}$$

$$=\frac{3110}{30}=103.67\,(\text{元})$$

我们注意到利用原始数据计算的均值为 103.5 元，而利用次数分配数据计算的均值为 103.67 元。前者是准确的均值，因为它的计算充分利用了原始资料的全部信息；后者则是一种近似的数值，原因是计算中利用组中值作为该组数值的代表，它是以大量数据并且数据在该组中均匀分布作为前提的。公式

(2.1)称为均值的简单式，(2.2)称为加权式。加权式还可以有另一种算法，即

$$\overline{X}=\frac{\sum_{i=1}^{k}X_if_i}{\sum_{i=1}^{k}f_i}=\frac{X_1f_1+X_2f_2+\cdots+X_kf_k}{\sum_{t=1}^{k}f_i}$$

$$=X_1\frac{f_1}{\sum_{i=1}^{k}f_i}+X_2\frac{f_2}{\sum_{i=1}^{k}f_i}+\cdots+X_k\frac{f_k}{\sum_{i=1}^{k}f_i}$$

$$=\sum_{i=1}^{k}X_i\frac{f_i}{\sum_{i=1}^{k}f_i} \qquad (2.3)$$

公式(2.3)的结果表明，均值的加权式还可以用求组中值X_i乘以其权系数$\frac{f_i}{\sum_{i=1}^{k}f_i}$的和的方式求出。该式清楚地说明均值不仅受到组中值$X_i$大小的影响，而且受到权系数$\frac{f_i}{\sum_{i=1}^{k}f_i}$大小的影响。

均值在测度集中趋势的统计量中应用最广，原因不仅表现为通过样本均值去估计和检验总体均值具有很好的性质(详见第 4 章)，而且均值本身也具有一些很好的数学性质，其中最重要的有两条，即数据观察值与均值的离差之和为零以及观察值与均值的离差平方和最小，亦即

$$\sum_{i=1}^{N}(X_i-\overline{X})=0\text{或}\sum_{i=1}^{k}(X_i-\overline{X})f_i=0 \qquad (2.4)$$

$$\sum_{i=1}^{N}(X_i-X_0)^2>\sum_{i=1}^{N}(X_i-\overline{X})^2 \qquad (2.5)$$

(2，5)式中的 X_0 是任一不等于 $\overline{X}$ 的数值，当 X_0 等于 $\overline{X}$ 时，(2.5)式两边相等。公式(2.4)表明均值是偶然性观察值抵消后的必然性数量特征，公式(2.5)是以后将要讲到的最小二乘法的基础。当然任何一种统计量都存在局限性，均值的局限性主要表现为易受极端值影响，某个极端大值或极端小值都会影响均值的代表性，同时还影响其对集中趋势侧度的准确性。

§2.3.2　中位数

将数据观察值 X_1，X_2，⋯，X_N 按其变量值由小到大的顺序排序后为 $X_{(1)}$，$X_{(2)}$，⋯$X_{(N)}$，其位置处于中间的变量数值即中位数，如果数据个数为奇数，则中位数数值恰为$\frac{N+1}{2}$位置上的数值。如果数据个数为偶数，则中位数数值为最中间位置上两个数值的均值。当数据经过整理，形成次数分配后，中位数的计算只能在假定各组观察值在各组中均匀分布的条件下近似地得出。用 M，表示中位数，其插值公式是

$$M_e = \frac{L + \frac{\Sigma f}{2} - S_{m-1}}{f_m} \cdot i \qquad (2.6)$$

式中：$\frac{\Sigma f}{2}$表示中位数位置，L 表示中位数所在组下限值，S_{m-1}表示中位数所在组以下各组的累积次数；f_m表示中位数所在组的次数；i 表示中位数所在组的组距。利用表 2.5 的资料，可由公式(2.6)算出

$$M_e = 100 + \frac{\frac{30}{2} - 10}{3} \times 10 = 103.85(\text{元})$$

用表 2.3 全部数据排序得到的中位数数值为 105 元,这里得出的 103.85 元只是近似值。

中位数比均值的计算简单，且不受极端值的影响，具有稳健性，是数据描述与集中趋势测度的另一个较好的统计量。

§2.3.3　众数

众数是被研究总体中出现次数最多的那个变量值，众数与中位数都不是通过全部观察值计算出的，而是将数据排序后从位置上直接得到的，因而这两种代表值也称为位置平均数。显然，只有当次数分配具有明显的集中趋势时，才有众数可言。如由表 2.3 数据直接观察，众数应为 106 元。因为得此工资额的共有 4 人，是人数最多的一种工资水平。如考虑表 2.5 次数分配数据，100—110 元组共 13 人，在该组中具有集中趋势，可用下式近似地计算众数数值，用 M_0表示众数

$$M_0 = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot i \qquad (2.7)$$

式中：Δ_1表示众数所在组与下一组次数之差；Δ_2表示众数所在组与上一组次数之差。将表 2.5 数据代入公式(2.7)，可得

$$M_0 = 100 + \frac{13-7}{(13-7)+(13-5)} \times 10 = 104.29(\text{元})$$

众数的近似公式(2.7)是假定众数所在组与相邻的上下两组次数之差反映了偏态分布陡峭上升而缓慢下降的特点而插值得出的。表 2.5 中实际资料 90—100 元的有 7 人，比 110—120 元的 5 人多，按此两组次数插值得 104.29 元，即从 105 元中值偏向 7 人这一边。实际个人工资额 106 元的人最多，其次数分配与标准偏态分配不符，因而误差较大。

众数不宜于进行总体参数的推断之用，但它有特殊应用的范围。如农贸市场上某种商品的价格水平是以众数作为代表值的；市场上各种尺码鞋子的需求量往往也是以众数为尺度的。

§2.3.4　几何均值

几何均值又称为几何平均数，主要应用于指数和平均发展速度的计算。用 M_g表示几何均值，其计算公式为

$$M_g = \sqrt[N]{X_1 \cdot X_2 \cdot \cdots X_N} \qquad (2.8)$$

即 N 个数值连乘积的 N 次方根。在计算水平法平均发展速度时，$x_1, x_2 \cdots X_N$不是实际观察值，而是实际观察值两两相除所得的比率。例如以 $a_0, a_1, a_2 \cdots$

a_N表示时间由基期至第N期的时间数列，$\frac{a_1}{a_0}$，$\frac{a_2}{a_1}$，$\frac{a_3}{a_2}$，$\frac{a_4}{a_3}$，…，$\frac{a_N}{a_{N-1}}$就是一列逐期的发展速度。由公式(2.8)可以计算出这N个发展速度的平均数

$$M_g = \sqrt[N]{\frac{a_1}{a_0}\cdot\frac{a_2}{a_1}\cdot\frac{a_3}{a_2}\cdots\frac{a_N}{a_{N-1}}} = \sqrt[N]{\frac{a_N}{a_0}}$$

由几何均值公式计算的结果就是 n 个发展速度的平均发展速度(详见第 7 章)。

§2.3.5　调和均值

调和均值是观察值倒数之平均数的倒数，故也称为倒数平均数。它广泛用于指数的计算中。用 M_H 表示调和均值，共计算公式为

$$M_H = \frac{1}{\frac{\sum_{i=1}^{N}\frac{1}{X_i}}{N}} = \frac{N}{\sum_{i=1}^{N}\frac{1}{X_i}} \tag{2.9}$$

加权的形式为

$$M_H = \frac{N}{\Sigma f_i \frac{1}{X_i}} \tag{2.10}$$

均值、几何均值及调和均值三者同属于均值体系。均值是直接对观察值进行平均；几何均值是对观察值取对数后的平均；调和均值则是对观察值取倒数后平均，然后再取倒数。一般情况下，三者的关系是

$$M_H \le M_g \le \overline{X}$$

§ 2.3.6　切尾均值

切尾均值是综合了均值和中位数两种计量优点的一种新的对集中趋势测度的计量。切尾均值现已广泛应用于电视大奖赛、体育比赛及需要由人们进行综合评价的竞赛项目，我们在电视中所熟悉的“去掉一个最高分，去掉一个最低分，最后得分是×分”就是利用切尾均值方法进行的评估，该方法的计算公式是

$$\overline{X}_a = \frac{X_{([n\bar{a}]+1)} + X_{([n\bar{a}]+2)} + \cdots + X_{(n-[n\bar{a}])}}{n-2\cdot[na]} \tag{2.11}$$

式中 :n 表示观察值的个数 ;a 表示由人们决定的大于等于 0 又小于$\frac{1}{2}$的系数，即 $0\le a<\frac{1}{2}$；[ ]表示取整数，$X_{(1)}$，$X_{(2)}$，…，$X_{(n)}$表示 $X_1,X_2,\cdots,X_n$ 的顺序统计量。切尾均值的计算特点是先将观察值两端的个别极大或极小值切去，然后再对中间的观察值进行平均。两端各切去几个数值，通过我们给定 a 值加以确定。当 a=0 时，切尾均值等于均值：当 a 接近$\frac{1}{2}$时，切尾均值接近于或等于中位数。例如某次比赛共有 11 名评委，对某位歌手的给分分别是 9.22,9.25,9.20,9.30,9.65,9.30,9.27,9.20,9.28,9.25,9.24 经整理，顺序统计量为

9.20,9.20,9.22,9.24,9.25,9.25,9.27,9.28,9.30,9.30,9.65 如去掉一个最高分，去掉一个最低分，则取 a＝$\frac{1}{11}$，切尾均值为

$$=\frac{X_{([11\cdot\frac{1}{11}]+1)}+X_{([11\cdot\frac{1}{11}]+2)}+\cdots+X_{(11-[11\cdot\frac{1}{11}])}}{11-2\cdot[11\cdot\frac{1}{11}]}$$

$$=\frac{X_{(2)}+X_{(3)}+\cdots+X_{(10)}}{11-2}$$

$$=\frac{9.20+9.22+\cdots+9.30}{9}$$

$$=9.26$$

这个平均得分避免了 9.65 分这个极端高分的影响。如取 a=0

$$\overline{X}_0=\frac{X_{(1)}+X_{(2)}+\cdots X_{(11)}}{11}=9.26$$

即$\overline{X}_0=\overline{X}$；如果$a=\frac{5}{11}$

$$\overline{X}_{\frac{5}{11}}=\frac{X_{(6)}}{1}=X_{(6)}=9.25$$

即 $X_{5>11}=M_e$。可见，切尾均值是综合了均值与中位数优点的一种较理想的统计量。

本节介绍了集中趋势测度的 6 种统计量或 6 种方法。均值是最基本的，同时也是应用最多的方法，其他 5 种测度方法都只是在一定条件下应用并用来补充均值这一主要测度方法的。

§2.4 离散程度的测度

前一节我们讨论了对一组数据集中趋势测度的方法。要把握一组数据的数量变化规律仅仅有集中趋势的测度值是不够的，与集中趋势相对应的，还要了解数据的离散程度。本节将介绍离散程度测度的主要方法：极差和方差。

§2.4.1 极差

极差也称为全距，是一组数据的最大值和最小值之差，即

$$R=X_{(n)}-X_{(1)} \tag{2.12}$$

显然，一组数据的差异越大，其极差也越大。以表 3.3 的数据为例，该研究所最高工资水平是 128 元，最低工资是 84 元，全所工资水平的极差是

$$R=128-84=44(\text{元})$$

用极差法测度离散程度十分简单，它广泛用于产品质量管理中控制质量的差异，但它有一定的局限性，因为它只计算一组数据中的两个数据，其余数据的信息没有利用，统计学中测度离散程度应用最多的是方差和标准差。

§2.4.2 方差和标准差

方差是观察值与其均值离差平方和的均值，它又有总体方差和样本方差之分。如果观察值数据是总体数据，则总体方差的计算公式为

$$\sigma^2=\frac{\sum_{i=1}^{N}(X_i-\overline{X})^2}{N} \tag{2.13}$$

式中：σ^2是总体方差的符号。标准差是方差的正平方根，即

$$\sigma=\sqrt{\frac{\sum_{i=1}^{N}(X_i-\overline{X})^2}{N}} \tag{2.14}$$

如果观察值是一组样本数据，其样本方差公式为

$$S^2=\frac{\sum_{i=1}^{n}(\overline{x}_i-\overline{x})^2}{n-1} \tag{2.15}$$

式中：S^2 表示样本方差；分母 n-1 表示自由度。之所以样本方差的自由度用 n-1 而不用 n，因为 n-1 自由度的样本方差 S^2 是总体方差σ^2 的无偏估计。样本标准差是样本方差的正平方根

$$S=\frac{\sqrt{\sum_{i=1}^{n}(\overline{x}_i-\overline{x})^2}}{n-1} \tag{2.16}$$

如果计算方差的资料是次数分配数据，在计算方差时要将各组权数 f_i 考虑进去，总体方差的公差为

$$\sigma^2=\frac{\sum_{i=1}^{k}(\overline{x}_i-\overline{x})^2 f_i}{\sum_{i=1}^{k}f_i} \tag{2.17}$$

样本方差的加权公式为

$$S^2=\frac{\sum_{i=1}^{k}(\overline{X}_i-\overline{X})^2 f_i}{\sum_{i=1}^{k} f_i-1} \tag{2.18}$$

方差的计算公式告诉我们，方差对离散程度的测度是以均值为中心的。它们互相依赖，互相补充，共同描述一组数据的数量变化特征，由于标准差具有同观察值相同的计最单位，在实际应用中多用标准差作为离散或差异程度的测度值。

公式(2.13)至(2.18)都是方差和标准差的定义公式，在实际计算时若计算每个观察值与均值的离差平方和则计算工作量很大，一般用方差的另一种表达式(2.19)计算。公式(2.19)式的推导过程如下

$$\begin{aligned}\sigma^2&=\sum_{i=1}^{N}(X_i-\overline{X})^2\\&=\frac{1}{N}(\sum_{i=1}^{N}X_i^2-2\overline{X}\sum_{i=1}^{N}X_i+N\overline{X}^2)\\&=\frac{1}{N}(\sum_{i=1}^{N}X_i^2-2\overline{X}\sum_{i=1}^{N}X_i+\overline{X}\sum_{i=1}^{N}X_i)\\&=\frac{1}{N}(\sum_{i=1}^{N}X_i^2-\overline{X}\sum_{i=1}^{N}X_i)\\&=\frac{\sum_{i=1}^{N}X_i^2}{N}-(\overline{X})^2 \qquad (2.19)\end{aligned}$$

对次数分配数据则有

$$\sigma^2=\frac{\sum_{i=1}^{k}X_i^2 f_i}{\sum_{i=1}^{k}f_i}-(\overline{X}) \tag{2.20}$$

为说明标准差的计算，我们仍用表 2.3 和表 2.5 数据为例。我们仍假定上述资料是总体数据，用公式(2.14)计算表 2.3 的数据，得到

$$\sigma^2=\sqrt{\frac{\sum_{i=1}^{N}(X_i-\overline{X})^2}{N}}=10.28(\text{元})$$

又用公式(2.20)计算表 2.5 数据，得到计算表 2.8。

表 2.8　　标准差计算表

月工资收入（元）	组中值 X（元）	人　数 f	Xf	X^2f
80 — 90	85	3	225	21675
90 — 100	95	7	665	63175
100 — 110	105	13	1365	143325
110 — 120	115	5	575	66125
120 — 130	125	2	250	31250
合　计	——	30	3110	325550

$$\sigma=\sqrt{\frac{\sum_{i=1}^{k}X_i^2f_i}{\sum_{i=1}^{k}f_i}-\left(\frac{\sum_{i=1}^{k}X_if_i}{\sum_{i=1}^{k}f_i}\right)^2}$$

$$=\sqrt{\frac{325550}{30}-\left(\frac{3110}{30}\right)^2}$$

$$=10.24(\text{元})$$

§2.5 茎叶法

对统计数据的整理，前面所介绍的传统方法是先将数据分组，然后形成次数分配,再进一步计算均值、中位数等值。茎叶法(Stem-and-Leaf Displays)可将统计分组和次数分配两项工作一次完成，图形直观且保留有原始信息，均值、中位数和众数均可依据原始信息准确方便地算出。在电子计算机日益普及的今天，这种方法更显示出其优越性。

我们仍以表 2.3 的数据为例以便于与传统分组和直方图法进行比较。茎叶法将分组之数量标志(即组距)视为树茎，将每一个观察值视为一个树叶，每个树叶按照树茎之要求长在应长的树茎上(见图 2.4)。我们将工资收入的十位和百位数作为树茎，如第一个树茎 8 表示十位数是 8，则个位数由 0—9 都应该长在该树茎上。在树茎确定并画好后，然后依次将每个研究人员工资的个位数写到对应的树茎上。如第 1 个研究人员收入 106 元，个位数的 6 应长在树茎 10 之上，第 2 个研究人员收入 84 元，个位数的 4 应长在树茎 8 之上。只要将各个叶子横竖对齐，所有的 30 个树叶长好后，就是一个横放的直方图，各树茎上的叶子数就是各组的次数，而且原始数据的信息全部保留在图中。均值可直接用 30 个原始数值算出，不必再用组中值做近似计算了。中位数只需在相应的树叶中寻找即可。本例中共 30 个观察值，中位数位置应是 $\frac{N+1}{2}=\frac{30+1}{2}=15.5$，就是顺序统计量的第 15 个和第 16 个数值的均值。不难得到第 15 个和第 16 个数值都是 105 元，中位数就是 105 元。众数在茎叶图中就是树叶出现最多的那个，本例中 106 元出现了 4 次，105 元出现了 3 次，107 元出现了 2 次，可以认为 106 元是数据中的众数。

图 2.4　某研究所研究人员月工资收入的茎叶图

上面介绍了简单数据的茎叶图画法，下面我们给出数据位数较多的人口普查资料(见表 2.9)

表2.9　　　　　　1982年全国人口普查数字

地区别	1982年普查人口数（人）	地区别	1982年普查人口数（人）
全国总计	1031882511	河南省	74422739
大陆29个省、市、自治区合计	1003937078	湖北省	47804150
湖南省	54008851	北京市	9230687
广东省	59299220	天津市	7764141
广西壮族自治区	36420960	河北省	53005875
四川省	99713310	山西省	25291389
贵州省	28552997	内蒙古自治区	19274279
云南省	32553817	辽宁省	35721693
西藏自治区	1892393	吉林省	22560053
陕西省	28904423	黑龙江省	32665546
甘肃省	19569261	上海市	11859748
青海省	3895706	江苏省	60521114
宁夏回族自治区	3895578	浙江省	38884603
新疆维吾尔自治区	13081681	安徽省49665724	
福建省	25873259	中国人民解放军现役军人	4238210
江西省	33184827	台湾省	18270749
山东省	74419054	香港、澳门地区同胞	5378627

我国 1982 年第三次全国人口普查统计资料中各省市人口数字多达 8 位数，树茎和树叶位数过多不仅不明了，而且树茎位数过多难以表示。我们将 8 位数的后 5 位数舍去，以 10 万人口为单位，以期反映各省市人口的差别，现将第 8 位数字(千万人)作为树茎，以第 7 和第 6 位两位数字作为树叶，于是将大陆 29 个省、市、自治区及解放军、台湾省、港澳地区共 32 个人口数据的茎叶图绘出(见图 2.5)。

图 2.5　我国各省、市、自治区的人口茎叶图

在舍去后五位数时，对第五位数采取的是“四舍六入五单双，奇进偶不进”的准则，即第五位数为 1—4 则舍，第五位数为 6—9 则八进，第五位数为 5 时，要看第六位数是奇数还是偶数，是奇数则进，是偶数则不进。例如上海市、贵州省和云南省人口数的第五位都是 5，则要看第六位数，上海市人口的第六位数是 8，属偶数则舍去，就在树茎“1”上画上 18 的树叶。贵州省和云南省的第六位都是 5，属奇数则进上去，分别在 2 和 3 的树茎上画上 86 和 26 的树叶。之所以采取这种舍入方法，是因为在大量数据处理时“四舍五入”准则平均偏大，这种方法的平均误差为 0。

在茎叶法中，技巧主要表现为设计适当的树茎。因为树茎定了，数据值就已经固定在某一组中了，只需将每一数据值的树叶部分填在对应树茎的树叶组中就行了。在人口年龄分组中，一般按 5 岁一组，即年龄由低到高依次分为：0—4 岁组，5—9 岁组，10—14 岁组，15—19 岁组……在这类不是每 10 个数值为一组的分组中，茎叶法常在树茎右边用一些符号表示。例如“0*”表示 0—4 岁，“0*”表示 5—9 岁，“1*”表示 10—14 岁，“1*”表示 15—19 岁，“2*”表示 20—24 岁等等，即树茎的数字仍表示十位数的数字，

“ * ”和“ * ”分别表示个位数的数字为 0—4 和 5—9。例如某居民组的人口年龄数据如下：6，13，0，39，25，10，19，24，……，若将前 8 个数据按 5 岁一组的方法分组，可得到如图 2.6 的茎叶图。

图 2.6　某居民组的年龄按每 5 个数值一组的茎时图

如果在以上的年龄分组中还希望分得更细些，每 2 个数值分为一组，即每 10 个个位数都分成五组。那么，与上面介绍的符号标志相似，我们用“ * ”表示个位数为 0—1，用“ t ”表示个位数为 2—3，用“ f ”表示个位数为 4—5，用“ s ”表示个位数为 6—7，用“ * ”表示个位数为 8—9。例如某托儿所和幼儿园孩子的年龄都在 7 岁以下，其中 0—1 岁上婴儿班，2—3 岁上小班，4—5 岁上中班，6—7 岁上大班。现有如下孩子的年龄数据

1，4，3，6，5，2，2，0，1，7，0，5……用茎叶法表示出来也就是对孩子进行分班，我们就可得到如图 2.7 的茎叶图。

如果数据有正又有负，则可以在树茎前加上正负号以示区别。例如有如下数据

5，-3，-9，0，12，-13……

做出茎叶图如图 2.8。

图 2.7　婴、幼儿年龄按每 2 个数值一组的茎叶图

图 2.8　带有正负值的茎叶图

我们在这里只给出了茎叶法的思想和一些简单作法，使用者可以自己设计树茎，充分利用这一方法解决实际问题。

习　题

1.某菜店元月份逐日销售额(元)资料如下：

257	276	297	252	228	310	240	228
265	278	271	292	261	281	301	274
267	280	291	258	272	284	268	303
273	282	263	322	249	269	290	

要求：

(1)计算该菜店元月份每日平均销售收入和标准差以及销售收入的中位数；

(2)试以 10 元为组距对该资料进行分组，分为 10 组，并画出直方图；

(3)利用分组数据近似计算该数据的均值、中位数和众数；

(4)若去掉两个最高收入额、去掉两个最低收入额，用切尾均值公式计算其均值；

(5)以 10 元为组距分为 11 组，将该数据用茎叶图表示出来；

(6)计算上述资料的两个四分位数。

2.若有两个正数，计算得到均值为 7.5，几何平均数为 7.1，调和平均数为 7.0，上述结果是否正确？

3.某班学生共 50 人，分为甲、乙两组。甲组学生 20 人，统计学平均成绩为 78 分，标准差为 8 分：乙组学生 30 人，统计学平均成绩为 72 分，标准差为 10 分。求全班 50 名学生的平均成绩及标准差。

第 3 章　概率与概率分布

在前面两章介绍了统计学的一些基本概念及其对数据的描述。如果获得的数据是所研究问题总体的全部数据，通过对数据的描述就可以直接得到表示总体数量规律性的参数及其分布。然而在实际研究中，由于种种原因，往往无法得到全部总体数据，只能是抽取部分总体单位作为样本，由样本所提供的信息对总体数量规律性做出推断，其理论基础是概率论与数理统计的估计理论。在这一章只简要介绍概率论的基本内容，以便为后面几章统计推断的内容打下基础。

§3.1 随机事件与概率

§3.1 样本空间与事件

对社会现象的观察和对自然现象的科学实验统称为试验，如果在相同的条件下试验可以重复进行，而且每次试验的结果不能事先确定，则称这样的试验为随机试验。例如掷一颗骰子，观察出现的点数，在一批产品中任意抽取一件，检验它的质量水平；记录某地一昼夜的最高温度和最低温度等都可以看作是随机试验。统计学就是通过随机试验来研究随机现象的数量规律性，下面我们以掷骰子的试验为例介绍一下样本空间和事件两个基本概念。掷一次骰子共有 6 种可能的点数结果，表示为

$$S=\{1,2,3,4,5,6\}$$

式中：S 表示样本空间，{ } 符号表示集合；集合中每个数字表示一个元素，则每个元素正好对应着一种可能的结果，由代表一试验全部可能结果的各元素组成的集合称为样本空间。样本空间的元素称为样本点。在试验中每一种可能结果都是一个随机事件，并且是该试验的最简单的随机事件，也称为基本事件。然而，在试验中人们通常只关心全部可能的结果中满足某些规定性质的结果。例如掷骰子出现点数是偶数的各种结果，我们用大写字母 A 表示，即 A＝ {2，4，6} 。又如全部产品中质量合格的产品 B＝ {全部合格品} ，如果说样本空间对应着集合论的全集的话，那么满足某些规定性质的结果的集合就是全集中的一个子集。我们一般称在给定样本空间 S 中，满足规定性质的结果的子集为事件。事件用大写字母 A、B、C…表示。若一个事件在每次试验中都必定发生，则称为必然事件。显然，只有样本空间 S(即全集)才是必然事件。与必然事件相对应的是不可能事件，也就是空集，用ϕ表示。

§3.1.2 事件的概率

概率是针对事件而言的。事件 A 的概率即事件 A 在样本空间 S 条件下发生可能性的大小，记为 P(A)。显然，必然事件发生的可能性是百分之百，所以它的概率是 1，而不可能事件发生的可能性是零，所以它的概率是 0，即

$$P(S)=1,\ P(\phi)=0$$

样本空间 S 有时也记为Ω，有 $P(\Omega)=1$。

概率的研究来自于欧洲中世纪的赌博或机会游戏。在数百年的发展中，许多数学家、统计学家对概率及其计算做出了巨大的贡献，从三个不同的角度对概率下了定义，做出了三种不尽相同的解释，即概率的古典定义、统计定义和主观定义。

1.概率的古典定义。人们最早研究概率是从掷硬币、掷骰子和摸球等游戏和赌博中开始的。这类游戏有两个共同的特点：第一，试验的样本空间元素有限。如掷硬币有正反两种结果，掷骰子有 6 种结果等；第二，试验中每个结果出现的可能性相同。如掷硬币出现正反的可能性各为$\frac{1}{2}$，掷骰子出现各种点数的可能均为$\frac{1}{6}$。

具有这种特点的随机试验称为古典概型或等可能概型。计算古典概型概率的方法称为概率的古典定义或古典概率。又由于样本空间有限，总能够用逻辑推理方法在试验之前推出各种事件的概率，因此古典概率也称为验前概率或逻辑概率。

定义 3.1 在古典概型中，事件 A 所包含的基本事件个数 m 与样本空间中基本事件总数 n 的比值称为事件 A 的概率，记作

$$P(A)=\frac{\text{事件A中包含的基本事件数}}{\text{样本空间中基本事件总数}}=\frac{m}{n} \tag{3.1}$$

例 3.1 一箱产品共 100 件，已知其中有 5 件次品。从中任取一件为次品的概率是多少？

解：令 A 事件为“任取一件是次品”。样本空间由 100 件产品组成，编号为 1，2，…，100。若 5 件次品的编号为 1，2，3，4，5。则事件 A 和样本空间的集合分别是

A= {1，2，3，4，5}

S= {1，2，3，…，100}

按定义，事件 A 发生的概率为

2.概率的统计定义。在概率的古典定义中，我们是假设每个结果具有等可能性，然后赋予相等的概率，最后再定义概率。这实际上是一个概率概念由另一个概率概念来定义。在多数实际问题中，要将全部结果列举出来往往是不可能的，同时试验结果等可能性的假定也很难成立。此时，人们只有利用实际频数数据来估计概率，这就是概率的统计定义，这种概率也称为频率概率或试验概率。

定义 3.2 在同一条件组下重复进行 n 次试验，当试验次数 n 充分大时，事件 A 发生的频率 $f(A)=\frac{m}{n}$ 趋向于某一数值 P 或稳定地在 P 值附近波动(0≤P≤1)，则定义 p 为事体 A 发生的概率，记作

$$P(A)=P$$

需要说明的是，频率是大量试验的结果，它是一个随着试验次数变化而变化的数值。而概率是一个确定的数值。频率随着试验次数的无限增加，以一种趋势无限接近概率。历史上很多人做出掷硬币的试验，表 3.1 的数据就证明了概率的统计定义。

表 3.1　　历史上数学家们掷硬币试验的数据

试验者	试验次数 n	正面向上次数 m	正面向上的频率 $f(A)=\frac{m}{n}$
蒲　丰	4040	2048	0.5069
K.皮尔逊	12000	6019	0.5016
K.皮尔逊	24000	12012	0.5005

表中试验结果正面向上的频率 f(A)随着次数 n 的增加，越来越接近 0.5 这个数值。因而可用 p＝0.5 作为事件的概率。

3.概率的主观定义。不论是古典概率还是频率概率都是按照大量重复试验的结果来确定概率的。然而在许多实际应用中，特别是在充满不确定性因素的经济问题中，不存在大量重复性过程，决策人所面对的是仅发生一次的事件或者在不相同条件下重复发生的事件，在该种情况下要对即将发生的某事件的可能性进行估计，就需要应用主观概率。

定义 3.3 面对不确定性，由个人判断某事件发生的可能性称为主观概率。主观概率有两个特点：一是由于主观概率直接依赖于观察者，因此对于同一事件，不同的人可能给出不同的概率，但这并不是说主观概率可以随意确定，它应该是以有理智的决策人的经验为根据的；二是由于主观概率依赖于个人判断，因而前人的经验、自己的知识及其对事件的分析都是作出判断的根据。

例3.2一位经济学家对本世纪最后10年某国在世界经济中的地位进行预测，他认为其经济地位“无变化”和“改善”的可能性大致相同，同时又估计“改善”的可能将是“恶化”可能性的 2 倍。他估计的样本空间是什么？每一样本点的概率是多少？

解：样本空间的集合

$$S=\{无变化，改善，恶化\}$$

由于经济状况只有这三种可能，因而这是一个完备的事件组，即

$$P(无变化)+P(改善)+P(恶化)=1$$

并且已知

$$P(无变化)=P(改善)=2\cdot P(恶化)$$

不难得出

$$P(无变化)=\frac{2}{5}，P(改善)=\frac{2}{5}，P(恶化)=\frac{1}{5}$$

以上三种概率定义都有其特定的应用范围，同时也有其局限性。古典概率的等可能性不能保证一定存在；频率概率的试验次数 n 无论取多大，其概率值也只能是一种近似估计；主观概率因人因事而异，似乎缺乏客观性。在概率论理论研究中，以上三种定义都缺乏严密性。随着集合论、测度论等数学理论的发展,在本世纪30年代,苏联数学家柯尔奠哥洛关(A.N. Kolmogorov)提出了概率论的公理化体系，从而使概率的定义纳入到近代数学一般概念的体系之中，为概率论数学理论严谨的逻辑推理打下了坚实的基础。

定义 3.4 设 E 为随机试验，S 是它的样本空间。对于 E 的每一事件 A 赋予一实数，记为 P(A)，称为事件 A 的概率，如果它满足下列条件

(1)对于每一事件 A 有 $0\le P(A)\le 1$

(2)$P(S)=1$

(3)对于两两互不相容的事件 A_i ($i=1，2，\cdots$)

则有

$$P(\sum_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i) \tag{3.2}$$

(3.2)式称为可列可加性或完全可加性。

§3.2　概率运算法则

由概率论公理化体系的三个条件(或称为三条公理)就可以推出概率的若干性质，这些性质同时又是概率运算的法则，其中最重要的法则就是加法公式和乘法公式。

§3.2.1　加法公式

在给出加法公式之前，先介绍事件之间的关系及其运算。

1.和事件。事件 A 与事件 B 中至少有一个发生，称为 A 与 B 的和事件，用 A+B 表示，其直观的表示见图 3.1。例如在 10 件产品中，有 8 件正品，2 件不合格品。从中任取 2 件，A 表示“恰有一件不合格品”，B 表示“恰有 2 件不合格品”，C 表示“至少有一件不合格品”，则有 C＝A＋B。

2.积事件。事件 A 与 B 均发生的事件称为 A 与 B 的积事件,用 AB 表示(见图 3.2)。例如某企业的职工中，A 表示女职工，B 表示管理人员，C 表示女管理人员，则有 C=AB。

3.事件 A 发生而事件 B 不发生的事件称为 A 与 B 之差，用 A－B 表示(见图 3.3)。例如在上例的某企业中，D 表示非管理人员的女职工，即 D=A-C。

4.如果事件 A 与事件 B 不可能同时发生，即 A 与 B 同时发生的事件为不可能事件，AB＝ϕ，则称 A 与 B 互斥(互不相容)(见图 3.4)。例如在上例中，非管理人员的女职工与管理人员不可能同时发生，即 DB＝ϕ。

在掌握了事件之间的运算关系后，概率的运算关系就很容易理解和记忆了。

性质 1　若 AB＝ϕ，则有

$$P(A+B)=P(A)+P(B) \tag{3.3}$$

(3.3)式将和事件的概率转化成了概率之和。同时，(3.3)式是(3.2)式的简化。

例 3.3 已知 10 个灯泡中有 3 个次品，现从中任取 4 个，那么至少有 2 个次品的概率是多少？

解：以 B 表示取出的 4 个灯泡中至少有 2 个次品；以 B1 表示取出的 4 个灯泡中恰有 2 个次品；以 B2 表示取出的 4 个灯泡中恰有 3 个次品。

显然，$B＝B_1＋B_2$，并且 $B_1B_2＝\phi$ 即 B_1 和 B_2 互斥，则 $P(B_1)$和 $P(B_2)$的概率可分别由组合的关系式求出

$$P(B_1)=\frac{C_3^2\cdot C_7^2}{C_{10}^4}=\frac{3}{10},P(B_2)=\frac{C_3^3\cdot C_7^1}{C_{10}^4}=\frac{1}{30}$$

再由加法公式得到

$$P(B)=P(B_1+B_2)=P(B_1)+P(B_2)$$
$$=\frac{3}{10}+\frac{1}{30}=\frac{1}{3}$$

性质 2　若 A＋B=S，AB=ϕ，则称 A、B 互为逆事件，用 $\overline{A}$ 表示 B。由加法公式(3.3)可知

$$1=P(S)＝P(A+\overline{A})＝P(A)＋P(\overline{A})$$

于是有

$$P(\overline{A})＝1－P(A)$$

或

$$P(A)=1-P(\overline{A}) \tag{3.4}$$

在例3.3中如果要计算最多有1个次品的概率，就可以通过求事件B概率的逆事件概率得出。

以A表示取出的4个灯泡中最多有1个次品，则A＋B＝S，AB＝ϕ，即AB互逆。A也可用$\overline{B}$表示。由(3.4)式有

$$P(A)=P(\overline{B})=1-P(B)=1-\frac{1}{3}=\frac{2}{3}$$

当然，也可以利用加法公式直接求出事件A的概率

$$P(A)=\frac{C_3^0\cdot C_7^4}{C_{10}^4}+\frac{C_3^1\cdot C_7^3}{C_{10}^4}=\frac{1}{6}+\frac{3}{6}=\frac{2}{3}$$

这一公式表明，如果直接计算事件A的概率比较困难，或者已知其逆事件$\overline{A}$的概率，可以利用(3.4)式求出事件A的概率。

性质3　在性质1中，若AB≠ϕ，即A、B为任意两个随机事件，则有

$$P(A+B)=P(A)+P(B)-P(AB) \tag{3.5}$$

性质3又称为广义的概率加法公式。

例3.4 某企业职工中女职工占60%，管理人员占20%。从该企业职工中任选一人，是女职工或管理人员的概率是多少？

解：令A表示女职工，B表示管理人员。由于任一职工既可能是女职工又可能是管理人员，则根据已知条件有

$$P(A)=0.60,P(B)=0.20,P(AB)=0.60\times0.20=0.12$$

根据(3.5)式，是女职工或管理人员的概率为

$$P(A+B)=0.60+0.20-0.12=0.68$$

§3.2.2　乘法公式

乘法公式与加法公式都是概率中的基本公式，要了解乘法公式，首先需要掌握条件概率。

1.条件概率。首先给出定义。

定义3.5 设A、B是两个事件，在已知B发生的条件下，A发生的概率称为A对于B的条件概率，用P(A|B)表示。例如在10件产品中有7件正品，3伴次品。在7件正品中有5件一等品，2件二等品。现从中任取一件，以B表示取到合格品，A表示取到一等品，则显然有 $P(A)=\frac{5}{10}$。若已知取到的是合格品，此时取到一等品的概率就应为 $P(AB)=\frac{5}{7}$。对于较一般的情况，需要利用乘法公式来计算条件概率。

2.乘法公式。首先来看公式

$$P(AB)=\frac{P(AB)}{P(B)} \tag{3.6}$$

(3.6)式给出了计算条件概率的一般方法。由图3.2不难看到，在事件B已经发生的条件下，事件A的概率应该是阴影部分(AB)占事件B的比率，即(3.6)式给出的比值。将(3.6)式进行转换，就是乘法公式的一般形式：

$$P(AB)=P(B)\cdot P(A|B) \tag{3.7}$$

当然，也可以将A和B的位置对调，即

$$P(AB)=P(A)\cdot P(B|A) \qquad (3.8)$$

例 3.5 在例 3.3 中，从中依次抽出两个灯泡，两个都是次品的概率是多少？

解：用 A_1 表示第一次抽到的是次品，A_2 表示第二次抽到的是次品，于是有

$$P(A_1)=\frac{3}{10},P(A_2|A_1)=\frac{2}{9}$$

由(3.8)式可得到

$$P(A_1A_2)=P(A_1)\cdot P(A_2|A_1)=\frac{3}{10}\times\frac{2}{9}\times\frac{1}{15}$$

3.独立性。首先给出定义。

定义 3.6 若 A、B 两个事件满足条件

$$P(AB)=P(A)\cdot P(B) \qquad (3.9)$$

则称 A、B 独立。

由 (3.9) 可得 $\frac{P(AB)}{P(B)}$=P(A) 或 $\frac{P(AB)}{P(A)}$=P(B)，对照 (3.6) 式，得到 P(AB)=P(A)，P(BA)=P(B)。这说明当两个事件独立时，A 发生后对 B 发生的概率没有影响，B 发生后对 A 发生的概率也没有影响。

例 3.6 在例 3.3 中，如果将第一次抽出的灯泡放回，连续抽两次且两次都是次品的概率是多少？

解：用 A1 和 A2 分别表示第一次和第二次抽到次品灯泡，由于第一次抽出次品灯泡后又将其放回，第二次仍是从 10 个灯泡中抽，故第一次抽出的结果对第二次没有影响，A1 与 A2 独立。由(3.9)式得

$$P(A_1A_2)=P(A_1)P(A_2)=\frac{3}{10}\times\frac{3}{10}=\frac{9}{100}$$

4.全概率公式。若事件 A1，A2，…AK 为一完备事件组，即满足条件

(1)$A_1,A_2,\cdots,A_k$ 两两互斥，$P(A_i)>0$；

(2)$A_1+A_2+\cdots+A_k=S$

则对任一事件 B 都有

$$\begin{aligned}P(B)&=P(A_1B)+\cdots+P(A_kB)\\&=\sum_{i=1}^{k}P(A_i)P(BA_i) \qquad (3.10)\end{aligned}$$

(3.10)式称为全概率公式。

例 3.7 某企业有甲、乙、丙三个车间生产同一种产品。各车间的产量分别占全厂总产量的 20%、30%、50%。根据过去产品质量检验记录知道甲、乙、丙车间的次品率分别为 4%、3%、2%，从该厂产品中随机抽选一件为次品的概率是多少？

解：令 A1、A2、A3 分别表示甲、乙、丙车间产品，B 表示产品为次品，则有

$P(A_1)=0.20,P(A_2)=0.30,P(A_3)=0.50$

$P(B|A_1)=0.04,P(BA_2)=0.03,P(BA_3)=0.02$

由公式(3.10)，有

$P(B)=P(A_1)\cdot P(BA_1)+P(A_2)\cdot P(B|A_2)+P(A_3)\cdot P(BA3)$

$=0.20\times0.04+0.30\times0.03+0.50\times0.02$

$=0.027$

§3.3 贝叶斯公式

贝叶斯公式是英国数学家 T·贝叶斯(T.Bayes)在 200 多年前提出的一个计算条件概率的公式，也称为逆概率公式。

如果事件组 A_1，A_2，…AK 为一完备事件组，则对任一事件 B(其 P(B)≠0)，有

$$P(A_iB)=\frac{P(A_i)\cdot P(BA_i)}{\sum_{i=1}^{k}P(A_i)P(BA_i)},i=1,2\cdots k \qquad (3.11)$$

(3.1)式即贝叶斯公式。由于

$$\begin{aligned}P(A_i|B)&=\frac{P(A_iB)}{P(B)}\\&=\frac{P(A_i)P(B|A_i)}{P(A_1B)+\cdots+P(A_kB)}\\&=\frac{P(A_i)P(B|A_i)}{P(A_1)P(BA_1)+\cdots+P(A_k)P(BA_k)}\\&=\frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{k}P(A_i)P(B|A_i)}\end{aligned}$$

不难发现贝叶斯公式的分子应用了乘法公式，分母既应用了加法公式也应用了乘法公式。在贝叶斯公式中，$P(A_i)$称为验前概率或先验概率；$P(B|A_i)$一般来自样本所提供的信息；$P(A_i|B)$称为后验概率。这样，贝叶斯公式实际上是综合利用先验概率和样本信息从而计算后验概率的一种计算方法。将这一基本公式广泛应用于参数估计、假设检验、统计预测及决策等，就形成了贝叶斯统计以及与频数学派相对立的贝叶斯统计学派。

例 3.8　在例 3.7 中，若从该企业的产品中任抽一件发现为次品，该产品是乙车间生产的概率是多少？

解：这个问题是在已知抽中次品 B 的条件下求乙车间生产的后验概率。应用例 3.7 的结果于(3.11)式，有

$$\begin{aligned}P(A_2B)&=\frac{P(A_2)P(BA_2)}{\sum_{i=1}^{3}P(A_i)P(BA_i)}\\&=\frac{0.30\times0.03}{0.20\times0.04+0.30\times0.03+0.50\times0.02}\\&=\frac{0.009}{0.027}\\&=\frac{1}{3}\\&=0.33\end{aligned}$$

用同一公式算出该产品是甲车间和丙车间生产的概率分别为 0.29 和 0.38。可以看到，甲、乙、丙三个车间生产的后验概率不同于先验概率，这是三个车间次品率不同的结果。

§3.4　随机变量及其分布

§3.4.1　随机变量

在本章第 1 节中讨论了随机试验，并把可在相同条件下重复进行，然而在事先不能确定其结果的试验称为随机试验。要对随机事件所构成的随机现象进行更深入的研究，就需要对随机试验的结果给予数量的描述。统计学中用随机变量对其进行描述并作为进一步研究推断的基础。

例 3.9 在 10 件同一类型产品中，有 2 件次品。现任取 2 件，则这 2 件中的次品数就是一个随机变量，用ξ或 X 表示。随机变量的取值是随机的，本例中可能取到 0，1，2 三个值。当“X＝0”时，表示任取的 2 件中没有次品；当“X＝1”时，表示任取的 2 件中有一件次品；当“X=2”时，表示任取的 2 件中全是次品。由本章前面所介绍的知识可以计算随机变量取各种值时的概率为

$$P(X=0)=\frac{C_8^2}{C_{10}^2}=\frac{28}{45}$$

$$P(X=1)=\frac{C_2^1C_8^1}{C_{10}^2}=\frac{16}{45}$$

$$P(X=2)=\frac{C_2^2}{C_{10}^2}=\frac{1}{45}$$

例 3.10 某一工人所生产的产品出现一级品的可能性为 P。现在连续生产 5 件产品，其中一级品件数就是一个随机变量。它的可能取值为 0，1，2，3，4，5。“X=i”意味着在 5 件产品中：“一级品有 i 件”。由于每件产品的生产是独立进行的，由概率乘法公式和独立性可得

$$P(X=i)=C_5^i p^i(1-p)^{5-i}, i=0,1,\cdots,5$$

从这两个例子可以看出，随机变量 X 取某一个值，就相当于某一随机事件。随机变量 X 取某一个值的概率就相当于某一随机事件的概率。显然，随机变量具有下列特征：第一，随机变量的取值是随机的，事先或试验前不知道取哪个数值；第二，随机变量取具体值的概率大小是确定的。

在本教材中，我们用大写 X，Y，Z，…表示随机变量，用小写 x，y，z，…表示随机变量的取值。

随机事件或随机现象从其数量测度的角度可以分为两类：一类是比较容易或可直接用数量加以描述的。例如测量误差的大小，某一产品的使用寿命，某一零件的规格尺寸等等：另一类似乎与数量无关或不易直接用数量加以描述，例如一台机床在 8 小时内连续工作是否会发生故障，某一新产品是否受到用户欢迎，一次技术测验是否及格等等。

对于前一类能用数量加以描述的随机事件，一般用随机变量 X 直接加以描述。对于不易直接用数量加以描述的随机事件，可以人为地给不同试验结果赋值，如可规定

$$X=\begin{cases}1 & \text{新产品受到欢迎}\\ 0 & \text{新产品不受欢迎}\end{cases}$$

这样经赋值后就可同前一类随机事件一样用随机变量加以描述并可进一步进行统计分析。

当用随机变最对随机事件或随机现象进行数量描述时，由于随机变量取值特点的不同，又可以将它分为两类：第一类是离散型随机变量，即所有随机变量的取值都可以一一列举。例如一批产品中的次品个数，某一新产品是否受到用户欢迎等，称之为离散型随机变量。第二类是连续型随机变量，即随机变量依照一定的概率规律在连续的数轴上取任一数值，它的取值不能一一列举，只能取某一区间内的全部数值。例如测量误差的大小，某一产品的使用寿命等等，这类随机变量都是连续型随机变量。

统计学中的随机变量与微积分中的变量不尽相同。在微积分中变量 X 的取值是确定的，如 X＝6。而统计学中的随机变量 X 的取值是随机的，即它是以一定的概率取某一数值。因而当说随机变量 X=6 时，总是伴随着一定的概率，即考虑 P(X=6)。描述随机变量这一特点的是随机变量的概率分布密度，不同类型的随机试验有不同的分布密度。

§3.4.2　分布密度

随讥变量的分布密度是对随机变量取值特点和取值规律的一种描述，它反映了一个随机变量取值变化的数量规律性。由于随机变量按其取值特点可以分为离散型和连续型随机变量，所以其分布密度也有相应的表现形式。

1.离散型随机变量的分布密度。

定义 3.6 离散型随机变量 X 的所有可能取值为离散值 x_i（$i＝1,2\cdots k\cdots$），并且已知 X 取 x_i 的概率为 p_i，即

$$P(X=x_i)=p_i \qquad i=1,2,\cdots k\cdots$$

则称函数

$$f(x)=\begin{cases} p_i & i=1,2,\cdots k\cdots \\ 0 & \text{其他} \end{cases} \tag{3.12}$$

为 X 的分布密度，简称密度，亦称为分布列，用 X～f(x)表示。

离散型随机变量的分布密度有如下性质

(1)　$f(x)\geq 0$,即 $p_i\geq 0$(因为概率 $p_i\geq 0$)

(2)　$\sum\limits_{i=1}^{\infty} f(x_i)=\sum\limits_{i=1}^{\infty} p_i=1$(因为必然事件的概率为1)

离散型随机变量的分布密度除了可用上述定义公式表述外，还可用表格和图形表示，例如在例 3.9 中已算出共分布密度

$$p_1=P(X=0)=\frac{28}{45}$$

$$p_2=P(X=1)=\frac{16}{45}$$

$$p_3=P(X=2)=\frac{1}{45}$$

可用表格和图形来分别表示(见表 3.2 和图 3.5)。

表 3.2　　抽中次品 X 的分布密度

$X=x_i$	$x_1=0$	$x_2=1$	$x_3=2$
$P(X=x_i)=p_i$	$\frac{28}{45}$	$\frac{16}{45}$	$\frac{1}{45}$

图 3.5　抽中次品 X 的分布密度图

2.连续型随机变量的分布密度。

定义 3.7 连续型随机变量 X 在(－∞，+∞)内取值，如果存在可积函数 f(x)，使得 X 依照概率规律

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx \qquad x_1, x_2 \in (-\infty, +\infty) \tag{3.13}$$

则称 f(x)为 X 的分布密度。

连续型随机变量的性质与离散型随机变量的性质相似

$$(1) f(x) \ge 0$$

$$(2) \int_{-\infty}^{+\infty} f(x)dx = 1$$

不论是离散型还是连续型随机变量，性质(1)和(2)都是分布密度的充要条件，只要满足这两条性质，它就是一个分布密度。

连续型随机变量的分布密度由于其随机变量的取值是连续的，不可列的，因而难于用表的形式表示，但可用图的形式表示其随机变量的规律。

§3.4.3　分布函数

随机变量的分布密度是对随机变量取值及其变化规律的最基本的描述。要计算随机变量取某些数值或某个区间的数值就需要利用分布密度进行求和或积分的计算，同时也就用到随机变量的分布函数。

定义 3.8 设 X 为一随机变量，x 为任意实数，称函数

$$F(x)=P(X<x) \qquad (-\infty<x<+\infty) \tag{3.14}$$

为随机变量 X 的分布函数，记为 X～F(x)。

本教材对分布函数的定义为 P(X＜x)，也有些书定义为 P(X≤x)。对于连续型随机变量，由于在 X＝x 点取值为零，因而这两种定义相同。对于离散型随机变量，两种定义的差别在于是否包括 x 这一点。考虑到在我国实际统计工作中分组时常用上组限不在内的原则.因而在这里采用 P(X<x)的定义。

从定义中不难发现，分布函数 F(x)实质是随机变量取值小于 x 的部分占全部的“百分比”。这一“百分比”是通过对分布密度的求和得到的。离散型与连续型分布函数的计算过程为

离散型　$F(x) = P(X < x_k) = \sum_{i=1}^{k-1} f(x_i)$

连续型　$F(x) = P(X < x) = \int_{-\infty}^{x} f(x)dx$

分布函数 F(x)有如下性质

(1) $0 \le F(x) \le 1$ (**因为**$F(x)$**是某种概率**)

(2) $F(+\infty) = \lim\limits_{x \to +\infty} p(X < x) = 1$

$F(-\infty) = \lim\limits_{x \to +\infty} p(X < x) = 0$

(3) **若取**$x_1 = a, x_k = b,$

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

或

$$\sum_{a \le x_i < b} f(x_i) = F(b) - F(a)$$

例如在例 3.9 问题中，我们已经计算了抽中次品 X 的分布密度，现在进一步计算其分布函数(见表 3.3)。

表 3.3　　抽中次品 X 的分布函数

$X=x_i$	**分布密度** $p_i=P(X=x_i)$	**分布函数** $F(x_i)=P(X<x_i)$
0	$\frac{28}{45}$	0
1	$\frac{16}{45}$	$\frac{28}{45}$
2	$\frac{1}{45}$	$\frac{44}{45}$
3	0	1
合计	1	—

§3.5 几种重要的随机变量分布

§3.5.1 离散型随机变量分布

1.两点分布。若互相独立的重复试验只有“成功”和“失败”两种结果，这种试验称为贝努里试验，可取

$$X=\begin{cases}1 & \text{成功}\\ 0 & \text{失败}\end{cases}$$

例如掷硬币、产品质量(合格品和次品)检验等问题。可用 P(X=1)=p 和 P(X=0)=1-p 表示，则分布密度为

$$P(X=x)=\begin{cases}p & x=1\\ 1-p & x=0\end{cases} \tag{3.15}$$

2.超几何分布。在 N 件产品中有 M 件次品，若从中任意抽取 n 伴，抽到的次品数 X 为一个随机变量并服从分布密度为

$$P(X=x)=\frac{C^{x}\cdot C_{N-M}^{n-x}}{C_{N}^{n}}\; x=0,1,\cdots \min(n,M) \tag{3.16}$$

的超几何分布。例 3.9 中概率的计算就是利用了(3.16)式。

3.二项分布。若重复进行 n 次贝努里试验，则“成功”次数 x 服从分布密度为

$$P(X=x)=C_{n}^{x}p^{x}(1-p)^{n-x} \quad x=0,1,\cdots,n \tag{3.17}$$

其中 p 满足 0<p<1 的二项分布，记作 X～b(n，p)。在例 3.10 中，连续生产 5 件产品，其中一级品件数 i 即服从二项分布。当超几何分布中的 N 很大，而 n 又比较小，即只要 $\frac{n}{N}$<0.10，就可用二项分布作为超几何分布的近似。

4.普哇松分布。当二项分布中的 n 相当大，P 又较小时，二项分布可用λ=np 的普哇松分布来近似。普哇松分布的密度为

$$P(X=x)=\frac{\lambda^{x}}{x\,!}e^{-\lambda} \tag{3.18}$$

$$x=0,1,2\cdots,\lambda>0$$

普哇松分布同超几何分布、二项分布一样，在经济管理中有着广泛的应用。例如对某种产品和服务的需求，电话交换台的呼叫次数，纺织品中的疵点数等都服从普哇松分布。

例 3.11 由某商店的销售记录可知平均每星期售出 4 台某种型号的电视机。并已知该销售量服从普哇松分布，该商店每星期仅售出两台电视机的概率是多少？

解：已知λ=4，则用(3.18)式有

$$P(X=2)=\frac{4^{2}}{2\,!}e^{-4}=0.15$$

即该商店恰售出两合电视机的概率为 0.15。

§3.5.2 连续型随机变量分布

1.正态分布。正态分布是最重要的随机变量的分布之一，特别是在统计方法的应用中，正态分布起着十分重要的作用。

首先，正态分布是最常见的分布，大量客观现象服从或近似服从正态分

布。人的生理特征如身高、体重、智商(IQ)、骨骼的长度；自然现象中如洋海的深度、积雪的厚度；产品的质量性能如电子管中的热噪声、电流、电压、纤维长度、细纱强度、钢的含碳量及抗拉强度；粮食作物的产量；某些产品寿命等。所有这些现象，其试验、观测和测量误差都服从正态分布。

其次，正态分布具有很好的数学性质。这一方面许多非正态分布以正态分布作为其渐近分布或极限分布，例如二项分布、普哇松分布等都以正态分布作为其渐近分布；另一方面任意样本均值的分布都以正态分布为其极限分布，这就是概率论的中心极限定理所证明的。具体他说，当总体是正态分布时，则不论样本容量大小，样本均值都服从正态分布；当总体不服从正态分布时，只要样本容量足够大，样本均值的分布也趋于正态分布。

正态分布的概率密度为

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e-\frac{(x-\mu)^2}{2\sigma^2} \qquad -\infty<x<+\infty \qquad \sigma>0 \tag{3.19}$$

简记作 $X \sim N(\mu,\sigma^2)$。式中μ,σ^2是两个参数，当$\mu=0$，$\sigma=1$ 时，相应的分布 N(0，1)称为标准正态分布。正态分布的几何形状如同一口古钟(见图 3.6)，因而也称为钟形分布。

从图 3.6 中不难看出正态分布具有如下特征

(1)正态分布密度以 $x=\mu$为对称轴，并在 $x=\mu$处达到最大值。

(2)当 $x\to\pm\infty$时，$f(x)\to 0$,故 f(x)以 x 轴为渐近线。

(3)用求导方法可知，$x=\mu\pm\sigma$为 f(x)的两个拐点，σ为拐点到对称轴σ的距离。如果σ较大，则曲线较平缓；如σ较小，则曲线较陡峭。

图 3.6 正态分布的概率密度曲线

正态分布对称轴μ左右曲线下的面积各为$\frac{1}{2}$，经计算知

$$P(\mu-\sigma \le X \le \mu+\sigma)=0.6826$$
$$P(\mu-2\sigma \le X \le \mu+2\sigma)=0.9545$$
$$P(\mu-3\sigma \le X \le \mu+3\sigma)=0.9973$$

为了方便，人们编制了$\mu=0$，$\sigma=1$ 的标准正态分布函数表，这样就可以直接从表中查出正态分布下各区间的面积。对于非标准正态分布的正态变量 X，需做如下变换

$$z=\frac{X-\mu}{\sigma} \tag{3.20}$$

变换后的 z 变量即标准正态变量，即 $z \sim N(0，1)$。于是标准正态变量的密度函数为

$$\varphi(z)=\frac{1}{\sqrt{2\pi}}e-\frac{z^2}{z} \tag{3.21}$$

标准正态变量的分布函数用$\Phi(z)$或$\Phi(x)$表示，其数学表达式为

$$\Phi(z)=\int_{-\infty}^{z}\varphi(z)dz=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e-\frac{t^2}{2}dt \tag{3.22}$$

这个慨率可从附表 1 中直接查到，即直接查表$\Phi(z)$或$\Phi(x)$。

例 3.12 某企业对生产中某关键工序进行调查,发现工人们完成该工序的时间服从正态分布。均值为 20 分钟，标准差为 3 分钟。

(1)从该工序生产工人中任选一人，其完成该工序时间少于 17 分钟的概率是多少？

(2)为安排和保证其它工序的连续性，要求以 95%的概率保证该工序生产时间不多于 25 分钟，这一要求能否保证？

(3)为鼓励先进帮助生产工人提高技术水平,拟奖励核工序生产工人中生产时间用得最少的 10%的工人，奖励的标准应定在什么时间以内？

解：已知 X～N(20,3^2)

$$(1)\quad P(X<17)=F(17)=\phi(\frac{17-20}{3})$$

$$=\phi(-1)=0.1587$$

从该工序生产工人中任选一人，生产时间少于 17 分钟的概率为 15.87%。

$$(2)\quad P(X\le 25)=F(25)=\phi(\frac{25-20}{3})$$

$$=\phi(1.67)=0.9525$$

由于计算概率 0.9525 大于要求概率 0.95，因而其要求可以得到保证。

$$(3)\quad P(X<x)=F(x)=\phi(\frac{x-20}{3})=0.10$$

查附表得 z=－1.28，即

$$Z=\frac{x-20}{3}=-1.28 \qquad x=16.16$$

该工序生产时间少于 16.16 分钟的概率为 10%，应奖励所用生产时间少于 16 分钟 10 秒的那些工人。

很明显，在正态分布中靠近均值周围出现的变量值频数较高，其变量值落在$\mu\pm3\sigma$外的概率不到 1%。这是一个很小的概率，就一次随机试验或观察来讲，它发生的可能性太小了，一般认为不可能发生。一旦这种小概率事件真正发生了，则有理由认为该变量值不是来自这一正态分布，而是有其它系统性原因导致。这就是著名的小概卒原理或 3σ准则，它广泛应用于产品质量控制、产品质量检验和日常生活之中。

2.指数分布。指数分布主要是在产品寿命分析中应用，其密度为

$$f(t)=\begin{cases}\lambda e^{-\lambda t} & t\ge 0,\\ 0 & t<0\end{cases} \qquad \lambda>0 \qquad (3.23)$$

称随机变量 T 服从参数为λ的指数分布。

例 3.13 在自动化生产线上某工序平均每分钟装配 4 件产品。任意两件产品间隔时间不超过半分钟的概率(已知间隔时间 T 服从指数分布)是多少？

解：令间隔时间为 T，已知λ=4

则

$$P(T\le 0.5)=\int_0^{0.5}\lambda e^{-\lambda t}dt=\int_0^{0.5}4e^{-4t}dt$$

$$=-e^{-4t}\Big|_0^{0.5}=-e^{-2}+1=0.8647$$

3.均匀分布。随机变量 X 在有限区间(a，b)内取值，且其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其他}\end{cases} \qquad (3.24)$$

则称 X 在区间(a，b)上服从均匀分布。

例 3.14 设电阻的阻值 R 是一个随机变量，均匀分布在 1000～1500 欧姆之间，R 的概率密度及 R 落在 1200—1300 欧姆之间的概率是多少？

解：根据(3.24)式，R 的概率密度为

$$f(r)=\begin{cases}\dfrac{1}{1500-100} & 1000<r<1500\\ 0 & \text{其他}\end{cases}$$

则有 $$P(1200<R<1300)=\int_{1200}^{1300}\frac{1}{500}dr=0.2$$

§3.6 二维随机变量及其分布

前面几节讨论的是一个随机变量的情况，然而在实际问题中，有些随机现象用一个随机变量来描述是有困难的。例如研究企业的经济效益，就可以用劳动生产率、成本降低率、销售利税率和资金利税率等多个变量来反映；又例如电视机的质量，就需要考虑耐用时间、画面清晰度、色彩准确度等多项指标。而且对于这些变量和指标不仅要逐个进行分析，还要进行综合的考虑。这时，就必须考虑多维随机变量的问题。本节只简单地给出二维随机变量及其分布。

§3.6.1 联合分布

设 X、Y 是两个随机变量，则称

$$F(x,y)=P(X<x,Y<y) \tag{3.25}$$

为(X，Y)的联合分布函数。

与一维随机变量的情况相同，二维随机变量(X，Y)也可分为离散型和连续型两种。

(1)如果(X,Y)只能取有限对或可列对数值(a_i,b_j)，i=1,2…k…，j=1,2…k…)，称(X,Y)为离散型的，于是

$$F(x,y)=\sum_{a_i<x}\sum_{b_j<y}P(X=a_i,Y=b_j) \tag{3.26}$$

(2)如果有密度 f(u,v)使随机变量(X,Y)

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv \tag{3.27}$$

成立，则称(X,Y)为连续型的。同样，称(3.26)式中的 $P(X=a_i,Y=b_j)=p_{ij}$ 和(3.27)式中的 f(u,v)为联合密度。对于连续型分布亦有

$$f(x,y)=\frac{\partial^2F(x,y)}{\partial x\partial y} \tag{3.28}$$

例 3.15 连续掷两枚完好的硬币，令 X 表示两次试验中正面出现的次数，则 X=0，1，2。令 y 表示第一次结果是哪一面。若第一次结果是反面，则 Y=0。若第一次结果是正面，则 Y＝1。于是我们可得到表 3.4 的联合分布。若两次试验结果都是正面，则从联合分布表中可找到 P(X＝2，y＝1)=$\frac{1}{4}$。

表 3.4　　　　（X,Y）的联合分布表

X=x	Y=y 0	1	P(x)
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
P(y)	$\frac{1}{2}$	$\frac{1}{2}$	1

§3.6.2 边缘分布

联合分布 $F(x,y)$ 反映的是 X 和 Y 分别取值的情况。若仅考虑二维随机变量中单个随机变量 X 或 Y 的自身分布 $F_x(x)$ 或 $F_y(y)$，就是 X 和 Y 的边缘分布。由(3.26)式，可分别求出 X 或 Y 的边缘分布

$$P(X=a_i)=\sum_j P(X=a_i, Y=b_j)=\sum_j p_{ij}$$

$$P(Y=b_j)=\sum_i P(X=a_i, Y=b_j)=\sum_i p_{ij}$$

例 3.15 的边缘分布见表 3.4 的最下一行和最右一行。

如果要做进一步的分析研究，还可应用条件概率的公式(3.6)求出二维随机变量的条件分布。

§3.7 随机变量的数字特征

在第 2 章中，曾讨论论过数据的代表性数值，如均值、方差、标准差等。本节要讨论的是随机变量的代表性数值，即期望和方差，数据的均值与随机变量的期望都是一般水平的代表值。不侗的是，数据的均值是从实际资料的角度对数据的一般性水平进行描述，而随机变量的期望则是从理论分布的角度对分布的一般性水平进行描述。数据的方差与随机变量的方差都是对其离散(或差异)程度的测度，不同点在于前者是测度实际数据，后者则是描述理论分布。

§3.7.1 期望与方差

定义 3.9 设 X 是离散型随机变量，X 取 x_1，…x_k…，相应的概率为 p_1…p_k…,则称

$$E(X)=\sum_i x_i p_i \tag{3.29}$$

为 X 的期望值。称

$$D(X)=\sum_i (x_i - E(X))^2 p_i \tag{3.30}$$

为 X 的方差。

设 X 是连续型随机变量，X～f(x)。则称

$$E(X)=\int_{-\infty}^{+\infty} x f(x)dx \tag{3.31}$$

为 X 的期望值。称

$$D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2 f(x)dx \tag{3.32}$$

为 X 的方差。

离散型随机变量的期望和方差与数据的均值和方差非常相似，不同之处只在于期望和方差中用概 p_i 代替了数据的比重权数 $\frac{f_i}{\Sigma f_i}$。当观察值或数据充分大时，比重权数 $\frac{f_i}{\Sigma f_i}$ 这也反映了实际数据与理论数值之间的渐近关系。

如同数据中均值与方差的辩证关系一样，随机变量的期望与方差是描述随机变量的最重要的两个数字特征，二者互相依存、互相补充，共同反映随机变量的变化规律及其特征。

§3.7.2 常用分布的期望与方差

对于本章前面所介绍的几种常用分布，应用公式(3.29)、(3.30)、(3.31)和(3.32)可以得到表 3.5 的结果，其推导过程见参考文献[5]。

对于二维随机变量，除分别研究 X 和 Y 的期望与方差，还需考虑 X 与 Y 之间相互关系的数字特征，主要是协方差和相关系数。

定义 3.10E ｛[X－E(x)][Y－E(Y)]｝称为随机变量 X 与 Y 的协方差，记为 Cov(X , y) , 而

表 3.5　　几种常用分布的数学期望和方差

分布	概率密度	期望	方差
两点分布	$P(X=x)=p^x(1-p)^{1-x}$ $x=0,1\quad 0<p<1$	p	pq
超几何分布	$P(X=x)=\dfrac{C_M^x\ C_{N-}^{n-x}}{C_N^n}$ $x=0，1，\cdots\min(n,M)$	np	$npq\dfrac{N-n}{N-1}$
二项分布	$P(X=x)=C_n^x p^x(1-p)^{n-x}$ $x=0，1，2\cdots n，0<p<1$	np	npq
普哇松分布	$P(X=x)=\dfrac{\lambda^x}{x!}e^{-\lambda}$ $x=0，1，2\cdots\lambda>0$	λ	λ
正态分布	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<\mu<\infty,\sigma>0$	μ	σ^2
指数分布	$f(t)\lambda e^{-\lambda t}$ $t\geq 0,\lambda>0$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
均匀分布	$f(x)=\dfrac{1}{b-a}$ @<x<b,	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$

$$\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$$

称为随机变量 X 与 Y 的相关系数，样本相关系数用 r 表示。相关系数 p 或 r 是表示两个随机变量线性关系程度的测度，在第 9 章相关与回归中还要做详细讨论。

习　题

1.设 A、B、C 表示 3 个随机事件，试将下列事件用 A、B、C 表示出来：

(1)A 出现，B、C 不出现；

(2)A、B 都出现，C 不出现；

(3)所有 3 个事件都出现；

(4)所有 3 个事件都不出现，

(5)3 个事件至少有 1 个出现；

(6)不多于 1 个事件出现；

(7)不多于 2 个事件出现。

2.某工厂生产过程中出现次品的概率为 0.05，每 100 件产品为一批，检查产品质量时是从中任取一半检查，如果发现次品不多于 1 个，则认为这批产品是合格的。每一批产品被认为合格的概率是多少？

3.设某种动物由出生到存活 22 岁的概率为 0.8，活到 25 岁的概率为 0.4，现在年龄为 20 岁的这种动物活到 25 岁的概率是多少？

4.播种时用的一级小麦种子中混有 30%的二级种子和 2%的三级种子。用一级、二级和三级种子长出的穗，含有 50 颗以上麦粒的概串分别是 0.5、0.25 和 0.1，这批种子所结的穗含有 50 颗以上麦粒的概率是多少？

5.某商业部门收购进甲、乙、丙三厂生产的同样规格的产品，在总收购量中甲、乙、丙三厂的产品各占 40%、35%和 25%。甲、乙、丙三厂生产的次品率分别为 1%、2%和 3%。若从总购量中任取 1 件检查，问：

(1)该件产品是次品的概率是多少？

(2)如果抽到的产品是次品,那么所抽到的产品恰好是甲厂生产的概率是多少？恰好是乙厂和丙厂生产的概率各是多少？

6.据某地过去气象记录，在 11 月的 30 天中平均有 2 天是雨天，假定 11 月每天是否下雨如同重复试验一样服从二项分布。

(1)这种假定是否合理？

(2)接二项分布计算，次年 11 月最多有 2 天是雨天的概率是多少？

7.应用普哇松分布计算 500 人中至多有 1 人在元旦出生的概率(假定 1 年是 365 天)。

8.设已知生产脆度过高的钻子(废品)的概率为 0.01，现每盒装 100 个钻子。问：

(1)在 1 盒中没有废品的概率是多少？

(2)在 1 盒中有 1 个废品的概率是多少？

(3)若要求以 99%的概率保证每盒有 100 个合格品，每盒至少要装多少钻子？

9.某厂生产一批小型装置，设已知该小型装置的平均寿命为 10 年，标准差为 2 年。如果该小型装置的寿命服从正态分布，问：

(1)整批小型装置不小于 9 年的比重是多少？

(2)整批小型装置不小于 11 年的比重是多少？

(3)如果工厂规定在保用年限期间遇有故障可免费换新,今要求免费换新率限制在 3%以内，保用年限有多长？

10.某种电开关寿命(年)具有失效率 $K=\frac{1}{2}$ 的负指数分布。若有 100 个此

种开关装在不同系统中，那么在第一年最多有 30 个失效的概率是多少？

11.设收获率服从正态分布,试由表 1 的样本资料计算收获面积的理论次数。

表 1　　　　　　　　某地水稻按收获率分组的面积资料

收获率(斤/亩)	**收获面积(亩)**
300 — 400	12
400 — 500	36
500 — 600	164
600 — 700	235
700 — 800	194
800 — 900	18
900 — 1000	24
合　计	683

第 4 章　参数估计与假设检验

统计学的目的是探索总体的数量规律性。统计方法的特点是通过随机样本对总体作出科学的推断。在对数据进行了描述，提取了样本数据中的信息，从而形成了推断总体的统计量(见第 2 章和第 3 章)以后，就可以依据概率分布和抽样分布的原理对总体进行推断。本章将讨论统计推断的两个最基本的同时又是互相联系的部分，即估计和假设检验的问题

§4.1　抽样分布

在进行统计推断时，由于应用的是随机样本，因而就要考虑样本的抽样分布。所谓抽样分布是指进行重复抽样时，样本的统计量所形成的概率分布。在统计推断中，应用最多的是正态总体中 $\overline{X}$ 的分布及 X^2 分布、t 分布和 F 分布。

§4.1.1　正态总体中 x 的分布

定理 4.1 设 x_1，x_2，…，x_n 是独立同分布随机变量，且每个随机变量服从正态分布 N(μ，σ^2)，则其均值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_4$ 服从正态分布 N($\mu,\frac{\sigma^2}{n}$)。

这个定理是容易理解的，因为独立正态变量之和 $\sum_{i=1}^{n}x_i$ 仍为 正态变量，乘上 $\frac{1}{n}$ 后的 $\bar{x}$ 也还是正态变量。

$$E(\bar{x})=E(\frac{1}{n}\sum_{i=1}^{n}x_i)=\frac{1}{n}\sum_{i=1}^{n}Ex_i=\frac{n\cdot\mu}{n}=\mu$$

$$D(\bar{x})=D(\frac{1}{n}\sum_{i=1}^{n}x_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(x_i)=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

所以 $\bar{x}$ 服从正态分布 N($\mu,\frac{\sigma^2}{n}$)。

§4.1.2　x^2 分布

定理 4.2 设 $x_1,x_2,\cdots,x_n$ 是独立同分布随机变量，而每一个随机变量服从标准正态分布 N(0，1)，则随机变量 $x^2=x_1^2+x_2^2+\cdots+x_n^2$=的分布密度是

$$f(x)=\begin{cases}\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}\quad e^{-\frac{x}{2}} & x>0\\ 0 & x\le 0\end{cases}\tag{4.1}$$

其中 $\Gamma(\frac{n}{2})$ 是 Gamma 函数(Γ函数)在 $\frac{n}{2}$ 处的值。这种分布称为自由度为 n 的 x^2 分布，记为 $x^2(n)$。随机变量 x^2 简称 x^2 变量。

x^2 分布可看成是Γ分布的一个特例，即当参数 a=$\frac{n}{2}$，$\beta=\frac{1}{2}$ 的Γ分布。x^2 分布的密度函数曲线如图 4.1，它随 n 的取值不同而不同。

若对于给定的 a(0＜a＜1)，存在 $x_a^2(n)$ 使

$$\int_{x_a^2(n)}^{\infty}f(x)dx=a$$

则称 $x_a^2(n)$ 为 x^2 分布的上侧分位数，其数值可查附录 2。

图 4.1 x^2 分布的密度曲线

x^2 分布具有如下性质

(1) $E(x^2)=n$，$D(x^2)=2n$

(2)x^2分布具有可加性。若两个x^2变量x_1^2和x_2^2相互独立，$x_1^2 \sim x_{n_1}^2, x_2^2 \sim x_{n_2}^2$，则

$$(x_1^2 + x_2^2) \sim x_{n_1+n_2}^2$$

(3)当 n→∞时，$x_n^2 \to N(n, 2n)$。

在统计推断中，我们常用到下面的定理。

定理 4.3 设 x_1，x_2，…x_n,相互独立，且都遵从 $N(\mu，\sigma^2)$，则有

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$$

且

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim x_{n-1}^2$$

$\bar{x}$与 S^2 相互独立。

若 x_1，x_2，…，x_n 相互独立且都遵从 N(0，1)，则

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \sim x_{n-1}^2$$

§4.1.3 t 分布

定理 4.4 设随机变量 X 服从标准正态分布 N(0，1)，随帆变量 Y 服从自由度为 n 的 x_2 分布，且 X 与 Y 相互独立，则

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

其分布密度为

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{t^2}{n}) - \frac{n+1}{2}, -\infty < t < \infty \tag{4.2}$$

这种分布称为自由度为 n 的 t 分布，简记为 t(n)。随机变量 T 简称 T 变量。

t 分布密度见图 4.2，它随 n 的不同取值而变化。由于(4.2)式中的函数是偶函数，所以 t 分布密度对 y 轴是对称的。由(4.2)式不难推出，当 n→∞时，$f(t) \to \frac{1}{\sqrt{2\pi}} e^{\frac{-t_2}{2}}$ 即，当 n 趋向于无穷时，t 分布趋向于标准正态分布。

图 4.2　t 分布的密度曲线

t 分布的期望和方差分别为

$$E(t) = 0 \qquad D(t) = \frac{n}{n-2} \qquad (n > 2)$$

若对于给定的 a(0<a<$\frac{1}{2}$)，存在正数 $t_a(n)$使

$$\int_{t_a(n)}^{\infty} f(t)dt = a$$

则称 $t_a(n)$为 t 分布的上侧分位数，其数值可查附录 3 一般当 n＞30，可查标准正态分布表作为其近似。

§4.1.4　F 分布

定理 4.5 设 X 和 Y 分别服从自由度为 n_1、n_2的 x^2分布，且 X 与 Y 相互独立，则

$$F = \frac{X / n_1}{Y / n_2}$$

的分布密度为

$$f(z) = \begin{cases} \dfrac{\Gamma(\frac{n_1 + n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}(\frac{n_1}{n_2})(\frac{n_1}{n_2}\cdot z)^{\frac{n_1}{2}-1}(1+\frac{n_1}{n_2}\cdot z)^{-\frac{n_1+n_2}{2}} \\ 0 \end{cases},$$

这种分布称为第一自由度为 n_1、第二自由度为 x_2的 F 分布，或自由度为(n_1，n_2)的 F 分布，记为 F(n_1，n_2)。随机变量 F 简称为 F 变量。

F 分布的密度曲线见图 4.3，图中曲线随 n_1、n_2取不同数值而不同。但 F 分布不以正态分布为其极限，而总是一个正偏形分布。

F 分布的数学期望和方差为

$$E(z) = \frac{n_2}{n_2 - 2} \qquad (n_2 > 2)$$

$$D(z) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)} \qquad (n_2 > 4)$$

图 4.3　F 分布的密度曲线

若对于给定的 a(0<a<1)，存在正数 F_a(n_1，n_2)使

$$\int_{F_a(n_1, n_2)}^{\infty} f(z)dz = a$$

则称 F_a(n_1，n_2)为 F 分布的上侧数，它的数值可查附录 4。F 分布有一个重要性质：如果 F 变量服从分布 F(n_1，n_2)，那么$\frac{1}{F}$服从分布 F(n_2，n_1)，因为$\frac{1}{F}=\frac{Y / n_2}{X / n_1}$。在查表时，有

$$\frac{1}{F_a(n_1 + n_2)} = F_{1-a}(n_2, n_1)$$

当两个正态总体方差$\sigma_1^2 = \sigma_2^2$相等，两个样本容量分别为山和 n2 时，两个样本方差之比服从自由度为(n_1-1，n_2-1)的 F 分布，即$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$。F 分布常用来检验两个正态总体的方差是否相等，并在方差分析中应用。

§4.2　参数估计

参数是总体分布数量规律性的特征值，一般用θ_i，表示。例如正态分布总体 N(μ，σ^2)中的均值μ和方差σ^2就是总体的参数，该总体分布的位置和形状的数量特征就由μ和σ^2决定。在实际问题中，总体参数θ_i通常都是未知的，这就需要通过样本数据所提供的总体的有关信息对参数加以推断。在第 2 章中，通过对样本数据的加工和信息的提取，形成了对样本数据具有代表性的统计量。在进一步的统计推断中，应用样本统计量作为估计量$\hat{\theta}_i$(x_1，x_2，…，X_n)去估计总体参数θ_i。如果在估计中直接用估计量$\hat{\theta}_i$作为固定的数值对参数θ_i做出估计，就是参数的点估计；如果在估计中要对参数θ_i做出带有某种可靠性的估计，就需要给出对应于这一可靠性或置信度的区间，即区间估计。由于区间估计是在点估计的基础上形成的，所以我们首先讨论点估计的方法。

§4.2.1　点估计方法

在参数点估计中，主要有代替原则和极大似然估计两种方法。此外，还有贝叶斯估计(见参考文献[8])、最小二乘估计等方法。最小二乘估计主要应用于线性统计模型中的回归分析和方差分析。将在第 6 章“相关与回归”中加以讨论。

1.代替原则。在代替原则中，又可以进一步分为频率代替原则和矩法两种方法。

(1)频率代替。如果一个观察有 K 种可能结果，当进行了 n 次观察时，发生各种结果的频数为 n_1，n_2，…，n_k，($n_1+n_2+\cdots+n_1=n$)。我们可以用观察的频率$\frac{n_i}{n}$ (i＝1，2…k)作为总体比率 P，的估计值，这一估计方法称为频率代替原则。该法主要用于处理离散数据，特别是变量为列名尺度和顺序尺度的数据。通常用的“成数”就是处理这两种尺度的频率方法。例如某饭店管理部门要了解该饭店的服务质量，随机抽取了 100 名旅客进行调查，其中有 18 人回答“非常满意”，有 52 人回答“比较满意”，有 23 人回答“不够满意”，有 7 人回答“很不满意”。由此样本调查结果可对该饭店综合服务质量做出推断和评价，即反映“比较满意”和“非常满意”的约占 70%，反映“不够满意”和“很不满意”的约占 30%。这一推断方法就是应用了频卒代替原则。调查结果表明，虽然多数旅客的反映是好的和比较好的，但该饭店的服务质量仍需改进。这种频率代替原则在经济管理中广泛应用。

(2)矩法。设 x1，…，x，是 N(μ,σ^2)的一个样本，因此 $E(x_i)=\mu,D(x_i)=\sigma^2,i=1,2,\cdots n$，$\mu$与$\sigma^2$是总体的一阶原点矩和二阶中心矩。由大数定律知道样本的矩依概卒收敛于总体的矩,因此用样本的一阶原点矩(即样本均值$\bar{x}$)就可以估计总体的均值μ。用样本的二阶原点矩(即$\frac{1}{n}\sum_{i=1}^{n}x_i^2$)可以去估计总体的二阶原点矩$\sigma^2+\mu$。于是可得到近似关系式

$$\bar{x}=\mu$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i^2=\mu^2+\sigma^2$$

由上式解出μ与σ^2的估计量

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases} \qquad (4.4)$$

如果是求标准差σ的估计值$\hat{\sigma}$，则由上式可得

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (4.5)$$

以上是正态分布的矩估计，对一般问题的矩估计可按下面步骤进行：

首先，确定要估计的总体参数。总体参数可以是一个参数，也可以是两个或几个参数形成的参数向量。如正态分布的两个参 数就可以形成参数向量$\theta=(\mu,\sigma^2)$；要估计的参数既可以是参 数本身，也可以是参数的函数，如要估计正态总体的变异系数$c = \frac{\sigma}{|\mu|}$，它就是μ和σ^2 的函数。因此要估计的参数总可以写成 $g(\theta)$的形式。

其次，要列出总体矩与参数的关系式，我们用μ_k表示总体的 k 阶原点矩，于是当$\mu_k<\infty$时，有

$$\mu_k = \int_{-\infty}^{\infty} x^k f(x;\theta)dx \qquad k = 1,2,\cdots s \qquad (4.6)$$

若θ中有 m 个参数，选 s＝m，有

$$\begin{aligned} \mu_k &= \int_{-\infty}^{\infty} x^k f(x;\theta_1,\cdots,\theta_m)dx \\ &= h(\theta_1,\cdots,\theta_m) k = 1,2,\cdots m \end{aligned} \qquad (4.7)$$

利用 m 等式(4.7)就可以将$\theta_1,\cdots,\theta_m$用$\mu_1,\cdots,\mu_m$表示出来，也就是解(4.7)给出的关于$\theta_1,\cdots,\theta_m$的方程，得到$\mu_1,\cdots,\mu_m$的函数θ；

$$\theta_j = f_j(\mu_1,\cdots,\mu_m) \qquad i = 1,2,\cdots,m \qquad (4.8)$$

再次，用样本的 k 阶原点矩$\frac{1}{n}\sum_{i=1}^{n} x_i^k$ 代人(4.8)式中的μ_k，得到θ_j 的估计值$\hat{\theta}_j$，再用$\hat{\theta}_j$代人要估计的参数 $g(\theta)=g(\theta_1,\cdots,\theta_m)$，就得到了 $g(\theta)$的估计值

$$g(\theta) = g(\hat{\theta}_1,\cdots,\hat{\theta}_m)$$

例 4.1 设 $x_1,\cdots,x_n$是抽自 $N(\mu,\sigma^2)$的随机样本，要估计参数之变异系数$c = \frac{\sigma}{|\mu|}$。

解：按照上面的步骤：

首先确定分布的参数是μ和σ^2，要估计的参数$c = \sqrt{\sigma^2}/|\mu|$

其次，总体的一、二阶原点矩为μ_1和μ_2，与参数μ和σ^2的关系式是

$$\mu_1=\mu,\mu_2=\sigma^2+\mu^2$$

因此解出

$$\mu=\mu_1,\sigma^2=\mu_2-\mu^2=\mu_2-\mu_1^2$$

再次，用样本原点矩代入上式得

$$\hat{\mu}=\bar{x},\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

将$\hat{\mu}$和$\hat{\sigma}^2$代入$c=\sqrt{\sigma^2}/|\mu|$得到

$$\hat{c}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\,/|\bar{x}|$$

2.极大似然估计(简记 MLE)。设 $x_1,\cdots,x_n$ 是抽自密度为 $f(x;\theta)$的一个样本，则 $x_1,\cdots,x_n$ 的联合密度为 $f(x_1;\theta)\cdot f(x_2;\theta)\cdots f(x_n;\theta)$。这时 $x_1,\cdots,x_n$ 被视为变量，θ被视为常量。如果把样本观察值 $x_1,\cdots,x_n$ 视为常量，将要估计的θ视为变量，$f(x_1;\theta)\cdot f(x_2;\theta)\cdots f(x_n;\theta)$被称为θ的似然函数，记为 $L(\theta;x_1,\cdots,x_n)$。即

$$L(\theta;x_1,\cdots,x_n)=\prod_{i=1}^{n}f(x_i;\theta) \tag{4.9}$$

由于在实际问题中，样本值 $x_1,\cdots,x_n$ 是已知的确定数值，而θ是未知的，因而将θ视为变量是合理的，似然函数反映了这种关系。极大似然估计法就是求似然函数的最大值点$\hat{\theta}$来作为对θ的估计量。

例 4.2 设有一批产品，其废品率为 p(0＜p＜1)。现从中随机抽出 100 个，其中有 10 个废品，试估计 p 的数值。

解：若正品用“0”表示，废品用“1”表示。则总体 X 的分布为样本观察值的联合分布为

$$P(X=x)=p^x(1-p)^{1-x} \qquad x=0,1$$

样本观察值的联合分布为

$$\begin{aligned}L(p;x_1,\cdots,x_{100})&=p^{x_1}(1-p)^{1-x}\cdot p^{x_2}(1-p)^{1-x_2}\cdot\cdots p^{x_{100}}(1-p)^{1-x_{100}}\\&=p^{\sum_{i=1}^{100}x_i}(1-p)^{100-\sum_{i=1}^{100}x_i}\\&=p^{10}(1-p)^{90}\end{aligned}$$

要得到 p 中使 $L(p;x_1,\cdots,X_n)$取最大值的$\hat{p}$，根据微积分的知识，要对似然函数求导，即

$$L(p;x_1,\cdots,x_n)=10\cdot p^9(1-p)^{90}-90\cdot p^{10}(1-p)^{89}=0$$

$$\hat{p}=\frac{10}{100}=0.1$$

这个结果与用代替原则计算的结果是一致的。

例 4.3 设 $x_1,\cdots,X_n$是抽自 $N(\mu,\sigma^2)$的随机样本，求μ,σ^2的极大似然估计。

解：样本观察值 $x_1,\cdots,X_n$ 的联合密度，即似然函数为

$$\begin{aligned}L(\mu,\sigma^2;x_1,\cdots,x_n)&=\prod_{i=1}^{n}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}\right)\\&=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}\end{aligned}$$

在一般情况下，由于 L 的最大值点与 lnL 的最大值点是相同的，可对 L 取对数

$\ln L(\mu,\sigma^2;x_1,\cdots,x_n)=-n\ln\sqrt{2\pi}-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$，由于似然函数中有两个未知参数，对其求偏导并令其为

$$\frac{\partial \ln L}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0$$

0

$$\frac{\partial \ln L}{\partial\sigma^2}=\frac{n}{2}\cdot\frac{1}{\sigma^2}+\frac{1}{2}\cdot\sum_{i=1}^{n}(x_i-\mu)^2\cdot\frac{1}{\sigma^4}=0$$

解此方程组得

$$\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$$

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

于是得到μ和σ^2的极大似然估计

$$\hat{\mu}=\bar{x},\ \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

这种方法与矩法结果相同。

以上两例用极大似然估计与用矩法得到的结果均相同，但不能由此得出两种方法结果均相同的一般性结论。在均匀分布中，两种方法的估计结果就不同(见参考文献［4］、［6］)。

§4.2.2　评价估计量好坏的标准

用什么标准来评价一个估计量的好坏呢？通常我们总是考虑估计量$\hat{\theta}$的均方误差 $E(\hat{\theta}-\theta)^2$，因为它反映了估计量$\hat{\theta}$和参数θ的差异之大小。将估计量$\hat{\theta}$的数学期望$\hat{\theta}$引人均方误差公式，就可以将均方误差进一步分解成两部分。换句话说，均方误差主要是受两个因素的影响，即

$$\begin{aligned}E(\hat{\theta}-\theta)^2&=E(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2\\&=E(\hat{\theta}-E(\hat{\theta}))^2+E(E(\hat{\theta})-\theta)^2\\&=D(\hat{\theta})+(E(\hat{\theta})-\theta)^2\end{aligned}\qquad(4.10)$$

式中：$D(\hat{\theta})$是估计量$\hat{\theta}$本身的方差，表示该估计量的精度；$(E\hat{\theta}-\theta)^2$是估计量期望值与其参数之间的偏差的平方，表示估计量估计的准确程度。这两个因素就是反映估计量好坏的两个主要标准，前者称为最小方差性或有效性，后者称为无偏性。

定义 4.1　若参数θ的估计量$\hat{\theta}$满足

$$E(\hat{\theta})=\theta$$

则称$\hat{\theta}$是θ的无偏估计(量)。

显然，如果一个估计量是无偏的，则(4.9)式中的第二项为 0。

定义 4.2　若$\hat{\theta}$，■都是参数θ的估计量，但有关系式

$$E(\hat{\theta}-\theta)^2\le E(■-\theta)^2$$

则称$\hat{\theta}$比■有效。

显然，如果$\hat{\theta}$对θ的均方误差比■对θ的均方误差小，以$\hat{\theta}$作估计量比■更有效、更理想。

现将这两条准则结合起来。在 θ 的任意无偏估计量中，如果能够找到一

个具有最小均方误差的$\hat{\theta}$，则称比$\hat{\theta}$是θ的一致最小方差无偏估计(量)，即UMVU估计量(Uniformly Minimum Variance Unbiased)。UMVU估计能使(4.10)式中第二项为 0，使第一项最小，就保证了该估计量均方误差的最小。UMVU估计在理论上是最好的，但实际求起来比较困难，一般要求估计量是完全充分统计量(完全性和充分性是统计量好坏的另外两个性质，见参考文献[7]、[9]。完全充分统计量求起来也很麻烦，一般借助于指数分布族。因为定理证明了指数分布族中的统计量是完全充分统计量，而一般的概率分布都可以写成指数分布族的形式。这样，通过指数分布族就可以找到完全充分统计量，并找到 UMVU 估计。另外一种寻找 UMVU 估计的方法是通过信息不等式找到无偏估计量方差的下界，从而得到 UMVU 估计(见参考文献[6]、[7]、[9]。

前面主要介绍了评价估计量的无偏性和有效性。此外，估计量还应考虑相合性(或一致性)、稳健性、完全性、充分性等，并由这些性质来评价和比较各种估计方法。统计学家研究了以 UMVU 估计为代表的无偏估计和极大似然估计(MLE)的特点，发现这两种估计法在指数分布族和大样本中无甚差别，具体选用哪种方法要根据计算的方便程度来决定。在小样本情况下，两种方法各具特点，统计学家也没有明确的结论。当然，这两种最优的估计方法也都存在一定的局限性。UMVU 估计理论上最优，但计算较困难。MLE 虽然比 UMVU 好求，但在具有多余参数的模型中却不理想。此外，这两种方法都不具有稳健性(见参考文献[31]。

§4.2.3 区间估计

前面介绍的参数点估计方法是用一个确定的值去估计未知的参数，看起来似乎很精确，实际上把握并不大。因为估计量是来自一个随机抽取的样本，总是带有随机性或偶然性，样本估计量$\hat{\theta}$刚好等于θ的可能性极小。但若说θ在$\hat{\theta}$附近，即估计θ在某一个小区间内，这就有把握多了。这种估计参数在某一区间内的方法就称为区间估计。先看一个例子，然后从中总结出一般的方法。

例 4.4 设 x_1，…，X_n 是测量物体长度θ的测量值，已知测量误差是各次独立的，都遵从 N(θ，σ^2)其中σ^2是已知的常数。以 99%的把握可以确定长度θ在什么范围之内？

解：由于$x_i=\theta+e_i$，i＝1, 2, …n；e_i 是 x_i 的测量误差。又由于 x_i …，X_n 与 e_1，…，e_n 的方差相同，所以 x_1，…，x_n，是独立同分布的，均服从 N(θ，σ^2)。由前面的知识知道θ的点估计量$\bar{x}\sim N(\theta,\frac{\sigma^2}{n})$，于是有

$$\begin{cases} P(\dfrac{|\bar{x}-\theta|}{\frac{\sigma}{\sqrt{n}}}\le 2)=0.9545 \\ P(\dfrac{|\bar{x}-\theta|}{\frac{\sigma}{\sqrt{n}}}\le 3)=0.9973 \end{cases} \quad (4.11)$$

也即以 0.9973 的概率断言不等式$|\bar{x}-\theta|\le 3\frac{\sigma}{\sqrt{n}}$成立，这个不等式就是

$\bar{x}-3\frac{\sigma}{\sqrt{n}}\le\theta\le\bar{x}+3\frac{\sigma}{\sqrt{n}}$或$\theta\in[\bar{x}-3\frac{\sigma}{\sqrt{n}},\bar{x}+3\frac{\sigma}{\sqrt{n}}]$的概率为 0.9973。这样我们就得到了θ的区间估计。

由(4.11)式可以看出以下两点：

(1)估计的区间越大，参数被包含在该区间的概率就越大，估计的可靠性也就越大。

(2)测量误差的方差越小(即测量值的精度越高)，相同的概率下，其区间越短。

可见区间估计总是与一定的概率保证相对应的。

定义 4.3　设 x_1，…，X_n 是抽自密度为了 f(x；θ)的一个样本，对给定的 0＜a＜1，如能求得统计量$\bar{\theta}$和$\bar{\theta}$使

$$P(\underline{\theta}\le\theta\le\bar{\theta})=1-a$$

则称［$\underline{\theta}$，$\bar{\theta}$］为θ的置信度为 1－a 的置信区间。$\underline{\theta}$和$\bar{\theta}$均是样本估计量的函数，被称为θ的置信下限和置信上限， 1－a 被称为置信度或置信概率，表示区间估计的可靠程度，通常取 90%，95%和 99%。a 称为显著性水平。

由例 4.4，我们可以得出区间估计的一般步骤：

(1)明确待估参数和置信度。在例 4.4 中参数是θ，置信度为 0.99。

(2)用参数的点估计法导出估计量的分布。例。4.4 中θ的估计量为$\bar{x}$该例已知$\bar{x}\sim N(\theta,\frac{\sigma^2}{n})$。

(3)利用估计量的分布给出置信区间。例 4.4 中利用$\bar{x}\sim N(\theta,\frac{\sigma^2}{n})$导出 P$(\bar{x}-\theta\le3\frac{\sigma^2}{n})$ ＝ 0.99 。 由 此 得 $\theta\in$［$\bar{x}$-$3\frac{\sigma}{\sqrt{n}},\bar{x}+3\frac{\sigma}{\sqrt{n}}]$的概率为$0.99,\underline{\theta}=\bar{x}-3\frac{\sigma}{\sqrt{n}},\bar{\theta}=\bar{x}+3\frac{\sigma}{\sqrt{n}}$。

下面我们按估计对象之不同分别讨论其区间估计的方法及应用。

1.总体均值”的区间估计。在对总体均值μ的估计中，最简单的情况是总体为正态分布，且方差σ^2已知(如例 4.4)，此时样本均值$\bar{x}$～N(μ，$\frac{\sigma^2}{n}$)，统计量

$$Z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0，1)\qquad(4.12)$$

当给定显著性水平 a 时，有

$$P(-Z_{\frac{a}{2}}\le\frac{|\bar{x}-\mu|}{\frac{\sigma}{\sqrt{n}}}\le Z_{\frac{a}{2}})=1-a$$

从而推出

$$P(\bar{x}-Z_{\frac{a}{2}}\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{x}+Z_{\frac{a}{2}}\frac{\sigma}{\sqrt{n}})=1-a$$

其应用可见例 4.4。

当总体分布未知或总体为非正态分布时，只要样本是大样本(一般认为 n ≥50 为大样本)，根据中心极限定理，样本均值 $\bar{x}$ 近似服从正态分布，并且用样本方差 S^2 来竹计总体方差 σ^2，因而也可以用 Z 统计量来进行区间估计。

例 4.5　某县 1987 年抽样调查了 400 户农民家庭的人均年化纤布消费量，得到均值为 3.3 米，标准差为 0.98 米。试以 95%的置信度估计该县 1987 年农民家庭年人均化纤布的消费水平。

解：由题意，虽然总体未知，但 n ＝400 属于大样本，可用(4.12)式的 Z 统计量进行区间估计，即

$$P(\bar{x}-Z_{\frac{a}{2}}\cdot\frac{S}{\sqrt{n}}\le\mu\le\bar{x}+Z_{\frac{a}{2}}\frac{S}{\sqrt{n}})=0.95$$

将各数值代入，得到

$$P(3.3-1.96\frac{0.98}{\sqrt{400}}\le\mu\le3.3+1.96\frac{0.98}{\sqrt{400}})=0.95$$

置信区间为

[3.204，3.396]

故我们有 95%的把握保证该县 1987 年农民家庭年人均化纤消费量为 3.204 米到 3.396 米。

如果在正态总体中，均值 μ 和方差 σ^2 均未知，要用小样本估计 μ 时，也要用样本方差 S^2 代替 σ^2，此时不能应用标准正态分布的 Z 统计量，而应该用 t 统计量，即

$$t=\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}\sim t_{n-1} \tag{4.13}$$

则均值μ的 1-a 的置信区间为

$$[\bar{x}-t\frac{a}{2}_{i\ n-1}\cdot\frac{S}{\sqrt{n}},\bar{x}+t\frac{a}{2}_{i\ n-1}\cdot\frac{S}{\sqrt{n}}]$$

例 4.6　铅的比重测量值是服从正悉分布的。现测量了 16 次，得到 $\bar{x}$=2.705.S=0.029。若置信度为 95%，试求铅的比重的置估区间。

解：根据已知 n ＝16，$\bar{x}$=2.705，S=0.029，且从附录查到 $t\frac{0.05}{2}$，15=2.131。从而有

$$\bar{x}-t_{\frac{a}{2}_{in-1}}\cdot\frac{S}{\sqrt{n}}=2.705-2.131\cdot\frac{0.029}{\sqrt{16}}=2.690$$

$$\bar{x}+t_{\frac{a}{2}_{n-1}}\cdot\frac{S}{\sqrt{n}}=2.705+2.131\cdot\frac{0.029}{\sqrt{16}}=2.720$$

铅比重的置信区间为[2.690，2.720]。

在经济管理中，经常要对总体比率进行推断。例如在产品质量的抽样检验中，要通过样本的不合格品率估计一批产品的不合格品率，并做出整批产品是否合格的判断。这类问题服从二项分布，其统计量 X 的均值和方差为

E(x)np,D(x)=np(1-p)

对于样本比率 $\bar{p}$ 有

$$E(\bar{p}) = E(\frac{x}{n}) = p$$

$$D(\bar{p}) = D(\frac{x}{n}) = \frac{1}{n^2} D(x) = \frac{1}{n} p(1-p)$$

根据德莫弗一拉普拉斯中心极限定理，在大样本时，二项分布可用正态分布来近似计算，即 $\bar{p} \sim N(p, \frac{1}{n} p(1-p))$

由于样本比率 $\bar{p}$ 是总体比率 P 的无偏估计，在 P 未知时，可用 $\bar{p}$ 代替。按照(4.12)式，有

$$Z = \frac{\bar{p} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \sim N(0，1) \tag{4.14}$$

于是可得到在 1－a 置信度下总体比率 P 的置信区间是

$$[\bar{p} - Z_{\frac{a}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + Z_{\frac{a}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}] \tag{4.15}$$

例 4.7 在一批产品中随机地抽取 100 件产品进行质量检验，其中有 4 件次品，试以 95%的置信概率估计整批产品的次品率。

解：记正品为“0”，次品为“1”，整批产品次品率为 p。由已知条件 $\bar{p} = \frac{x}{n} = \frac{4}{100}$，$n = 100$，查得 $Z_{\frac{a}{2}} = 1.96$。

代入(4.14)式有

$$\underline{\theta} = \bar{p} - Z_{\frac{a}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= 0.04 - 1.96 \cdot \sqrt{\frac{0.04 \times 0.96}{100}} = 0.002$$

$$\bar{\theta} = \bar{p} + Z_{\frac{a}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= 0.04 + 1.96 \sqrt{\frac{0.04 \times 0.96}{100}} = 0.078$$

所有 95%的把握断定整批产品的次品率在［0.002，0.078］内。

2.两总体均值之差的区间估计。在对两个总体均值之差进行区间估计时，根据总体和样本的不同，所用的统计量也是不同的。

当两个样本都是大样本时，则不论总体分布的情况如何，$\bar{x}_1$ 和 $\bar{x}_2$ 分别近似服从 $N(\mu_1，\frac{\sigma_1^2}{n_2})$ 和 $N(\mu_2，\frac{\sigma_2^2}{n_2})$。由样本的独立性可知 $\bar{x}_1$ 和 $\bar{x}_2$ 是独立的，且有

$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$

$$D(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

经标准化，可得

$$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \tag{4.16}$$

近似地服从标准正态分布 N(0，1)，并可用此估计量进行区间估计。若(4.16)式中 σ_1^2和σ_2^2 未知，则由大样本方差具有相合性这一性质，可用样本方差 S_1^2和S_2^2代替σ_1^2和σ_2^2，该统计量仍近似服从 N(0，1)分布。这时，统计量为

$$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \tag{4.17}$$

在给定 a 后，可由(4.16)式推出μ_1-μ_2的置信区间

$$[\bar{x}_1-\bar{x}_2-Z_{\frac{a}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}，\bar{x}_1-\bar{x}_2+Z_{\frac{a}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}] \tag{4.18}$$

例 4.8　从某市近郊区和远郊区各自独立地抽取了 50 户农民家庭，调查每户年末手存现金和存款余额。经计算得：$\bar{x}_1$＝650 元，$\bar{x}_2$＝480 元，S_1＝120 元，S_2＝106 元。试以 95%的概率估计该市近郊区与远郊区农民平均每户年末手存现金和存款余额之差的置信区间。

解：虽然两总体分布未知，但由于 n_1＝n_2＝50，属于大样本，故可用 2 统计量近似计算。直接代人(4.17)式有

$$[650-480-1.96\sqrt{\frac{120^2}{50}+\frac{106^2}{50}}，650-480-1.96\sqrt{\frac{120^2}{50}+\frac{106^2}{50}}]$$

经计算得到[125.62，214.38]，即该市近远郊农民平均每户年末手存现金和存款余额相差大约 125.62 元至 214.38 元其可靠性为 95%。

若两个样本是小样本，两个总体均为正态分布，总体方差σ_1^2和σ_2^2未知，但已知$\sigma_1^2=\sigma_2^2$，则可用样本方差来替代。这时，该统计量不再服从正态分布，而应采用 t 统计量，即

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{(\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2})\cdot\frac{(n_1+n_2)}{n_1n_2}}}\sim tn_1+n_2-2 \tag{4.19}$$

例 4.9　为了估计磷肥对某种农作物增产的作用，现选 20 块土壤条件大致相同的土地。其中 10 块不施磷肥.另外 10 块施磷肥，得到亩产量如表 4.1。

表 4.1　　某种农作物在 20 块地上的亩产量　　（单位：斤）

不施磷肥亩产	560	590	560	570	580	570	600	550	570	550
施磷肥亩产	620	570	650	600	630	580	570	600	600	580

已知不施磷肥的亩产量与施磷肥的亩产量都服从正态分布，且方差相同。试以置信度 0.95 对两种平均亩产之差做出区间估计。

解：已知 $n_1=n_2=10$，经计算

不施磷肥 $\bar{x}_1=570$，$(n_1-1)S_1^2=\sum\limits_{i=1}^{n1}(x_{1i}-\bar{x}_1)^2=2400$

施磷肥 $\bar{x}_2=600$，$(n_2-1)S_2^2=\sum\limits_{i=1}^{n2}(x_{2i}-\bar{x}_2)^2=6400$

查得 $t_{\frac{a}{2}}(18)=2.1009$，所以 μ_2-μ_1 的置信下限为

$$\bar{\theta}=\bar{x}_2-\bar{x}_1-t_{\frac{a}{2}}(n_1+n_2-2)\cdot\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1n_2}}$$

$$=600-570-2.1009\cdot\sqrt{\frac{8800}{18}\cdot\frac{20}{100}}$$

$$=30-20.8$$

$$=9.2$$

μ_2-μ_1 的置信上限为

$$\bar{\theta}=\bar{x}_2-\bar{x}_1+t_{\frac{a}{2}}(n_1+n_2-2)\cdot\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1n_2}}$$

$$=600-570+2.1009\cdot\sqrt{\frac{8800}{18}\cdot\frac{20}{100}}$$

$$=30+20.8$$

$$=50.8$$

故施磷肥与不施磷肥平均亩产之差的置信区间是[9.2，50.8]。在上例中若总体方差未知且 $\sigma_1^2\neq\sigma_2^2$，则不再简单服从 t 分布，讨论起来较复杂(见参考文献[3])。当两个总体分布未知时就要用非参数的方法，详见本书第 8 章。

3.正态总体方差 σ^2 的区间估计。设正态总体的分布是 $N(\mu,\sigma^2)$，其中 μ 和 σ^2 均未知。从总体中任意抽一样本，试对总体方差 σ^2 进行区间估计

由本章定理 4.3 可知

$$x^2=\frac{(n-1)S^2}{\sigma^2}\sim x^2(n-1)\text{或}x_{n-1}^2$$

给定置信度 1-a 时，$x^2(n-1)$ 的分布密度如图 4.4

图 4.4　$x^2(n-1)$ 的分布密度

在图 4.4 中，左右两侧面积都等于 $\frac{a}{2}$，即

$$P\left\{x^2 \geq x^2_{\frac{a}{2}}(n-1)\right\} = \frac{a}{2} \text{和} P\left\{x^2 \leq x^2_{1-\frac{a}{2}}(n-1)\right\} = \frac{a}{2}$$

则有

$$P\left\{x^2_{1-\frac{a}{2}}(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq x^2_{\frac{a}{2}}(n-1)\right\} = 1-a$$

可写成

$$P\left\{\frac{(n-1)S^2}{x^2_{\frac{a}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{x^2_{\frac{a}{2}}(n-1)}\right\} = 1-a \tag{4.20}$$

例 4.10 某灯具厂为提高产品质量,降低其主要产品 C 型灯泡使用寿命的不稳定程度，随机抽取了 40 个 C 型灯泡，测得其平均使用寿命为 4800 小时，样本标准差为 300 小时。试以 95%的可靠性估计该型灯泡寿命方差的置信区间。(已知其寿命服从正态分布)。

解：由题意知 n=40,S=300,查表

$$x^2_{\frac{a}{2}}(n-1) = x^2_{0.025}\ (39) = 58.120$$

$$x^2_{1-\frac{a}{2}}(n-1) = x^2_{0.975}\ (39) = 23.654$$

则由(4.19)式得σ^2的置信区间为

$$[\frac{(n-1)S^2}{x^2_{\frac{a}{2}}(n-1)}, \frac{(n-1)S^2}{x^2_{1-\frac{a}{2}}(n-1)}] = [\frac{39 \cdot 300^2}{58.120}, \frac{39 \cdot 300^2}{23.654}]$$

$$= [60392.29, 23.654]$$

4.两个正态总体方差比的区间估计

征两个正态总体 N(μ_1, σ_1^2)和 N(μ_2, σ_2^2)中，μ_1、μ_2、σ_1^2、σ_2^2 均未知。从两总体中独立地各取一个样本，其方差分别为S_1^2和S_2^2。我们要对总体方差之比$\frac{\sigma_1^2}{\sigma_2^2}$做出区间估计，由本章 定理 4.3 知$\frac{(n_1-1)S_1^2}{\sigma_1^2}$和$\frac{(n_2-1)S_2^2}{\sigma_2^2}$分别服从自由度为($n_1$-1)和$(n_2-1)$的 x^2 分布。注意到S_1^2和S_2^2相互独立，由 F 分布的定义知

$$F = \frac{\frac{(n_2-1)S_2^2}{\sigma_2^2} / (n_2-1)}{\frac{(n_1-1)S_1^2}{\sigma_1^2} / (n_1-1)} = \frac{S_2^2 / \sigma_2^2}{S_1^2 / \sigma_1^2} = \frac{\sigma_1^2 / \sigma_2^2}{S_1^2 / S_2^2} \tag{4.21}$$

服从自由度为(n_2-1，n_1-1)的 F 分布。

在给定置信度 1-a 时，图 4.5 左右两侧面积各为$\frac{a}{2}$，即

图 4.5 F(n_2-1，n_1-1)的分布密度

$$P\left\{F \geq F_{\frac{a}{2}}(n_2-1,n_1-1)\right\}=\frac{a}{2}$$

$$\text{和}P\left\{F \leq F_{1-\frac{a}{2}}(n_2-1,n_1-1)\right\}=\frac{a}{2}$$

可写成

$$P\left\{F_{1-\frac{a}{2}}(n_2-1,n_1-1) \leq F \leq F_{\frac{a}{2}}(n_2-1,n_1-1)\right\}=1-a$$

将(4.21)**式代入则有**

$$P\left\{F_{1-\frac{a}{2}}(n_2-1,n_1-1)\frac{S_1^2}{S_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq F_{\frac{a}{2}}(n_2-1,n_1-1)\frac{S_1^2}{S_2^2}\right\}$$
$$=1-a$$

例 4.11 两种不同型号的电阻分别服从 $N(\mu_1,\sigma_1^2)$和 $N(\mu_2,\sigma_2^2)$的总体,参数均未知。依次抽取容量为 25 和 15 的两独立样本，测得阻值样本方差依次为 6.38 和 5.15，求两总体方差σ_1^2/σ_2^2的 90%的置信区间。

解：由题已知 n_1=25，n_2＝15，故 n_1-1=24，n_2-1=14。

又$1-a=0.90$，$\frac{a}{2}=0.05$，$1-\frac{a}{2}=0.95$，**查表得**$F_{0.05}(24,14)=2.35$，

$$F_{0.95}(24,14)=\frac{1}{F_{0.05}(14,24)}=\frac{1}{2.13}$$

又算得

$$\frac{S_1^2}{S_2^2}=\frac{6.38}{5.15}=1.24$$

所以，σ_1^2/σ_2^2**的置信区间为**

$$[\frac{1.24}{2.35},(1.24)(2.13)]=[0.528,2.64]$$

最后，我们将主要的参数区间估计小结于表 4.2 中。

表 4.2 **参数区间估计小结表**

估计对象	总体(或样本的)	所用统计量及其分布	
均值μ	正态总体σ^2已知	$Z=\dfrac{\bar{x}-\mu}{\dfrac{\sigma}{\sqrt{n}}}\sim N(0,1)$	$[x-Z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}},\bar{x}+Z$
	大样本	$Z=\dfrac{\bar{x}-\mu}{\dfrac{\sigma}{\sqrt{n}}}\overset{近似}{\sim} N(0,1)$	$[\bar{x}-Z_{\frac{\alpha}{2}}\dfrac{S}{\sqrt{n}},\bar{x}+Z$
	正态总体σ^2未知	$t=\dfrac{\bar{x}-\mu}{\dfrac{S}{\sqrt{n}}}\sim t(n-1)$	$[\bar{x}-t_{\frac{\alpha}{2}}(n-1)\cdot$
均值之差 μ_1-μ_2	两正态总体方差σ_1^2、σ_2^2已知	$Z=\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$[\bar{x}_1-\bar{x}_2-Z_{\frac{\alpha}{2}}\sqrt{\dfrac{\sigma_1^2}{n_1}}$
	大样本	$Z=\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}\overset{近似}{\sim} N(0,1)$	$[\bar{x}_1-\bar{x}_2-Z_{\frac{\alpha}{2}}\sqrt{\dfrac{S_1^2}{n_1}}$

§4.3 假设检验

统计推断的参数估计问题，已在前一节进行了讨论，下面继续讨论假设检验问题。

§4.3.1 区间估计与假设检验的对偶性

在进行统计推断时，如果总体分布的形式是已知的，只是参数未知，则统计推断问题就可归结为推断总体参数的问题。例如在产品质量检验中，通过随机抽取的样本不合格品率推断出总体(即全部产品)的不合格品率，如果我们要以一定的概率把握程度估计总体不合格品率，这就是一个参数估计的问题。更准确地说，就是一个区间估计问题；如果要以一定的概率判断这整肚产品是否合格，这就是一个假设检验的问题。这两个问题对同一实例用的是同一个样本，同一个统计量，同一种分布，因而可由区间估计问题转换成假设检验问题，也可由假设检验问题转换成区间估计问题。这种互相转换形成了区间估计与假设检验的对偶性。

例 4.12 根据过去的测试知某种电子元件的使用寿命服从 $N(\mu_0, \sigma^2)$。经过该产品技术改进后，随机抽取了 n 个独立同分布的产品 x_1，x_2，…，xn 进行使用寿命测试，得到平均寿命(假设分布方差不会改变)。根据以上资料，可以进行两类统计推断：

如果要对技术改进后的产品使用寿命μ进行统计推断，则可在给定 1-a 置信度时对总体均值μ做出区间估计，即

$$P(\bar{x} - Z_{\frac{a}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\frac{a}{2}}\frac{\sigma}{\sqrt{n}}) = 1 - a$$

它表时区间 $[\bar{x} - Z_{\frac{a}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\frac{a}{2}}\frac{\sigma}{\sqrt{n}}]$ 覆盖μ的概率为 1-a。

如果要问改进后的平均使用寿命μ是否与改进前的μ_0有明显的差别，则可在给定显著性水平α时对总体参数μ进行假设检验。首先建立原假设 H_0：$\mu = \mu_0$ 和备择假设 H_1：$\mu \ne \mu_0$。原假设 $\mu = \mu_0$ 表明技术改进前后的平均使用寿命没有明显差别，备择假设 $\mu \ne \mu_0$ 表明二者间有了明显差别。由区间估计和假设检验的对偶性，可以通过置信区间构造一个水平为α的检验。即当且仅当μ_0落在区间之内，就接受原假设 H_0。若令

$$[\bar{x} - Z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}}]$$

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

则当且仅当 $$-Z_{\frac{a}{2}} \le Z \le Z_{\frac{a}{2}}$$

时接受 H_0；若 $|Z| \ge Z_{\frac{a}{2}}$ 时，则应拒绝 H_0。这种检验称为双边检验。从这个例子中看到，区间估计中的置信区间所对应的就是假设检验中的接受域。置信区间之外所对应的就是拒绝域。

§4.3.2 假设检验的步骤

第一步：建立假设。在一般情况下，总是将检验者的目的放在备择假设 H_0上，这样可以保证有充分的把握拒绝原假设 H_0，达到检验者的目的。如果原假设中只有一个点，即 $H_0：\theta=\theta_0$ 称为简单原假设；如果原假设中不只一个点，即 $H_0：\theta\leq\theta_0$ 或 $H_0：\theta\geq\theta_0$，称为复合原假设。

第二步：寻找检验 H_0 的统计量。这一步骤通过区间估计与假设检验的对偶性可从表 4.2 中查到。

第三步：选择显著性水平 α。α 表示检验中所犯的第一类错误，即将原假设成立误判为备择假设成立。在检验中，还存在第二类错误 β，即将备择假设成立误判为原假设成立。这两类错误的图示见图 4.6。

图 4.6　假设检验中犯两类错误的图示

在这两类错误中，理论工作者研究后认为第一类错误(即 α)重要，第二类错误(即 β)相对次要。原因是：

首先，在实际问题中，原假设是什么常常是明确的，而备择假设是什么则常常是模糊的。如例 4.12 中产品改进后，产品质量有了差别(即 $\mu\neq\mu_0$)就是模糊的。此时，我们不知道是 $\mu>\mu_0$ 还是 $\mu<\mu_0$。若 $\mu>\mu_0$，到底大多少，也不清楚。因而若错误地将 H_0 成立当成 H_1 成立，而 H_1 是什么并不清楚，这种错误是危险的，所以不犯第一类错误重要。

其次，原假设常常是简单假设，而备择假设常常是复合假设，备择假设的成立往往是不明确的，如 $H1：\theta>\theta_0$。由于 θ 可以任意逼近 θ_0，θ 比 θ_0 大到什么程度不明确，也说明不犯第一类错误重要。

由于两类错误的性质不同，将严重的错误放在 α 上，这同将检验者的目的放在备择假设上是一致的。当 α 确定后，检验之临界值或拒绝域就确定了。

第四步：用检验统计量值与临界值比较，若统计量值超过临界值，说明落人拒绝域中，就可以比较有把握地拒绝原假设，因为此时犯错误的概率是清楚的，只有 α 这么大。若统计量值落人接受域中，意味着没有把握拒绝原假设，尚需进一步检验。此时不能做出接受原假设的肯定结论，因为接受原假设时可能犯第二类错误，其大小为 β。

根据以上介绍的检验思想和检验步骤我们来讨论下面两类主要检验问题。

§4.3.3　总体均值的假设检验

1.正态总体均值是否为已知数 μ_0 的检验。在总体均值的检验中，若总体方差已知(如例 4.12)，可用 Z 统计最进行检验，若总体方差未知，就只能用样本方差来代替。此时只能用 t，统计量进行检验。

例 4.13　某食品厂用自动装罐机装罐头食品，每罐标准重量为 500 克。现随机抽取 10 罐来检查机器工作情况，这 10 罐的重量为(单位：克)

495，510，505，498，503，492，502，512，497，506。

假定重量服从正态分布,试问这段时间机器工作是否正常(给定显著性水平 5%)。

解：检验者检验的目的是试图发现机器布正常的情况。因而按检验步骤，建立假设

$H_0：\mu=500$　　$H_1：\mu\neq500$

由样本处得$\bar{x}=502$，$S=6.50$，查表$t_{\frac{\alpha}{2}}(10-1)=t_{\frac{0.05}{2}}(9)=2.2622$，检验统计量

$$t=\frac{|\bar{x}-500|}{S/\sqrt{n}}=\frac{|502-500|}{6.50/\sqrt{10}}=0.97$$

由于 t=0.97<$t_{\frac{\alpha}{2}}$(n-1)=2.26，因而不能拒绝原假设，故机器没发现异常。

与估计问题相同，如果检验统计量是来自大样本，则按照中心极限定理，不论其总体分布是否已知、是否正态，均可用 Z 统计量近似。

2.两正态总体均值是否相等的检验。当检验两正态总体均值是否相等时，所用统计量与估计问题中两正态总体均值之差的统计量相同。若两正态总体的方差已知，用 Z 统计量检验；若两正态总体的方差未知，虽可用样本方差代替，但必须假定两总体方差相等，此时可用 t 统计量进行检验。

例 4.14 现有两种不同热处理方法对金属材料做抗拉强度试验,得到试验数据如下：(单位：公斤/厘米^2)

甲法:31，34，29，26，32，35，38，34，30，29，32，31。

乙法:26，24，28，29，30，29，32，26，31，29，32，28。

设两总体均为正态总体，且方差相等。在给定显著性水平α＝0.05 时，问两种方法得到的抗拉强度有无显著差异。

解：将两种热处理加工的金属材料抗拉强度分别视为总体，建立假设 H_0：$\mu_1=\mu_2$　　H_1：$\mu_1\neq\mu_2$

通过计算有

$$\bar{x}_1=31.75 \qquad \bar{x}_2=28.67$$

$$(n_1-1)S_1^2=112.25 \qquad (n_2-1)S_2^2=66.64$$

$$n_1=n_2=12$$

又知 a=0.05，查表$t_{\frac{\alpha}{2}}$(22)=2.074

则有检验统计量

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{n_1+n_2}{n_1n_2}}}$$

$$=\frac{31.75-28.67}{\sqrt{\frac{112.25+66.64}{12+12-2}}\sqrt{\frac{24}{144}}}$$

$$=2.646$$

由于 t=2.646>$t_{\frac{\alpha}{2}}$(22)=2.074，统计量落入拒绝域中。故两种热处理法加工的金属材料抗拉强度有显著差异。

同单个总体均值检验相同,若两个样本均为大样本(即 $n_1\geq50$,$n_2\geq50$ 则可用 Z 统计量做两总体均值相等的检验。

§4.3.4　总体方差的假设检验

1.正态总体方差$\sigma^2=\sigma_0^2$的检验。

设 X～N(μ_1,σ^2)，建立假设

H_0：$\sigma 2=\sigma_0^2$　　　　H_1：$\sigma 2\neq\sigma_0^2$

应用 x^2 统计量

$$x^2 = (n-1)\frac{S^2}{\sigma^2}$$

当给定显著性水平α，可查表得到两个邻界值 $x^2_{\frac{a}{2}}(n-1)$ 和 $x^2_{1-\frac{a}{2}}(n-1)$。当

$$x^2_{\frac{a}{2}}(n-1) \le x^2 \le x^2_{1-\frac{a}{2}}(n-1)$$

统计量 X^2 落入接受域，即不能拒绝原假设；否则就拒绝原假设。

例 4.15　已知某纺纱车间纺出细纱的支数服从正态分布，其总体标准差为 1.2 支。从某日纺出的一批细纱中，随机地抽出 16 缕进行支数测量，得到样本标准差 S 为 2.1 支，问该日纱的均匀度与平时有无显著差别(取α=0.05)。

解：建立假设

H_0：$\sigma^2=(1.2)^2$　　　H_1：$\sigma^2\neq(1.2)^2$

计算统计量

$$x^2 = (n-1)\frac{S^2}{\sigma_0^2} = 15\times\frac{(2.1)^2}{(1.2)^2} = 45.94$$

查表得　$x^2_{0.975}(15) = 6.262,\ x^2_{0.025}(15) = 27.488$

由于 $x^2 = 45.94 > x^2_{0.025} = 27.488$，$x^2$ 检验量落入拒绝域中，说明这一天细纱均匀度与往日相比有显著差别。

2.两正态总体方差 $\sigma_1^2 = \sigma_2^2$ 的检验。

设 X_1 和 X_2 分别抽自正态总体 $N(\mu_1,\ \sigma_1^2)$ 和 $N(\mu_2,\ \sigma_2^2)$。现要检验假设

H_0：$\sigma_1^2=\sigma_2^2$　　H_1：$\sigma_1^2\neq\sigma_2^2$

由本章第 1 节可知

$$F = \frac{S_1^2 / S_1^2}{S_2^2 / \sigma_2^2} = \frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$$

当给定显著性水平α时，可查表得到两个邻界值 $F_{1-\frac{a}{2}}(n_1-1,n_2-1)$ 和 $F_{\frac{a}{2}}(n_1-1,n_2-1)$。当

$F_{1-\frac{a}{2}}(n_1-1,n_2-1) \le F \le F_{\frac{a}{2}}(n_1-1,n_2-1)$ 时，

统计量 F 落人接受域，即不能拒绝原假设；否则就拒绝原假设。

例4.16　现用两台机床加工同一种轴。从这两台机床的产品中分别随机地抽取若干根，测得直径(单位：毫米)如下：

机床甲：20.5119.8，19，7，20.4，20.1，20.0，19.0，19.9。

机床乙：19.7，20.8，20.5，19，8，19，4，20.6，19.2。

假定各台机床加工轴的直径分别服从正态总体，试比较这两台机床加工的精度间有无显著差异。(取α＝0.05)。

解：经计算得样本值

$$n_1 = 8, \bar{x}_1 = 19.93, S_1^2 = 0.216$$
$$n_2 = 7, \bar{x}_2 = 20.00, S_2^2 = 0.397$$

检验统计量

$$F = \frac{S_1^2}{S_2^2} = \frac{0.216}{0.397} = 0.544$$

表 4.3 **假设检验小结表**

对总体(或样本)的要求	假 设	检验统计量及其分布	H_0 的拒
正态总体 σ^2 已知	H_0：$\mu=\mu_0$ H_1：$\mu\neq\mu_0$	$Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0，1)$	$\lvert Z\rvert \geq Z$
	H_0：$\mu\geq\mu_0$ H_1：$\mu<\mu_0$		$Z \leq -Z$
	H_0：$\mu\leq\mu_0$ H_1：$\mu>\mu_0$		$Z \geq Z$
(单总体)大样本	H_0：$\mu=\mu_0$ H_1：$\mu\neq\mu_0$	$Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \overset{近似}{\sim} N(0，1)$	$\lvert Z\rvert \geq Z$
	H_0：$\mu\geq\mu_0$ H_1：$\mu<\mu_0$		$Z \leq -Z$
	H_0：$\mu\leq\mu_0$ H_1：$\mu>\mu_0$		$Z \geq Z$

续前表

对总体(或样本)的要求	假　设	检验统计量及其分布	H_0 的拒
正态总体 σ^2 未知	H_0：$\mu=\mu_0$ H_1：$\mu\neq\mu_0$	$t=\dfrac{\bar{x}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$	$\lvert t\rvert\geq t$
	H_0：$\mu\geq\mu_0$ H_1：$\mu<\mu_0$		t ≤ -t
	H_0：$\mu\leq\mu_0$ H_1：$\mu>\mu_0$		t ≥ t
两正态总体方差已知	H_0：$\mu_1=\mu_2$ H_1：$\mu_1\neq\mu_2$	$Z=\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\lvert Z\rvert\geq Z$
	H_0：$\mu_1\geq\mu_2$ H_1：$\mu_1<\mu_2$		Z ≤ -Z
	H_0：$\mu_1\leq\mu_2$ H_1：$\mu_1>\mu_2$		Z ≥ Z

续前表

对总体(或样本)的要求	假　设	检验统计量及其分布	H_0 的拒
(两总体)大样本	H_0：$\mu_1=\mu_2$ H_1：$\mu_1\neq\mu_2$	$Z=\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}\overset{\text{近似}}{\sim}N(0,1)$	$\lvert Z\rvert\geq Z$
	H_0：$\mu_1\geq\mu_2$ H_1：$\mu_1<\mu_2$		Z ≤ -Z
	H_0：$\mu_1\leq\mu_2$ H_1：$\mu_1>\mu_2$		Z ≥ Z
两正态总体方差未知但相等	H_0：$\mu_1=\mu_2$ H_1：$\mu_1\neq\mu_2$	$t=\dfrac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\right)\cdot\left(\dfrac{n_1+n_2}{n_1n_2}\right)}}\sim t(n_1+n_2-2)$	$\lvert t\rvert\geq t$
	H_0：$\mu_1\geq\mu_2$ H_1：$\mu_1<\mu_2$		t ≤ -t
	H_0：$\mu_1\leq\mu_2$ H_1：$\mu_1>\mu_2$		t ≥ t

续前表

对总体(或样本)的要求	假　设	检验统计量及其分布	H_0 的拒绝
正态总体	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$x^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim x^2(n-1)$	$x^2_{\frac{a}{2}}(n-1) \leq x^2 \leq x$
	$H_0: \sigma^2 \geq \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$		$x^2 \leq x^2_{1-\frac{a}{2}}(n$
	$H_0: \sigma^2 \leq \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$		$x^2 \geq x^2_a(n$
两正态总体	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$	$F_{1-\frac{a}{2}}(n_1-1, n_2-1)$ $F_{\frac{a}{2}}(n_1$
	$H_0: \sigma_1^2 \geq \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$		$F \leq F_a(n_1-1,$
	$H_0: \sigma_1^2 \leq \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$		$F \leq F_{1-a}(n_1-1$

由 a=0.05，查表得 $F_{0.025}(7,6)$=5.70，$F_{0.975}(7,6)$

$= \frac{1}{F_{0.025}(6,7)} = \frac{1}{5.12} = 0.195$。**由于 $F_{0.975}(7,6)$<$F_{0.025}(7,6)$，故结论是两正态总体方差无显著差异。**

与参数估计一样，最后我们将主要的假设检验法小结于第 109—112 页表 4.3。

习　题

1.设母体 X 的分布密度为

$$f(x)=\begin{cases}\theta x^{\theta-1}, & 0<x<1\\ 0\ , & \text{其他}\end{cases}$$

其中θ＞0。

(1)求θ的极大似然估计量；

(2)用矩法求θ的估计量。

2.设母体 X 的密度为

$$f(x)=\frac{1}{2\sigma}e^{-\frac{|x|}{\sigma}},\qquad -\infty<x<\infty$$

试求σ的极大似然估计；并问所得估计量是否是σ的无偏估计。

3.设母 X 调服从正态分布 N(μ，1)。(X_1，X_2)是从母体中抽取的一个样本。试验证下面三个估计量：

$$(1)\hat{\mu}_1=\frac{2}{3}x_1+\frac{1}{3}x_2$$

$$(2)\hat{\mu}_2=\frac{1}{4}x_1+\frac{3}{4}x_2$$

$$(3)\hat{\mu}_3=\frac{1}{2}x_1+\frac{1}{2}x_2$$

都是μ的无偏估计，并求出每个估计量的方差。问哪一个方差最小。

4.随机地从一批钉子中抽取 16 枚，测得其长度(单位：㎝)为

2.14	2.10	2.13	2.15
2.13	2.10	2.13	2.10
2.15	2.12	2.14	2.10
2.13	2.11	2.14	2.11

设钉长分布为正态的，试求总体均值μ的置信概率为 90%的置信区间，其中

(1)若己知α＝0.01(㎝)

(2)若σ未知

5.假定每次试验时，出现事件 A 的概率 P 相同但未知。如果在 60 次独立试验中，事件 A 出现 15 次，试求概率 P 的置信区间(α＝0.05)。

6.方差σ^2为己知的正态总体，问需抽取容量 n 为多大的样本。才能使总体均值μ的置信概率为 1-a 的置信区间长度不大于 L。

7.设统计量调服从总体为 N(μ，σ^2)的正态分布，$\bar{x}$ 和 S_n^2 分别是样本$\bar{x}_1$，$\bar{x}_2\cdots\bar{x}_n$，的均值和方差，又设$\bar{x}_{n+1}$～N(μ,σ^2)且与$\bar{x}_1$，$\bar{x}_2\cdots\bar{x}_n$，独立。试求统计量抽样分布。

8.对某农作物两个品种 A、B 计算了 8 个地区的亩产量，(单位：公斤)，结果如下：

品种 A:86，87，56，93，84，93，75，79

品种 B:80，79，58，91，77，82，76，66

假定这两个品种的亩产量分别服从正态分布，且方差相等。试求平均亩产量之差置信概率为 95%的置信区间。

9.在一批货物容量为 100 的样本中，经检验发现 5 个次品。试求这批货

物次品率的单侧置信上限(置信概率为 95%)。

10.从正态总体 N(μ，1)中抽取 100 个样本观察值，计算得 x＝5.32。

(1)试检验 H_0：μ＝5 是否成立(a＝0.01)；

(2)计算上述检验在μ＝4.8 时所犯第二类错误的概率。

11.某产品的次品率为 0.17。现对该产品进行新工艺试验，从中抽取 400 件检验，发现有次品 56 件。能否认为这项新工艺明显有效(a＝0.05)？

12.某切割机正常工作时，切割每段金属棒的平均长度为 0.5cm。今在某段时间内随机地抽取 15 段进行测量，结果为(单位：cm)

10.4	10.6	10.1	10.4	10.5
10.3	10.3	10.2	10.9	10.6
10.8	10.5	10.7	10.2	10.7

假定全属棒长度服从正态分布，此段时间内该机工作是否正常(a＝0.05)？

13.有甲、乙两合机床加工同样产品，从这两台机床加工的产品中随意地抽取若干件，测得产品直径(单位：mm)为

机床甲：20.5　19.8　19.7　20.4　20.1　20.0　19.0　19.9

机床乙：19.7　20.8　20.5　19.8　19.4　20.6　19.2

假定这两台机床加工的产品直径都服从正态分布，且方差相等。试比较这两台机床加工产品直径有无显著差异(α＝0.05)。

14.某电工器材厂生产一种保险丝。假定其保险丝熔化时间服从正态分布。在通常情况下，熔化时间的方差为 400。今从某天产品中抽取容量为 25 的样本，测得其熔化时间 $\bar{x}$＝62.24，S^2＝404.77，这天保险丝熔化时间的离散程度与通常有无显著差异(α＝0.01)？

15.检验 13 题中甲、乙两台机床加工精度有无显著差异(a＝0.05)。

16.测得两批电子器材电阻(单位：欧姆)的观察值为

A 批：0.140　0.138　0.143　0.142　0.144　0.137

B 批：0.135　0.140　0.142　0.136　0.138　0.140

设这两批器材的电阻分别服从分布 $N(\mu_1，\sigma_1^2)$和 $N(\mu_2，\sigma_2^2)$。

(1)检验 H_0：$\sigma_1^2=\sigma_2^2$ (a=0.05)

(2)检验 H_0：$\mu_1=\mu_2$(a=0.05)

第 5 章　方差分析

§5.1　方差分析的基本思想

在现实的生产和经营管理过程中，影响产品质量、数量或销量的因素往往很多，如农作物的收获量受作物品种、肥料种类及数量等的影响；不同地区、不同时期对某种产品的销量有影响等等。在众多因素中，有些因素影响大些，有些则小些。现实中常常需要分析哪几种因素对生产或销售起显著影响，并需知道起显著作用的因素在什么时候起最好的影响作用。方差分析是解决这些问题的一种有效方法。

§5.1.1　一个实际例子

例 5.1　某公司为了研究三种内容的广告宣传对某种大型机械销售量的影响，进行了调查统计。经广告广泛宣传后，按寄回的广告上的订购数计算，一年四个季度的销售量如表 5.1。

表 5.1　　机械销售量数据　　（单位：台）

广告类型	季度			
	一	二	三	四
A_1	163	176	170	185
A_2	184	198	179	190
A_3	206	191	218	224

表中 A_1 是强调运输方便性的广告，A_2 是强调节省燃料的经济性的广告，A_3 是强调噪音低的优良性的广告。试判断：新闻广告的类型对该种机械的销售量是否有显著影响？若影响显著，哪一种广告内容为好？

在这里，新闻广告是所要检验的因子(因素)，三种不同广告的内容可看作是三种水乎，因而这是一个单因子三水平的试验。若这三种广告内容的宣传对机械销售量的影响没有显著差异，那么从中采取一种既经济又方便的广告即可；若有显著差异，则希望从中选取一种较优的方案，以便对提高机械的销售量更为有利。

从例 5.1 可以看到以下两点：

(1)虽然用同一种广告，在同一年份不同季度的销售量也是不相等的(该机械本身无季节性)，其差别可以看作是由随机原因造成的。

(2)不同的新闻广告所引起的机械销售量也是不同的,这可能是由于广告的内容不同造成的，也可能是由于随机原因造成的。可以假设三种不同的广告为三个不同的总体。

要辨别随机误差和新闻广告这两个因素中哪个是造成销售量有差异的主要因素，这一问题可归结为判断三个总体是否具有相同分布的问题。在实际中，经常假定遇到的是具有正态分布的总体。同时，在进行方差分析时，除了所关心的因素(如例 5.1 中的新闻广告)外，其他条件总是尽可能使其保持一致，这样可以认为每个总体的方差是相同的。因而，推断几个总体是否具有相同分布的问题，就简化为检验几个具有相同方差的正态总体均值是否相等的问题。

§5.1.2　应用方塞分析的条件

方差分析是通过对误差的分析研究来判断多个正态总体均值是否相等的一种方法，它与前面所讲述的比较两组平均值的 t 检验法一样，有基本的假定条件：

(1)设检验的因子有 r 种水平，X_1，X_2，…，X_r，是 r 个相互独立的正态总体，分别服从于 $N(\mu_i, \sigma^2)$分布，i＝1,2，…，r。μ_i，第 i 个总体的均值，σ^2为方差。

(2)x_{ij}(i=1,2,…,r；j=1,2,…,n)是分别从总体 X_i 中抽得的简单随机样本。

这就是说，应用方差分析时要求符合下面两个条件：

(1)各个水平的测量观察数据,要能够被看作是从服从正态分布的总体中随机抽得的样本。

(2)各组测量观察数据，是从具有相同方差的相互独立的总体中抽得的。

通常，对第一个条件的要求并不苛刻，而第二个条件则要求必须满足。

§5.1.3　显著性检验

对于上面所提出的多个正态总体均值是否相等的问题，也就是要求检验假设

H_0：$\mu_1 = \mu_2 = \cdots = \mu_r$

如果说$n = \sum_{i=1}^{r} n_i$，那么有

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \tag{5.1}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij} = \frac{1}{n}\sum_{i=1}^{r} n_i \bar{x}_i \tag{5.2}$$

式中：$\bar{x}_i$是从第 i 个总体抽得的样本的平均值，通常称为组平均值；$\bar{x}$ 称为总平均值；n 是从 r 个总体中抽得的样本的总数目。

由(5.1)、(5.2)可以推得

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(\bar{x}_j - \bar{x}) = 0$$

由此可得

$$\begin{aligned} Q &= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 \\ &= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^{r} n_i(\bar{x}_i - \bar{x})^2 \\ &= Q_1 + Q_2 \end{aligned} \tag{5.3}$$

式中　$Q_1 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$

$$Q_2 = \sum_{i=1}^{r} n_i(\bar{x}_i - \bar{x})^2 \tag{5.4}$$

由(5.3)式可知，Q 是所有测量观察数据 x_{ij} 与总平均值的离差平方和，它是描述所获得的全部数据离散程度的一个指标，亦被称为总离差平方和。

(5.3)式是总离差平方和的分解式，它可以分解为两项，第一项 Q_1 是每个测量观察数据与其组平均值离差的平方和，反映了数据 x_{ij} 抽样误差的大小程度，也被称为误差平方和，它反映的是随机误差；第二项 Q_2 是各组平均值与总平均值离差的平方和，反映了各总体的样本平均值之间的差异程度，在一定程度上也反映了各总体均值 μ_1 之间的差异程度，它可以反映系统误差。因而通过 Q_2 可以反映原假设 H_0 是否成立。若 Q_2 显著地大于 Q_1，说明各 $\bar{x}_i$ 之间的差异大小显著地大于抽样误差，那么 H_0 可能不成立。这种比较方差大小以判断原假设 H_0 是否成立的方法就是方差分析法名称的由来。

那么，Q_2/Q_1 的比值大到什么程度，可以否定 H_0 呢？根据数理统计中关于正态母体子样均值和方差分布定理(柯赫伦分解定理)的推论可知，统计量

$$F = \frac{Q_2 / r-1}{Q_1 / n-r} \sim F(r-1, n-r) \tag{5.5}$$

可以作为判断 H_0 是否成立的检验统计量。

给定显著性水平 α，在 F 分布表中查找第一自由度 f_1=r-1、第二自由度 f_2＝n-r 相应的临界值 F_a(r-1，n-r)。若

$$F > F_a$$

则拒绝 $H_0(\mu_1=\mu_2=\cdots=\mu_r)$，也就是说有 1-α 的把握认为因　子对指标有显著影响，即 μ_I 之间的差异是显著的。若

$$F \le F_a$$

则不能拒绝 H_0，不能认为 μ_I 之间有显著差异，可认为因子对指标无显著影响。

§5.2　单因子方差分析

在方差分析中，通常称因子或因素的不同状态为水平。在例 5.1 中，考虑的是一个因子(新闻广告)，这个因子具有三个水平(A_1，A_2，A_3)。当涉及的因子只有一个时，这种方差分析称为单因子方差分析；涉及到的因子有两个时，称为双因子方差分析或二因子方差分析；涉及到的因于为两个或两个以上时，统称为多因子方差分析。例 5.1 是单因子方差分析的例子。

单因子方差分析的基本思想在上节已作过介绍。本节主要介绍它的具体作法。

§5.2.1　方差分析表

当所要分析研究的问题满足应用方差分析的条件时，我们可以建立原假设

H_0：$\mu_1=\mu_2=\cdots=\mu_r$

根据提供的数据，利用(5.4)式分别计算得到 Q_1 和 Q_2，然后计算 F 值，再与临界值 F_a 比较，对是否接受 H_0 作出判断。实际应用中，为方便起见，常用方差分析表代替计算过程，其表式如表 5.2。

例 5.1 的计算结果如表 5.3，此时 r=3，n＝12。给定显著性水平 α=0.05，在 F 分布表中查得 $F_{0.05}(2,9)=4.16$。因为 F= 10.93＞F_a=4.16，所以拒绝 H_0，即认为广告内容不同对销售量的影响是显著的。

表 5.2　　单因子方差分析表

方差来源	平 方 和	自由度	方　差	F　值
因子影响	$Q_2=\sum_{i=1}^{r} n_i(\bar{x}_i-\bar{x})^2$	r-1	S_2^2=Q2/r-1	S_2^2/S_1^2
误　　差	$Q_1=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$	n-r	S_1^2=Q1/n-r	
总　　和	$Q=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$	n-1	S^2=Q/n-1	

表 5.3　　方差分析表

方差来源	平 方 和	自由度	方　差	F　值
因子影响	2668.17	2	1334.09	10.93
误　　差	1098.50	9	122.06	
总　　和	3766.67	11		

§5.2.2　数学模型

上面的单因子方差分析方法是建立在一定的数学模型基础上的。若将考察的因素用字母 A 表示，则 A_1，A_2，…，A_i(i=1，2，…，r)表示试验所选择的水平，x_{ij} 表示在第 i 个水平下的第 j 个试验数据。如例 5.1 中有三种水平：A_1，A_2，A_3。每种水平有四个试验数据，它可以用表 5.4 来表示。

表 5.4 中 i 表示试验的水平；j 表示第 j 个试验数据；x_{ij} 表示第 j 个水

平下第 j 个数据的观察值。即使在同一水平下的观察值，如 x_{11}、x_{12}、x_{13}、x_{14} 之间也是有差异的，这是不可控制的因素影

表 5.4 试验方案及其结果

试验ζ（j）	水 平 (i)		
	A_1	A_2	A_3
1	x_{11}	x_{21}	x_{31}
2	x_{12}	x_{22}	x_{32}
3	x_{13}	x_{23}	x_{33}
4	x_{14}	x_{24}	x_{34}

响的结果，表现为随机误差，即有

$$x_{ij}=\mu_1+\varepsilon_{1j} \qquad (j=1,2,3,4)$$

其中μ_1表示第 A_1 水平下试验数据的平均值，ε_{1j} 表示第 j 次试验中一些随机因素的总效应，它服从正态分布。其他两种水平下的数据也可以有类似的分析。因而，一般地第 i 列即相应于 A_i 水平下的数据 x_{ij} 可表示为

$$x_{ij}=\mu_i+\varepsilon_{1j} \tag{5.6}$$

式中：$\varepsilon_{1j}\sim N(0,\sigma^2)$，它表示在因子的第 i 个水平下第 j 次试验结果数据 x_{ij} 相应的试验误差，μ_i 表示第 i 个水平下试验结果数据的理论均值。

(5.6)式数据的结构式即数学模型。为便于分析，还可以将其进一步分解为

$$\mu=\frac{1}{n}\sum_{i=1}^{r}n_i\mu_i$$

$$a_i=\mu_i-\mu \qquad (i=1,2,\cdots,r)$$

称 a_i 为第 i 个水平对试验结果(试验指标)的效应值，它反映因子第 i 个水平对试验指标的“纯”作用大小。于是，(5.6)式可以写成效应分解式

$$x_{ij}=\mu+a_i+\varepsilon_{ij} \tag{5.7}$$

$$i=1,2,\cdots,r；j=1,2,\cdots,n_i$$

对比(5.6)式与(5.7)式可知，μ_i 之间的差异性同 a_i 之间的差异性是等价的，并且

$$\sum_{i=1}^{r}n_ia_i=0$$

因此，检验：$\mu_1=\mu_2=\cdots=\mu_r=\mu$ 也可以归结为检验：$\alpha_1=\alpha_2=\cdots=\alpha_r=0$，即依据样本数据检验 $\alpha_i=0(i=1，2，\cdots，r)$是否成立。

§5.2.3　统计分析

因子的第 i 种效应 α_i，是指除去因子对试验指标的平均影响后，因子的第 i 种水平对试验指标的特殊影响。例 5.1 中试验指标为机械销售量，广告因子共有三种水平，即三种内容的广告。除去三种广告对机械销售量的平均影响以后，剩下的是每一种内容的广告对机械销售量的特殊影响，这就是相应水平对试验指标的特殊效应。从效应的角度看，方差分析要解决的问题是：检验因子各水平对试验指标的影响是否显著，若存在显著的差异，表明因于

各水平的效应不完全相同，可以从中选出效应最大的水平即最优水平作为实施方案；若无显著差异，表明因子各水平对试验指标的影响一样，差异是随帆引起的，这时，可以从中选择支出最少或费用最低的水平作为实施方案。

对于数学模型(5.7)式来说，统计分析的任务除了判断 H_0：$\alpha_1=\alpha_2=\cdots=\alpha_r$ 外，还可以求出 α_i，(i=1，2，…，r)的点估计和 μ_i 的区间估计。

因为 $\bar{x}_i$ 是 μ_i 的无偏估计，而

$$E(\bar{x}) = E(\frac{1}{n}\sum_{i=1}^{r} n_i \bar{x}_i)$$

$$= \frac{1}{n}\sum_{i=1}^{r} n_i \mu_i = \mu$$

所以 $\bar{x}$ 是 μ 的无偏估计。记 $\bar{a}_i = \bar{x}_i - \bar{x} \qquad (i = 1,2,\cdots,r)$

由于

$$E(\bar{a}_i) = E(\bar{x}_i - \bar{x}) = \mu_i - \mu$$

$$= a_i \qquad (i = 1,2,\cdots,r)$$

因此，$\bar{a}_i$ 是 a_i 的无偏估计。作为 a_i 的点估计有

$$\bar{a}_i = \bar{x}_i - \bar{x} \tag{5.8}$$

若原假设 H_0 被否定，则因子各水平的效应 a_i 之间有显著性差异，从而可以挑选出最优水平，并对最优水平下试验指标的观察值进行预测，同时作出区间估计。根据给定的显著性水平 a，第一自由度 f_1＝1，第二自由度 f_2=n-r，可查表得到临界值 $F_a(1，n-r)$，则 μ_i 的 1-a 置信区间为

$$[\bar{x}_i - \frac{S_e}{\sqrt{n_i}}\sqrt{F_a(1,n-r)}, \bar{x}_i + \frac{S_e}{\sqrt{n_i}}\sqrt{F_a(1,n-r)}] \tag{5.9}$$

式中：S_e 表示误差的标准差，即 $S_e^2 = S_1^2 = Q_1/n-r$。

例 5.2　对例 5.1 提供的资料进一步作统计分析。

解：由例 5.1 的方差分析表(见表 5.3)可知，广告内容对销售量的影响是显著的，即因子各水平对试验指标的影响有显著差异，原假设 H0 被否定。

由(5.8)式可得到各水平效应值 α_i 的点估计

$$\bar{a}_1 = \bar{x}_1 - \bar{x} = 173.5 - 190.33 = -16.83$$

$$\bar{a}_2 = \bar{x}_2 - \bar{x} = 187.75 - 190.33 = -2.58$$

$$\bar{a}_3 = \bar{x}_3 - \bar{x} = 2.9.75 - 190.33 = 19.42$$

计算结果表明，效应值 $\bar{a}_3$ 最大。这说明，广告 A_3 引起的销售量最多，水平 A_3 为最优水平。在今后的广告宣传中，应多宣传噪音低的优良性，这对增加销量有好处。同时，应进一步进行工艺改革以降低噪音。

在这个最优水平下试验对指标的观察值进行预测，即求出 μ_3 的区间估计。由(5.9)式可以计算得到 μ_3 的 95%置信区间为

$$\bar{x}_3 = 209.75$$

$$S_e^2 = \frac{1098.5}{9} = 122.06$$

$$F_{0.05}(1,9) = 5.12$$

$$\bar{x}_3 \pm \frac{S_e}{\sqrt{n}}\sqrt{F_{0.05}(1,9)} = 209.75 \pm 12.5$$

即 μ_3 的置信区间为[197.25222.25]。由此可知，在采用广告 A3 的情况下，机械的销售量将有 95%的把握达到 197 台到 223 台。

下面再通过一个例子来说明实际中方差分析的应用。

例 5.3　某工厂的锻造零件由 3 个地区制造，该零件的重要特性是强度。若其质量差异大，对生产将有影响。因此，当不同地区制造的零件强度有较大差异时，应从中选择一个最适合的。表 5.5 是从 3 个地区制造的零件中各随机抽取 4 个，由同一操作者在同一试验条件下，按随机的顺序进行强度试验(破坏性试验)所得的结果。试分析 3 个地区的零件强度是否有显著差异。若有不同，选择哪个地区的零件为宜。

表 5.5　　零件强度数据　　（单位：100kg）

地　区	序		号	
	1	2	3	4
A_1	115	116	98	83
A_2	103	107	118	116
A_3	73	89	85	97

解：可以把每个地区生产的零件看作是一个总体，而每个总 体零件的强度是服从正态分布、具有等方差且相互独立的随机变量，满足方差分析的基本假定。因此，零件强度差异性的检验可以采用方差分析方法。

根据表 5.5 提供的统计数据，可以计算得到方差分析表，如表 5.6。

表 5.6　　方差分析表

方差来源	平方和	自由度	方差	F 值
因子	1304	2	652	4.92
误差	1192	9	132.4	
总和	2496	11		

令显著性水平 a=0.05，查表得：$F_{0.05}(2,9)=4.26$ 因为

$$F=4.92>F_{0.05}(2,9)=4.26$$

所以，拒绝 H0。即有 95%的把握认为不同地区的零件强度有显著差异。

不同地区的零件强度平均值为

$$\bar{x}_1 = 103，\bar{x}_2 = 111，\bar{x}_3 = 86$$

因此，$\bar{x}_2$ 所代表的地区为最优地区，工厂应选择 A_2 地区生产的零件。

由于 $S_e^2 = 132.4$，n=4，$F_{0.05}(1,9)=5.12$，因此，A_2 地区零件强度均值的

95%的置信区间为

$$\bar{x}_2 \pm \frac{S_e}{\sqrt{n_2}}\sqrt{F_{0.05}(1,9)} = 111 \pm \sqrt{\frac{132.4}{\sqrt{4}}}\sqrt{5.12}$$

$$= 111 \pm 13.02$$

即　　　[97.98,124.02]

计算结果表明，A_2 地区生产的零件强度在 98 公斤到 124 公斤之间，可靠性为 95%。

§5.3　双因子方差分析

在许多实际问题中，往往不能只考虑单一因子的各个水平的影响，而必须同时考察几种因子的影响作用。如例 5.1 中，对机械销售量的影响因素，不仅有新闻广告，还有机械的销售价格。如果想通过试验，选取使机械销售量达到尽可能高的新闻广告形式和机械销售价格，同时研究两种因子对试验指标的影响，就是双因子方差分析问题，也称两因子方差分析。

由于存在两个因子的影响，就产生一个新问题：不同内容的广告和不同的销售价格对销售量的影响是否正好是它们每个因子分别对销售量影响的迭加？也就是说，是否会产生这样的情况，分别使销量达到最高的广告和销售价格(在不低于成本的情况下)结合起来，会使销售量增加的幅度超过二者分别增加幅度之和或低于二者分别影响之和。这种各个因子的不同水平的搭配所产生的新的影响在统计学上称为交互作用。各因子间是否存在交互作用是多因子方差分析新产生的问题。两个因子间交互作用的概念，反映了单因子方差分析与多因子方差分析的本质区别。多因子问题的分析比较复杂，但其解题的思想和基本方法类同，因而本章仅介绍双因子方差分析。

§5.3.1.　无交互作用的双因子方差分析

在双因子方差分析中，只有在每个因子的不同水平上进行重复试验时，才能分析出是否存在交互影响。为更好地了解双因子方差分析，首先考虑没有交互作用的情况。

1.数据结构。设因子 A 有 r 个不同水平

$$A_1,\ A_2,\ \cdots,\ A_r,$$

因子 B 有 S 个不同水平

B_1，B_2，…，B_s，在其他因子都加以控制的条件下，因子 A 的每一个水平和因子 B 的每一个水平组成一组试验条件，每种情况进行一次独立试验，共得到 rs 个试验结果 x_{ij}($i=1, 2, \cdots, r$；$j=1, 2, \cdots, s$)，其数据结构如表 5.7。

表 5.7　　双因子方差分析的数据结构

		B 因子				平均值 $\bar{x}_i$
		B_1	B_2	…	B_s	
A	A_1	x_{11}	x_{12}	…	x_{1s}	$\bar{x}_1$
因	A_2	x_{21}	x_{22}	…	x_{2s}	$\bar{x}_2$
子	M	M	M	…	M	M
	A_r	x_{r1}	x_{r2}	…	x_{rs}	$\bar{x}_r$
平均值 $\bar{x}_{\cdot j}$		$\bar{x}_{\cdot 1}$	$\bar{x}_{\cdot 2}$	…	$\bar{x}_{\cdot s}$	$\bar{x}$

表 5.7 中

$$\bar{x}_i = \frac{1}{S}\sum_{j=1}^{S} x_{ij} \qquad i = 1,2,\cdots,r$$

$$\bar{x}_j = \frac{1}{r}\sum_{i=1}^{r} x_{ij} \qquad j = 1,2,\cdots,s$$

$$\bar{x} = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{S} x_{ij}$$

2.数学模型。若因子 A 的第 i 种效应记作 a_i，因子 B 的第 j 种效应记作 β_j，试验误差记作 e_{ij}，那么

$$x_{ij} = \bar{x} + a_i + \beta_j + e_{ij} \quad (5.10)$$

$$i = 1,\ 2,\ \cdots,\ r; \qquad j = 1,\ 2,\ \cdots,\ s$$

式中：a_i，β_j 满足

$$\sum_{i=1}^{r} a_i = 0, \qquad \sum_{j=1}^{S}\beta_j = 0,$$

$$a_i = \bar{x}_{i.} - \bar{x} \qquad i = 1,\ 2,\ \cdots,\ r$$

$$\beta_j = \bar{x}_{.j} - \bar{x} \qquad j = 1,\ 2,\ \cdots,\ s$$

(5.10)式是双因子方差分析(无交互作用)的数学模型，并且假定

$$e_{ij} \sim N(0,\sigma^2)$$

3.方差分析。对于数学模型(5.10)式进行统计分析的原理与单因子方差分析的原理相同，判断因子 A 的影响是否显著等价于检验假设

H_{01}：$a_1=a_2=\cdots=a_r=0$

判断因子 B 的影响是否显著等价于检验假设

H_{02}：$\beta_1=\beta_2=\cdots=\beta_s=0$

为检验这些假设，将离差总平方和 Q 进行分解

$$\begin{aligned}
Q &= \sum_{i=1}^{r}\sum_{j=1}^{S}(x_{ij} - \bar{x})^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{S}[(\bar{x}_{i.} - \bar{x}) + (\bar{x}_{.j} - \bar{x}) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})]^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{S}(\bar{x}_{i.} - \bar{x})^2 + \sum_{i=1}^{r}\sum_{j=1}^{S}(\bar{x}_{.j} - \bar{x})^2 + \sum_{i=1}^{r}\sum_{j=1}^{S}(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 \\
&= S\sum_{i=1}^{r}(\bar{x}_{i.} - \bar{x})^2 + r\sum_{j=1}^{S}(\bar{x}_{.j} - \bar{x})^2 + \sum_{i=1}^{r}\sum_{j=1}^{S}(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 \\
&= Q_1 + Q_2 + Q_3 \qquad (5.11)
\end{aligned}$$

式中 $$Q_1 = S\sum_{i=1}^{r}(\bar{x}_{i.} - \bar{x})^2$$

$$Q_2 = r\sum_{j=1}^{S}(\bar{x}_{.j} - \bar{x})^2$$

$$Q_3 = \sum_{i=1}^{r}\sum_{j=1}^{S}(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2$$

与单因子方差分析的方法一样，可以采用统计量

$$\mathrm{F_A} = \frac{\mathrm{S}_1^2}{\mathrm{S}_3^2} = \frac{Q_1 / r-1}{Q_3 / (r-1)(s-1)} \tag{5.12}$$

和

$$F_B = \frac{\mathrm{S}_2^2}{\mathrm{S}_3^2} = \frac{Q_2 / r-1}{Q_3 / (r-1)(s-1)} \tag{5.13}$$

对 H_{01} 和 H_{02} 进行检验。

为方便起见，通常也用方差分析表，如下页表 5.8。

例 5.4 某公司分别在五个地区建立了某种小型车床的销售点 A_1、A_2、A_3、A_4、A_5，共记录了 5 个时期的销售量资料(见表 5.9)。试分析不同地区以及不同时期对销售量是有显否著影响。

解：设地区为因子 A，分为 5 个水平。第 i 个水平对销售量的特殊效应为 α_i(i＝1，2，3，4，5)，时期的因子为 B，也分为 5 个水平。第 j 个水平对销售量的特殊效应为 β_j(j=1，2，3，4，5)，则原假设为

H_{01}：$\alpha_1=\alpha_2=\alpha_3=\alpha_4=\alpha_5=0$

H_{02}：$\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=0$

方差分析的数据计算采用(5.11)式既不方便，又容易带来累积误差，因而实际计算时常采用下面的公式

表 5.8 双因子(无交互作用)方差分析表

方差来源	平 方 和	自 由 度
A 的影响	$Q_1 = S\sum_{i=1}^{r}(\bar{x}_{i.} - \bar{x})^2$	r-1
B 的影响	$Q_2 = r\sum_{j=1}^{S}(\bar{x}_{.j} - \bar{x})^2$	s-1
误 差	$Q_3 = \sum_{i=1}^{r}\sum_{j=1}^{S}(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2$	(r-1)(s-1)
总 和	$Q = \sum_{i=1}^{r}\sum_{j=1}^{S}(x_{ij} - \bar{x})^2$	rs-1

方差来源	方 差	F 值
A 的影响	$S_1^2 = Q_1 / r - 1$	$F_A = S_1^2 / S_3^2$
B 的影响	$S_2^2 = Q_2 / s - 1$	$F_B = S_2^2 / S_3^2$
误 差	$S_3^2 = Q_3 / (r-1)(s-1)$	
总 和		

表 5.9 销售量试验结果数据 (单位：百台)

		时期					平均值 $\bar{x}_{i}.$
		B_1	B_2	B_3	B_4	B_5	
地区	A_1	6.5	1.8	3.6	3.7	7.6	4.64
	A_2	14.2	7.1	10.8	8.9	12.6	10.72
	A_3	13.4	9.4	7.2	8.6	7.5	9.22
	A_4	2.4	1.5	1.7	2.3	2.8	2.14
	A_5	6.2	4.8	4.9	4.6	5.2	5.14
平均值 $\bar{x}_{.j}$		8.54	4.92	5.64	5.62	7.14	6.37

$$SS_T = \sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij} - \bar{x})^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} x_{ij}^2 - CT \tag{5.14}$$

$$SS_A = S\sum_{i=1}^{r}(\bar{x}_i . - \bar{x})^2 = \frac{1}{S}\sum_{i=1}^{r}(\sum_{j=1}^{S} x_{ij})^2 - CT \tag{5.15}$$

$$SS_B = r\sum_{j=1}^{S}(\bar{x}._j - \bar{x})^2 = \frac{1}{r}\sum_{j=1}^{S}(\sum_{i=1}^{r} x_{ij})^2 - CT \tag{5.16}$$

$$SS_e = \sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij} - \bar{x}_i . - \bar{x}._j + \bar{x})^2$$

$$= SS_T - SS_A - SS_B \tag{5.17}$$

式中 $CT = \frac{1}{rs}(\sum_{i=1}^{r}\sum_{j=1}^{s} x_{ij})^2$

根据表 5.9 中的数字，按照上面的公式可以计算得到

$$CT = \frac{1}{25}(159.3)^2 = 1015.06$$

$$SS_T = \sum_{i=1}^{r}\sum_{j=1}^{S} x_{ij}^2 - CT$$

$$= 1344.25 - 1015.06 = 329.19$$

$$SS_A = \frac{1}{5}\sum_{i=1}^{S}(\sum_{j=1}^{S} x_{ij})^2 - CT$$

$$= 1262.28 - 1015.06 = 247.22$$

$$SS_B = \frac{1}{5}\sum_{j=1}^{S}(\sum_{i=1}^{S} x_{ij})^2 - CT$$

$$= 1057.56 - 1015.06 = 42.50$$

$$SS_e = SS_T - SS_A - SS_B$$

$$= 329.16 - 247.22 - 42.50 = 39.47$$

方差分析表如表 5.10。

表 5.10　　　　　　　方差分析表

方差来源	平方和	自由度	方差	F 值
因子 A	$Q_1=SS_A=247.22$	4	$S_1^2=61.81$	$F_A=25.02$
因子 B	$Q_2=SS_B=42.50$	4	$S_2^2=10.63$	$F_B=4.03$
误　差	$Q_3=SS_e=39.47$	16	$S_3^2=2.47$	
总　和	$Q=SS_T=329.19$	24		

若给定显著性水平 α=0.05，查表得到临界值 $F_{0.05}(4，16)=3.01$；如果显著性水平 α=0.01，查表得到的临界值 $F_{0.01}(4，16)=4.77$。因为

$$F_A=25.02>F_{0.05}(4,16)=3.01$$

$$F_B=4.30>F_{0.05}(4,16)=3.01$$

所以拒绝 H_{01} 和 H_{02}，即根据现有的数据资料，可以有 95%的把握说不同地区对销售量的影响是极为显著的，不同时期对销售量也有影响。但当 α＝0.01 时

$$F_A=25.02>F_{0.01}(4,16)=4.77$$

$$F_B=4.30>F_{0.01}(4,16)=4.77$$

因而，根据现有数据资料，可以有 99%的把握断定不同地区对销售量的影响显著，但没有这样的把握来评价不同时期对销售量的影响，因为其 F 统计量小于临界值，不能拒绝 H_{02}。

例 5.5　为提高某种产品的合格率，考察原料用量和来源地对其是否有影响。原料来源地有三个：甲、乙、丙。原料用量有三种：现用量、增加 5%、增加 8%。每个水平组合备作一次试验，得到的数据如表 5.11。试分析原料用量及来源地对产品合格率的影响是否显著。

表 5.11　　　　　　　　　产品合格率数据　　　　　　　　(单位：%)

		原　料　用　量		
		B_1(现用量)	B_2(增加 5%)	B_3(增加 8%)
原料来源地	甲地 A_1	59	70	66
	乙地 A_2	63	74	70
	丙地 A_3	61	66	71

解：设原料来源地为因子 A，三个地区为因子的三个水平，第 i 个水平对合格率的特殊效应为 α_i(i=1，2，3)；原料用量为因子 B，三种用料量为因子的三个水平，第 j 个水平对合格率的特殊效应为 β_j(j=1,2,3)，则原假设为

H_{01}：$\alpha_1=\alpha_2=\alpha_3=0$　　H_{02}：$\beta_1=\beta_2=\beta_3=0$

利用(5.14)至(5.17)式可以计算得到

$$CT=\frac{1}{9}(600)^2=40000$$

$$SS_T=40200-40000=200$$

$$SS_A=\frac{1}{3}(38025+42849+39204)-40000=26$$

$$SS_B=\frac{1}{3}(33489+44100+42849)-40000=146$$

SS_e=200-26-146=28

方差分析表如表 5.12。

表 5.12　　　　　　　　　方差分析表

方差来源	平方和	自由度	方　差	F　值
因子 A	$Q_1=SS_A=26$	2	$S_1^2=Q_1/2=13$	$F_A=S_1^2/S_3^2=1.86$
因子 B	$Q_2=SS_B=146$	2	$S_2^2=Q_2/2=73$	$F_B=S_2^2/S_3^2=10.43$
误　差	$Q_3=SS_e=28$	4	$S_3^2=Q_3/4=7$	
总　和	$Q=SS_T=200$	8		

若显著性水平α=0.05,查表得到临界值

$$F_{0.05}(2,4)=6.94$$

因为

$$F_A=1.86<F_{0.05}(2,4)=6.94$$

$$F_B=10.43>F_{0.05}(2,4)=6.94$$

所以，不能拒绝 H_{01}，但能拒绝 H_{02}。即根据现有数据资料，有 95%的把握推断原料来源地对产品合格率影响不大，而原料用量对合格卒有显著影响。

可以计算出效应值β_2＝3.33，这是β_1、β_2、β_3中的最大者，说明原

料用量对产品合格率影响显著，B_2水平为最优水平。原料来源地对产品合格率影响不显著，因而可考虑以方便且运费省为准则选择水平。如果乙地最方便且运输距离短，那么最优条件为(A_2B_2)，即采用乙地原料并在原有用料量上增加 5%，这一方案为最佳。

由于只有因子 B 对合格率影响显著，故只需计算因子 B 各水平均值的区间。区间半径

$$\delta=\sqrt{\frac{S'_e\cdot F_a(1,f'_e)}{n_e\cdot f'_e}} \tag{5.18}$$

式中 $S_e^{'}$=误差平方和+不显著因子偏差平方和

$f_e^{'}$=误差的自由度+不显著因子的自由度

$$n_e=\frac{\text{试验总次数}}{1+\text{显著因子的自由度}}$$

1-α的置信区间为

$$\bar{x}_{\cdot j}\pm\delta \tag{5.19}$$

由前面的计算可知

$$S_e^{'}=SS_e+SS_A=54$$
$$f_e^{'}=4+2=6$$
$$n_e=\frac{9}{1+2}=3$$
$$F_{0.05}(1,6)=5.99$$

所以，区间半径

$$\delta=\sqrt{\frac{54\times5.99}{3\times6}}=4.24$$

因子 B 第二水平的区间估计为

70%±4.24%

即 [65.76%,74.24%]

有 95%的把握推断，原料增加 5%,产品合格率将在 65.76%到 74.24%之间。

§5.3.2 有交互作用的双因子方差分析

所谓交互作用，简单来说就是不同因子对试验所考察的指标的复合作用。在实际问题中常常遇到这种情况，因子 A 的作用随因子 B 的水平而变化。例如大豆亩产量，在不施氮肥的情况下，施加磷肥 4 斤，产量增加 50 斤；而在用氮肥 6 斤时，同样施加磷肥 4 斤，亩产就增加 l30 斤。这表明氮肥和磷肥对大豆亩产量有交互作用，或者说 A 和 B 两因素的综合效应不是二因素效应的简单相加。

当需要考虑两个因子的交互作用时，每组试验条件的试验至少做两次，否则无法将交互作用的平方和从误差平方和中分出来。

1.数据结构。设因子 A 取 r 个水平，因子 B 取 s 个水平，共有 rs 个水平组合 A_iB_j(i=1，2，…，r；j=1，2，…，s)。每个水平组合(试验条件)A_iB_j重复 l 次试验，每次试验结果的数据用 x_{ijk}(k=1，2，…，l)表示。其数据结构如表 5.13。

表 5.13　　　　　　有交互作用的双因子方差分析数据结构

		因子 B			
		B_1	B_2	⋯	B_s
因	A_1	$x_{111},\cdots,x_{111}$	$x_{121},\cdots,x_{121}$	⋯	$x_{1s1},\cdots,x_{1s1}$
子	A_2	$x_{211}\cdots,x_{211}$	$x_{221},\cdots,x_{221}$	⋯	$x_{2s1},\cdots,x_{2s1}$
A	M	M	M	⋯	M
	A_r	$x_{r11},\cdots x_{r11}$	$x_{r21},\cdots x_{r21}$	⋯	$x_{rs1},\cdots,x_{rs1}$

2.数学模型。若记$\bar{x}_{ij}$表示在因子 A 的第 i 个水平与因子 B 的第 j 个水平结合的水平组合 A_iB_j 下试验结果的均值，e_{ijk}表示在水平组合 A_iB_j 下第 k 次重复试验的试验误差，则对试验结果的数据 x_{ijk}有下面的关系

$$x_{ijk} = \bar{x}_{ij} + e_{ijk} \tag{5.20}$$

$$i=1，2，\cdots，r；j=1，2，\cdots，s；$$

$$k=1，2，\cdots，l$$

式中$\bar{x}_{ij} = \dfrac{1}{l}\sum_{k=1}^{l} x_{ijk}$

(5.20)式就是有交互作用的双因子方差分析的数学模型，对它有以下基本假定：总体 x_{ij} 服从 $N(\bar{x}_{ij}，\sigma^2)$，且 x_{ij} 之间相互独立(i=1，2，⋯，r；j=1，2，⋯，s)。x_{ijk}是总体 x_{ij} 的容量为 l 的样本。e_{ijk}是相互独立且服从 $N(0，\sigma_e^2)$的随机变量。

为了进一步分析因子 A、B 以及 A、B 间的交互作用对试验指标的影响，需对$\bar{x}_{ij}$进行分解。

令 $$\bar{x} = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\bar{x}_{ij} = \frac{1}{rsl}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}x_{ijk}$$

$$\bar{x}_{i\cdot} = \frac{1}{s}\sum_{j=1}^{s}\bar{x}_{ij}(i=1，2，\cdots，r)$$

$$\bar{x}_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r}\bar{x}_{ij}(j=1，2，\cdots，s)$$

表示各组水平组合对试验指标的平均贡献，它是度量各种效应的一个方便的起始点。取α_i为因子 A 第 i 水平 A_i 对试验指标的效应。

$$\alpha_i = \bar{x}_{i\cdot} - \bar{x} \qquad i=1,2,\cdots,r$$

i 为因子β第 j 个水平β_j对试验指标的效应。

$$\beta_j = \bar{x}_{\cdot j} - \bar{x} \qquad j=1,2,\cdots,s$$

显然

$$\sum_{i=1}^{r}\alpha_i=0 \qquad \sum_{j=1}^{s}\beta_j=0$$

$$\begin{aligned}\bar{x}_{ij}&=\bar{x}+(\bar{x}_{i}.-\bar{x})+(\bar{x}._{j}-\bar{x})+(\bar{x}_{ij}-\bar{x}_{i}.-\bar{x}._{j}+\bar{x})\\&=\bar{x}+\alpha_i+\beta_j+(\alpha\beta)_{ij}\end{aligned}$$

式中 $(\alpha\beta)_{ij}=\bar{x}_{ij}-\bar{x}_{i}.-\bar{x}._{j}+\bar{x}$

于是(5.20)**式可以写成效应分解式**

$$x_{ijk}=\bar{x}+\alpha_i+\beta_j+(\alpha\beta)_{ij}+e_{ijk} \tag{5.21}$$

在效应分解式(5.21)**式中**

$$\begin{aligned}(\alpha\beta)_{ij}&=(\bar{x}_{ij}-\bar{x})-(\bar{x}_{i}.-\bar{x})-(\bar{x}._{j}-\bar{x})\\&=(\bar{x}_{ij}-\bar{x})-\alpha_i-\beta_j\end{aligned}$$

式中，($\bar{x}_{ij}$-$\bar{x}$)表示反映水平组合 A_iB_j 对试验指标的总效应。总效应减去 A_i 的效应 α_i 和 B_j 的效应 β_j，所得的$(\alpha\beta)_{ij}$称为 A 与 B_j 对试验指标的交互效应。在多因子方差分析中，通常把因子 A 与因子 B 对试验指标的交互效应设想为某一因子的效应，这个因子记作 A×B。$(\alpha\beta)_{ij}$表示由于 A_i 与 B_j 的搭配所产生的一种 A_i 和 B_j 单独都不能起到的新作用。当试验指标数据越大越好时，$(\alpha\beta)_{ij}>0$ 称为有利的交互作用；$(\alpha\beta)_{ij}<0$ 时称为不利的交互作用。显然

$$\sum_{j=1}^{s}(\alpha\beta)_{ij}=\sum_{i=1}^{r}(\alpha\beta)_{ij}=0$$

3.方差分析。要判断因子 A、B 的影响以及交互作用，即因子 A×B 的影响是否显著，这分别等价于检验假设

H_{01}：$\alpha_1=\alpha_2=\cdots=\alpha_r=0$

H_{02}：$\beta_1=\beta_2=\cdots=\beta_s=0$

H_{03}：$(\alpha\beta)_{ij}=0$ 这一切 $i=1,2,\cdots,r$；$j=1,2,\cdots,s$

为了检验这些假设，要将离差的总平方和 Q 进行分解

$$\begin{aligned}Q&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}(x_{ijk}-\bar{x})^2\\&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}[(\bar{x}_{i}.-\bar{x})+(\bar{x}._{j}-\bar{x})+(\bar{x}_{ij}-\bar{x}_{i}.-\bar{x}._{j}+\bar{x})+(x_{ijk}-\bar{x}_{ij})]^2\\&=Q_1+Q_2+Q_3+Q_4\end{aligned} \tag{5.22}$$

式中 $Q_1=sl\sum_{i=1}^{r}(\bar{x}_{i}.-\bar{x})^2$

$Q_2=rl\sum_{j=1}^{s}(\bar{x}._{j}-\bar{x})^2$

$$Q_3 = l\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$$

$$Q_4 = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{j=1}^{l}(x_{ijk} - \bar{x}_{ij})^2$$

与单因子方差分析一样，可以采用统计量

$$F_A = \frac{S_1^2}{S_4^2} = \frac{Q_1 / r-1}{Q_4 / rs(l-1)} \tag{5.23}$$

$$F_B = \frac{S_2^2}{S_4^2} = \frac{Q_2 / s-1}{Q_4 / rs(l-1)} \tag{5.23}$$

$$F_{AB} = \frac{S_3^2}{S_4^2} = \frac{Q_3 / (r-1)(s-1)}{Q_4 / rs(l-1)} \tag{5.23}$$

对 H_{01}、H_{02}、H_{03} 进行检验。为方便起见，通常也采用方差分析表，如表 5.14。

表 5.14　　双因子(有交互作用)方差分析表

方差来源	平方和	自由度
因子 A	$Q_1 = sl\sum_{i=1}^{r}(\bar{x}_{i\cdot} - \bar{x})^2$	r-1
因子 B	$Q_2 = rl\sum_{j=1}^{s}(\bar{x}_{\cdot j} - \bar{x})^2$	s-1
因子 A × B	$Q_3 = l\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$	(r-1)(s-1)
误差	$Q_4 = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{j=1}^{l}(x_{ijk} - \bar{x}_{ij})^2$	rs(l-1)
总和	$Q = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}(x_{ijk} - \bar{x})^2$	rsl-1

方差来源	方　差	F　值
因子 A	$S_1^2 = Q_1 / r-1$	$F_A = S_1^2 / S_2^4$
因子 B	$S_2^2 = Q_2 / s-1$	$F_B = S_2^2 / S_4^2$
因子 A × B	$S_3^2 = Q_3 / (r-1)(s-1)$	$F_{AB} = S_3^2 / S_4^2$
误　差	$S_4^2 = Q_4 / rs(l-1)$	
总　和		

例 5.6　为了研究三种不同工艺方法和三种不同的灯丝配方对灯泡寿命的影响，对每种水平组合进行了两次试验，得到数据如表 5.15。试分析工艺方法和灯丝配方对灯泡寿命是否有显著影响。

表 5.15　　　　　　　　灯泡寿命数据　　　　　　　　(单位：百小时)

		因子 B		
		I	II	III
因子 A	甲	13.20　15.00	16.10　17.30	18.00　17.00
	乙	14.40　15.60	13.70　14.30	14.50　15.70
	丙	14.00　13.60	16.30　17.10	17.10　16.10

解：工艺和配方之间可能有交互作用，试验重复做了 2 次，因而可以进行有交互作用的方差分析。设工艺方法为因子 A，三种工艺方法为三个水平，第 i 个水平对灯泡寿命的特殊效应为 $\alpha_i(i=1, 2, 3)$；灯丝配方为因子 B，三种配方为二个水平，第 j 个水平对灯泡寿命的特殊效应为 $\beta_j(j=1, 2, 3)$；工艺和配方的交互作用为因子 A×B，$(\alpha\beta)_{ij}$ 为 A_i 与 B_j 对灯泡寿命的交互效应。原假设为

H_{01}：$\alpha_1=\alpha_2=\alpha_3=0$

H_{02}：$\beta_1=\beta_2=\beta_3=0$

H_{03}：$(\alpha\beta)_{ij}=0$　对 i=1,2,3；j=1,2,3

方差分析采用(5.22)式不便于进行计算，可以采用下式

$$SS_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}(x_{ijk}-\bar{x})^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}x_{ijk}^2-CT \tag{5.26}$$

$$SS_A=sl\sum_{i=1}^{r}(\bar{x}_i.-\bar{x})^2$$

$$=\frac{1}{sl}\sum_{i=1}^{r}(\sum_{j=1}^{s}\sum_{k=1}^{l}xi_{jk})^2-CT \tag{5.27}$$

$$SS_B=rl\sum_{j=1}^{s}(\bar{x}._j-\bar{x})^2$$

$$=\frac{1}{sl}\sum_{j=1}^{s}(\sum_{i=1}^{r}\sum_{k=1}^{l}xi_{jk})^2-CT \tag{5.28}$$

$$SS_e=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}(x_{ijk}-\bar{x}_{ij})^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}x_{ijk}^2-\sum_{i=1}^{r}\sum_{j=1}^{s}[\frac{1}{l}(\sum_{k=1}^{l}x_{ijk})^2] \tag{5.29}$$

$$SS_{A\times B}=l\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{ij}-\bar{x}_i.-\bar{x}._j+\bar{x})^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}[\frac{1}{l}(\sum_{k=1}^{l}x_{ijk})^2]-\frac{1}{sl}\sum_{i=1}^{r}(\sum_{j=1}^{s}\sum_{k=1}^{l}x_{ijk})^2-\frac{1}{rl}\sum_{j=1}^{s}(\sum_{i=1}^{r}\sum_{k=1}^{l}x_{ijk})^2+CT$$

$$=SS_l-SS_A-SS_B-SS_e \tag{5.30}$$

式中$CT=\frac{1}{rsl}(\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{l}x_{ijk})^2$

利用上面的几个式子可以计算得到

$$SS_T=4361.06-4324.5=36.56$$
$$SS_A=4330.74-4324.5=6.24$$
$$SS_B=4338.54-4324.5=14.04$$
$$SS_e=4361.06-4355.7=5.36$$
$$SS_{A\times B}=SS_T-SS_A-SS_B-SS_e=10.92$$

列出方差分析表如表 5.16。

若显著性水平 $\alpha=0.05$，查表可得到临界值 $F_{0.05}(2，9)=4.26$，$F_{0.05}(4,9)=3.63$。因为

$$F_A=5.24>F_{0.05}(2,9)=4.26$$
$$F_B=11.79>F_{0.05}(2,9)=4.26$$
$$F_{A\times B}=4.58>F_{0.05}(4,9)=3.63$$

所以，拒绝 H_{01}、H_{02}、H_{03}，即根据现有数据资料，有 95%的把握可以断定工艺方法和灯丝配方对灯泡寿命的影响是显著的，工艺和配方之间也存在交互作用。

表 5.16　　方差分析表

方差来源	平立和	自由度
因子 A	$Q_1=SS_A=6.24$	3-1=2
因子 B	$Q_2=SS_B=14.04$	3-1=2
因子 A×B	$Q_3=SS_{A\times B}=10.92$	(3-1)(3-1)=4
误　差	$Q_4=SS_e=5.36$	3×3×(2-1)=9
总　和	$Q=SS_l=36.56$	3×3×2-1=17
方差来源	**方　差**	**F　值**
因子 A	$S_1^2=3.12$	$F_A=5.24$
因子 B	$S_2^2=7.02$	$F_B=11.79$
因子 A×B	$S_3^2=2.73$	$F_{A\times B}=4.58$
误　差	$S_4^2=0.5956$	
总　和		

从表 5.15 提供的试验数据看，在工艺甲下采用配方III生产，灯泡寿命最长。

习 题

1.某电视台既想多赚广告费，又不愿播放广告太多而失去很多观众。为分析广告节目播放时间对观众人数是否有显著影响，该合就本台播放的 4 个时段广告对观众进行了调查。计算机对调查所得数据进行计算得到表 1。试说明广告节目播放时间对观众人数是否有显著影响。

表 1

方差来源	平方和	自由度
因子	367	3
误差	294	16
总和	661	19

2.利用表 2 资料分析华北地区不同行业的全民所有制职工工资水平是否有显著差异。

表 2 （单位：元/人·年）

地区	工业	建筑业	交通运输、邮电通讯业	教育、文化艺术和广播电视事业
北京	1327	1737	1529	1244
天津	1161	1631	1503	1162
河北	1113	1428	1245	1026
山西	1201	1493	1297	1108
内蒙	1246	1499	1332	1199

3.利用表 3 资料分析研究不同地区和不同时间对农民家庭人均纯收入的影响。

表 3 （单位：元）

时 间	地 区				
	北京	天津	河北	山西	内蒙古
1980 年	290.46	277.92	175.78	155.78	181.32
1981 年	350.67	297.77	204.41	179.53	225.14
1982 年	432.63	326.12	235.73	227.18	273.03
1983 年	519.48	411.69	298.07	275.78	294.20
1984 年	664.16	504.64	345.00	305.50	336.12
1985 年	775.08	564.55	385.23	358.32	360.41

第 6 章　相关与回归

§6.1　相关分析

§6.1.1　变量之间的关系

在社会经济现象中，变量之间的关系大致可分为两种：函数关系和统计关系。

1.函数关系。变量之间依一定的函数形式形成的一一对应关系称为函数关系。若两个变量分别记作 Y 和 X，则当 Y 与 X 之间存在函数关系时，Y 值一旦被指定，Y 值就是唯一确定的。

例 6.1　试分析出租汽车费用和行驶里程的关系。

解：设 Y 为出租汽车费用，X 为行驶里程，Y 与 X 之间的关系可以表述为

$$Y=2.80+1.20X$$

式中，2.80 是出租汽车的固定服务费用；1.20 是每公里的出租费用。图 6.1 反映了变量 Y 与 X 之间的这种函数关系，它表现为平面坐标上的一条直线。当出租汽车的行驶里程为 2 公里时，所应支付的费用为 5.20 元；行驶里程为 3 公里时，费用为 6.40 元。在平面坐标图上表现为两个点(2，5.20)和(3，6.40)，这两个点在一条直线上，若行驶里程为 5 公里，支付费用则为 8.80 元，也在同一直线上。

2.统计关系。两个变量之间存在某种依存关系，但变量 Y 并不是由变量 X 唯一确定的，它们之间没有严格的一一对应关系。两个变量间的这种关系就是统计关系，亦称相关关系。例如，同

图 6.1　出租汽车费用与行驶里程相关图

样收入的家庭，用于食品的消费支出往往并不相同。这是因为对家庭食品费用的影响，不仅仅有家庭收入的多少，还有家庭人口、生活习惯等因素。因此，家庭食品费用支出与家庭收入之间不是函数关系，而是统计关系。

例 6.2　试分析企业年设备能力和年全员劳动生产率的关系。

解：设 Y 为全员劳动生产率，X 为年设备能力，根据 14 个企业的统计数字，可以绘制成图 6.2。当 X=4.0 千瓦/人，时，Y＝9.1 千元/人，在平面坐标图上就是点(4.0，9.1)。从图 6.2 可以看出，企业的劳动生产率与设备能力之间有着某种类似直线的关系，但在所有点间连出直线时(如图 6.2)，这些点散落在直线周围，并未全部落在直线上。劳动生产率并不由设备能力唯一确定。这是一个线性统计关系的例子。

例 6.3　试分析企业电视机单机成本与产量的关系。

解：设 Y 为电视机单机成本，X 为电视机月产量，根据某电视机厂 16 个月的统计数字，绘制成图 6.3。由图 6.3 可知，单机成本与产量之间有着某种关系。一般来说，产量愈多单机成本就愈低，但不象例 6.2 是线性关系，而是一种非线性关系。

图 6.2　劳动生产率与设备能力的相关图

图 6.3　单机成本与产量的相关图

两个变量之间若存在线性关系称为线性相关，存在非线性关系称为曲线

相关。很多非线性关系通过适当的变量变换，常可转换为线性关系形式(见§6.2.5)。因此，线性统计模型的应用十分广泛。

§6.1.2　简单相关系数

变量 Y 与变量 X 之间线性相关的程度可以用简单相关系数 r 度量。它的计算公式为

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2} \cdot \sqrt{\Sigma(y-\bar{y})^2}} \tag{6.1}$$

式中，$\bar{x}$和$\bar{y}$分别是变量 Y 和 Xn 个样本数据的平均值。由(6.1)式很容易推导出简单相关系数 r 的另一个计算公式

$$r = \frac{\Sigma xy - n\bar{x}\bar{y}}{\sqrt{\Sigma x^2 - n\bar{x}^2} \cdot \sqrt{\Sigma y^2 - n\bar{y}^2}} \tag{6.2}$$

简单相关系数 r 的取值为若大的 X 值趋于同大的 Y 值相关联，小的 X 值同小的 Y 值相关联，表明变量 X 与 Y 之间是正相关，这时，0＜r＜1，其散点图如图 6.4(a)；若大的 X 值趋于同小的 Y 值相关联，小的 X 值同大的 Y 值相关联，表明变量 X 与 Y 之间是负相关，这时，-1<r<0，其散点图如图 6.4(b)；若 Y 的取值几乎完全不受调值的影响，表明变量 X 与 Y 之间没有相关关系，这时，r=0，其散点图如图 6.4(c)。

图 6.4　变量 Y 与 X 的相关图

若 r=+1，变量 Y 与 X 是完全正相关；r=-1，变量 Y 与 X 是完全负相关。在这种情况下，两个变量之间的关系是函数关系。因而，函数关系可以看作是统计关系的一种特例。另外，变量 Y 与 X 的简单相关系数 r=o 时，也只能说明 Y 与 X 之间不存在线性统计关系，但可能存在非线性统计关系。

§6.1.3　相关系数的显著性检验

一般说来，简单相关系数 r 是对变量 Y 与 X 之间线性关系密切程度的一个测量。r 的值越接近于±1，反映 Y 与 X 之间的线性关系越紧密。但是线性相关系数通常是根据样本数据计算得到，因而带有一定的随机性。样本容量越小其随机性就越大，如当变量 Y 与 X 各只有 2 个样本数据时，其相关系数总是 1，但这并不等于两个变量完全相关。因此，简单相关系数也有一个显著性检验问题，即通过样本相关系数 r 对总体相关系数 ρ 作出推断。由于相关系数 r 的分布密度函数比较复杂，实际应用中需要对 r 作变换。对总体相关系数 ρ=0 的假设检验，就是对总体是否相关作出推断，令

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{6.3}$$

则统计量 t 服从 t(n-2)分布。将 r 更换成 t 后，可以用 t 检验方法检验 ρ=0 是否成立。

根据给定的显著性水平 α (通常 α＝0.05)和自由度 n-2，查 t 分布表[①]，找到相应的临界值 tα/2。若

$$|t| \geq t_{a/2}$$

表明 r 在统计上是显著的，它的值可以作为两个变量之间是否存在线性关系

[①] 相本系数的显著性检验是双尾检验。

的证据。若

$$|t|<t\alpha/2$$

表示 r 在统计上是不显著的，它的值不能作为两个变量之间是否存在线性关系的证据。

例 6.4　对例 6.2 中企业年设备能力与劳动生产率的相关系数进行显著性检验。

解：计算得到的相关系数 r=0.98，n＝14，故 r 的 t 值为

$$t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}=\frac{0.98\sqrt{14-2}}{\sqrt{1-0.98^2}}=17.2792$$

选择显著性水平 α=0.05，根据自由度 df＝14-2＝12，查找 t 分布表得：$t_{0.05/2}$=2.2792

结论：由于 t＝17.2792＞t0.05/2＝2.1788，表明简单相关系数 r 是显著的，它能够用来说明企业年设备能力和劳动生产率之间存在线性关系。

§6.2　一元线性回归

如果两个变量之间存在线性关系，其中一个是自变量，另一个是因变量，利用它们的样本数据，建立起表述它们之间关系的数学模型，对模型进行各种统计检验，并利用这一模型进行预测和控制，就是一元线性回归。

§6.2.1　一元线性回归的数学模型

如果两个变量之间存在相关关系，并且一个变量的变化会引起另一个变量按某一线性关系变化，则两个变量间的关系可以用一元线性回归模型表述。

1.总体回归模型。

例 6.2 中企业劳动生产率若可解释为主要由年设备能力的变化所致，则劳动生产率作为因变量(被解释变量)Y，年设备能力作为自变量(解释变量)X，二者之间有下面的数学结构式：

$$Y=\alpha+\beta X+\varepsilon \qquad (6.4)$$

式中，α、β是总体回归参数，ε是随机项，表示除年设备能力以外其它影响劳动生产率变化的因素。(6.4)式被称为总体回归模型。若能搜集到变量 Y、X 的 n 个数据，则对每一组 Yi、Xi 都存在上面的关系，即有

$$Y_i=\alpha+\beta X_i+\varepsilon_i \qquad i=1,2,\cdots,n \qquad (6.5)$$

为了便于进行统计推断,变量 Y 与 X 所建立的(6.5)式一元线性回归模型有以下几点基本假设：

(1)变量 Y 与 X 之间存在着“真实的”线性关系；

(2)变量 X 为非随机变量；

(3)随机项ε_i(i=1,2,…,n)～N(0,σ^2),且相互之间独立，即

$$\mathrm{Cov}(\varepsilon_i,\varepsilon_j)=\begin{cases}\sigma^2 & i=j\\ 0 & i\neq j\end{cases}$$

$$i=1,2,\cdots,n,\quad j=1,2,\cdots,n$$

2.样本回归模型。(6.4)式中的参数α、β是理论上总体的值，实际上是不知道的，通常只能利用变量 Y、X 的样本数据，依据某种准则，得到他们的估计值 a、b。与(6.4)式相对应的样本回归模型为

$$Y=a+bX+e \qquad (6.6)$$

式中，a、b 是总体回归参数α、β的估计值，称为回归系数，α是回归直线的截距，b 是回归直线的斜率，二者结合反映变量 Y 与 X 之间的一种变动关系；e 是误差项。与(6.5)式相对应的样本回归模型是

$$Y_i=a+bX_i+e_i \qquad (6.7)$$

§6.2.2　参数的最小二乘估计

(6.6)式中回归系数 a、b 的估计通常采用最小二乘准则，即

$$e_i=Y_i-(a+bX_i)$$

$$\Sigma e_i^2=\min$$

由此得到的 a、b 称为α、β的最小二乘估计，它们所确定的直线$\hat{Y}_i$＝a＋bX$_i$(i=1，2，…，n)称为 Y 对 X 的线性回归方程。它表述了变量 Y 同给定的变量 X 的诸值间的平均变动关系。Y 对 X 的回归系数 b 表示 X 每变动一个单位所引起的 Y 的平均变动。

由于Σe_i^2是 a、b 的二次函数且非负，所以存在最小值，根据微积分学的极值定理，可求解得到

$$\begin{cases} b = \dfrac{n\Sigma xy - \Sigma x \cdot \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} \\ a = \bar{y} - b\bar{x} \end{cases} \tag{6.8}$$

§6.2.3　回归模型的检验

利用变量 Y 与 X 的 n 对样本数据得到 a、b 的最小二乘估计值，由此建立的回归模型 $Y_i=a+bX_i+e_i$ 是否满足回归模型的基本假设，需要进行检验。

1.回归系数的显著性检验。由于回归系数 b 是总体回归系数β的一个估计值，通过 b 可以检验β的值与 0 是否存在显著差异。若β=0，意味着总体回归模型中无 X 项，即 Y 并不随 X 变动而变动，因此 Y 与 X 之间不存在线性关系，违背了一元线性回归模型的基本假设；若β≠0，说明变量 Y 与 X 之间存在线性关系，符合原来的假设。回归系数的显著性检验，通常是通过回归系数 b 的 t 值检验，验证变量 Y 与 X 间是否有“真实的”线性关系。其检验步骤如下：

建立原假设　H_0：β=0

对立假设　H_1：β≠0

计算回归系数 b 的 t 值

$$t_b = \frac{b}{S_b} \tag{6.9}$$

式中：S_b是回归系数 b 的标准差，可以利用公式

$S_b = \sqrt{S_y^2 / \Sigma(x_i - \bar{x})^2}$求得；$S_y$是回归的估计标准误差，由公式

$S_y^2 = \Sigma(y_i - \hat{y}_i)^2 / n-2 = \Sigma e_i^2 / n-2$计算得到(见公式(6.15))。

根据给定的显著性水平α(通常α＝0.05)和自由度 df=n-2，查 t 分布表，得到临界值 ta/2[①]，若

$$|t_b|>ta/2$$

拒绝 H_0。若 a=0.05，则表明回归系数β＝0 的可能性小于 5%，可得出β≠0 的结论。若

$$|t_b|\le ta/2$$

不能拒绝 H_0，回归系数β有可能为零，说明β与 0 的差异是不显著的。因而只有接受 H_0 的结论。

2.回归方程的显著性检验。变量 Y 与 X 之间的线性关系还可以通过回归方程这一整体来进行检验判定。回归方程的显著性检验是以方差分析方法为基础，对 Y 与 X 间存在“真实”线性关系的检验，也被称为回归方程的 F 检验。其检验步骤如下。

建立原假设 H_0：β= 0

对立假设 H_1：β≠0

计算回归方程的 F 统计量

[①] 回归系数的 t 检验是双尾检验.

$$F=\frac{\Sigma(\hat{y}-\bar{y})^2/1}{\Sigma(y-\hat{y})^2/n-2} \tag{6.10}$$

式中：$\bar{y}$是因变量 Yn 个样本数据的平均值。

根据给定的显著性水平α(通常 a=0.05)和两个自由度：$df_1=1$，$df_2=n-2$，查 F 分布表，得到临界值 F_a。若

$$F>F_a$$

拒绝 H_0，意味着线性回归模型中的一次项是必不可少的。这时，称回归方程的回归效果是显著的。若

$$F\le F_a$$

不能拒绝 H_0，意味着线性回归方程的回归效果不显著。

3.D.w 检验。回归模型的假设条件中有 $Cov(\varepsilon_i,\varepsilon_i)=0(i\neq j)$，即随机项是独立的。这一假设可以通过回归模型的误差(残差)序列(动态数据进行回归时，残差序列也是按时间顺序的)是否相互间独立来进行检验。若误差序列诸项间相互独立，则序列各项之间没有相关关系；若序列诸项之间有相关关系，误差序列不满足线性回归模型的基本假设，所建立的回归模型就不能表述变量 Y 与 X 间的真实变动关系。D.w 检验是残差序列的相关检验，其检验步骤如下。

计算残差序列的 d 统计量(D.w 值)

$$d=\sum_{i=2}^{n}(e_i-e_{i-1})^2 \tag{6.11}$$

根据给定的显著性水平α(α=0.05 或α=0.01)，自变量个数 K 和样本数据个数 n,查 D.w 表，得到下限值 d_l 和上限值 d_u，若

$$d_u<d<4-d_u$$

则残差序列无自相关，各项间相互独立，若

$$0<d<d_1$$

或 $$4-d_1<d<4$$

或表明残差序列存在正自相关或负自相关，各项之间不是相互独立的，违背线性回归模型的基本假设，D.w 检验未通过；若或则尚无法断定是否存在自相关。

4.拟合程度的测定。变量 Y 的各个观测值点聚在回归直线$\hat{Y}$＝a＋bX 周围的紧密程度，称作回归直线对样本数据点的拟合程度，通常用可决系数(亦称测定系数)r^2来表示。

变量 Y 的各个观测值 y_1 与其均值$\bar{y}$的离差$(y_i-\bar{y})$可以分解为两部分：回归值$\hat{y}_i$与均值$\bar{y}$的离差$(\hat{y}_i-\bar{y})$，这是由回归所能解释的部分；观测值 y_i 与回归值$\hat{y}_i$的离差$(y_i-\hat{y}_i)$亦即残差 e_i，这是不能由回归加以解释的部分。对某一项观测值 y_i 有

$$y_i-\bar{y}=(\hat{y}_i-\bar{y})+(\bar{y}-\hat{y}_i) \tag{6.12}$$

这一关系如图 6.5 所示。由(6.12)式可以得到

图 6.5　变差分解示意图

$$\Sigma(y_i-\bar{y})^2=\Sigma(\hat{y}_i-\bar{y})^2+\Sigma(\hat{y}_i-\hat{y}_i)^2+2\Sigma(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)$$

能够证明 $\Sigma(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=0$ **因此有**

$$\Sigma(y_i-\bar{y})^2=\Sigma(\hat{y}_i-\bar{y})^2+\Sigma(\hat{y}_i-\hat{y}_i)^2 \tag{6.13}$$

以 SST 表示总的变差平方和 $\Sigma(y_i-\bar{y})^2$，SSR 表示回归平方和 $\Sigma(\hat{y}_i-\bar{y})^2$，SSE 表示剩余平方和(残差平方和)$\Sigma(y_i-\hat{y}_i)^2$。显然，变量 Y 的各个观测值点与回归直线越靠近，SSR 在 SST 中所占的比重越大，因而定义

$$r^2=\frac{SSR}{SST}=\frac{\Sigma(\hat{y}_i-\bar{y})^2}{\Sigma(y_i-\bar{y})^2} \tag{6.14}$$

它可以用来测定回归直线对各观测值点的拟合程度。若全部观测值点 y_i (i=1，2，…，n)都落在回归直线上，则剩余平方和 SSE=0，r^2=1；若 X 完全无助于解释 Y 的变差，则回归平方和 SSR=0，r^2=0。显然，r^2 越接近于 1，用 X 的变化解释 Y 的变差的部分就越多,表明回归直线与各观测值点越接近，回归线的拟合程度越高。可决系数 r2 在[0，1]上取值。

回归直线对样本数据点拟合程度的另一测度是线性相关系数 r。在一元线性回归中，线性相关系数 r 实际上是可决系数 r^2 的平方根，即

$$r=\pm\sqrt{r^2}$$

r 的正负号与回归系数 b 的正负号相同[①]。|r|越接近于 1，表明回归直线对样本数据点的拟合程度越高。

5.估计标准误差。可决系数 r^2 和线性相关系数 r 描述了回归直线对样本数据点的拟合程度，但没有表示出变量 Y 诸观测值 i 与回归直线 $\hat{y}_i$＝a＋bX，的绝对离差数额。定义

$$S_y^2=\frac{\Sigma(y_i-\hat{y}_i)^2}{n-2}=\frac{\Sigma e_i^2}{n-2}$$

为最小二乘残差 ei 的方差，

$$S_y=\sqrt{\frac{\Sigma(y_i-\hat{y}_i)^2}{n-2}}=\sqrt{\frac{\Sigma e_i^2}{n-2}} \tag{6.15}$$

为变量 Y 对 X 的最小二乘回归的估计标准误差，简称估计标准误。S_y^2 和 S_y 可以作为诸 y 值与回归直线变差的测度。S_y 的计量单位与变量 Y 的单位相同。显然，S_y 越小表明误差越小。

例 6.5　根据表 6.1 提供的统计数字，建立某地区居民对某产品的需求量与居民收入的回归方程。

解：今需求量为因变量 Y，居民收人为自变量 X，根据表 6.1 的数据，绘制两个变量之间的散点图，如图 6.6。从图 6.6 中可以

[①] **由(6.8)式中 b 的计算公式可以得到**

表 6.1　　某地区居民对某产品的需求量和居民收入

年份	需求量（千件）	居民收入（万元）	年分	需求量（千件）	居民收入（万元）
1972	116.5	255.7	1980	146.8	330.0
1973	120.8	263.3	1981	149.6	340.2
1974	124.4	275.4	1982	153.0	350.7
1975	125.5	278.3	1983	158.2	367.3
1976	131.7	296.7	1984	163.2	381.3
1977	136.2	309.3	1985	170.5	406.5
1978	138.7	315.8	1986	178.2	430.8
1979	140.2	318.8	1987	185.9	451.5

图 6.6　某产品需求量与居民收入散点网

看出，二者之间呈线性关系，采用最小二乘法建立一无线性回归方程

$$\hat{Y}=27.9123+0.3524X$$

对模型进行各种检验

(1)t 检验。

因为
$$S_y^2=\frac{\Sigma e_i^2}{n-2}=\frac{17.57}{14}=1.255$$

$$S_b=\sqrt{S_y^2/\Sigma(x-\bar{x})^2}=0.0055$$

所以　$t_b=b/S_b=64.2069$

根据显著水平α=0.05，自由度 df=14，查 t 分布表得 $t_{0.05/2}=2.1448$。由于

$$t_b=64.2069>t_{0.05/2}=2.1448$$

表明回归系数 b 是显著的，居民收入与居民对某产品的需求量之间存在线性关系。

(2)F 检验。

因为　$SSR=\Sigma(\hat{y}_i-\bar{y})^2=5173.78$

$SSE=\Sigma(y_i-\hat{y}_i)^2=17.57$

所以　$F=\dfrac{5173.78/1}{17.57/14}=4122.53$

根据显著水平α=0.05，$df_1=1$，$df_2=14$，查 F 分布表得到 $F_{0.05}(1,14)=4.60$。由于

$$F=4122.53>F_{0.05}(1,14)=4.60$$

表明回归方程的 F 检验通过，回归方程的回归效果显著。

可以验证：$F=t^2$，$F_\alpha＝t\alpha/2$，所以在一元线性回归中，F 检验与 t 检验的结果相同。

(3)D.W 检验。用公式(6.11)可以计算得到残差序列的 d 统计量：d＝0.68。根据显著性水平α=0.05，自变量个数 1，样本数据个数 16，查 D.W 表得到：$d_l=1.10$,$d_u=1.37$。由于

$$0<d=0.68<d_l=1.10$$

D.W 检验未通过，残差序列存在正自相关。

(4)拟合程度测定。

$$r^2=\frac{SSR}{SST}=\frac{5173.78}{5191.35}=0.9966$$

r^2值很接近于 1，表明回归直线对样本数据点的拟合程度很高。

(5)估计标准误差。由前面的计算知 S_y＝1.2550，表明回归标准误差较小。

结论：该回归模型虽然其他检验都已通过，担 D.W 检验没有通过，表明残差序列存在正自相关，前面应用最小二乘法的结果中S_y^2可能低估了真正的σ^2，因而 t 检验、F 检验不再有效。这种情况下，应分析查找残差序列自相关的原因，并采取相应措施加以解决，以建立更适宜的回归方程。

一般来说，导致残差序列自相关的原因有以下三种。(1)选择的数学模型不合适，变量间不是线性关系而建立了线性模型。这种情况应进一步选择合适的模型；(2)模型中包含的自变量数目不合适，或是遗漏了某些重要的影响因素，或是包含了不必要的影响因素；(3)序列中包含有很强的趋势分量。通常可以采用迭代法或差分法进行补救。由于经济的时间序列常常有自相关现象，因此在计量经济学教科书中都有较详细的讨论。

§6.2.4　利用回归方程进行预测

利用变量 Y 与 X 的 n 对样本数据建立的回归方程

$$\hat{Y}=a+bX \tag{6.16}$$

如果通过了上述的各种检验，即可用来预测。所谓预测问题，就是在确定自变量的某一个 X_0 值时求相应的因变量 Y 的估计值，其中又可以分为点预测和区间预测。

1.点预测。将自变量的预测值 X_0 代人回归模型(6.16)式所得到的因变量 Y 的值$\hat{Y}_0$，作为与 X_0 相对应的 Y_0 的预测，就是点预测。可以证明$\hat{Y}_0$是无偏预测。

2.区间预测。对于与 X_0 相对应的值 Y_0，$\hat{Y}_0$＝a＋bX 可以作为 $Y_0=\alpha+\beta X+\varepsilon$ 的一个点估计值。但不同的样本会得到不同的 a 、 b ，因此，$\hat{Y}_0$与 Y_0 之间总存在一定的抽样误差。在回归模型的假设条件下，可以证明($\hat{Y}_0$-Y_0)～N[0,σ^2(1+$\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}$)]，因此，$Y_0$的概率为 1-a 的预测区间为

$$\hat{Y}_0\pm t_{a/2}\cdot\sigma\cdot\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\Sigma(x_i-\bar{x})^2}}$$

当 x_0 取值在$\bar{x}$附近，n 又比较大时，可以近似地认为($\hat{Y}_0$-Y_0)～N(0，S_y^2)。因而 Y_0 的概率为 1- a 的预测区间为

$$\hat{Y}_0\pm t_{a/2}\cdot S_y$$

实际应用时，常常采用这一区间作为因变量 Y 相对应于自变量 X_0 的回归预测区间。当 a =0.05 时，Y_0 的 95%预测区间为若 a =0.01，则 Y_0 的 99%预测区间为

$$\hat{Y}_0\pm 2S_y$$

若 a=0.01，则 Y_0 的 99%预测区间为

$$\hat{Y}_0 \pm 3S_y$$

以上是对利用回归方程进行预测方法的介绍。下面，以一个完整的例子加以说明。

例 6.6　根据表 6.2 提供的数据，分析预测 1981 年到 1985 年我国国民收入以 4.5%的速度递增，钢材消费量将达到的水平。

解：令钢材消费量为因变量 Y，国民收人为自变量 X，根据表 6.2 的数据绘制散点图，如图 6.7。从图可以看出，变量 Y 与 X 之间呈线性关系。利用最小二乘法建立一元线性回归方程

$$\hat{Y} = -460.5282 + 0.9840X$$

表 6.2　　我国钢材消费量与国民收入*

年份	钢材消费量 (万吨)	国民收入 (亿元)	年份	钢材消费量 (万吨)	国民收入 (亿元)
1964	698	1097	1973	1765	2286
1965	872	1284	1974	1762	2311
1966	988	1502	1975	1960	2003
1967	807	1394	1976	1902	2435
1968	738	1303	1977	2013	2625
1969	1025	1555	1978	2446	2948
1970	1316	1917	1979	2736	3155
1971	1539	2051	1980	2825	3372
1972	1561	2111			

* 国民收入按 1975 年价格计算

图 6.7　钢材消费量与国民收入散点图

对模型进行各种检验。

(1)t 检验。经计算得 S_y^2 18348.9240，由此可得

S_b＝0.0497，因而 $t_b=b/S_b=0.9840/0.0497=19.78$ 根据显著性水平 a＝0.05，df＝15，查 t 分布表得 $t_{0.05/2}$＝2.1310。由于

$$t_b=19.78>F_{0.05/2}=2.1310$$

所以，回归系数 b 的 t 检验通过，表明回归系数 b 是显著的，变量国民收入能够解释变量钢材消费量的变化。

(2)F 检验。计算回归方程的 f 值为 F＝391.27。根据显著性水平 a=0.05，$df_1=1,df_2=15$，查 F 分布表得 $F_{0.05}(1，15)$＝4.54。由于

$$F=391>F_{0.05}(1,15)=4.54$$

所以 F 检验通过，表明回归方程的回归效果显著。

(3)D.W 检验。计算 d 统计量：d＝2.0326。根据 a＝0。05，自变量个数 1，样本数据个数 17，查 D.W 表得 d_u＝1.38。由于 $d_u=1.38<d=2.0326<4-d_u=2.62$，所以 D，W 检验通过，表明残差序列无自相关。

(4)其他检验。r^2＝0.9631，接近千 1，表明回归直线对样本数据点的拟合程度很高。

S_y＝135.4582，虽然并不接近于 0，但其与因变量样本数据平均值 $\bar{y}$ ＝

1585.4710 的比值为：$S_y/\bar{y}$=0.0854，小于 10%，可以认为比较小。

上述分析说明，回归方程通过了各种统计检验，可以用来表述钢材消费量和国民收入之间的回归关系。

预测：

若 1981 年至 1985 年国民收入以 4.5%的速度递增，利用回归方程可以得到相应的钢材消费量点预测值及 95%的预测区间，如表 6.3。

利用变量 Y 与 X 的样本数据建立的回归方程能否用于预测，除了需要通过各种统计检验外，还应考虑变量之间结构关系的稳定性。若变量间的结构关系比较稳定，这种关系又能保持到预测期，那么回归方程用于预测是适宜的。否则，应慎重使用。建立回归预测方程时，样本数据不宜过少，因为小样本也许不能真实反映变量之间的结构关系。

表 6.3　　钢材消费量的预测结果

年份	国民收入 (亿元)	钢材消费量 (万吨)	钢材消费量预测区间 (万吨)
1981	3523.7	3006.83	2741.33—3272.33
1982	3682.31	3162.86	2897.36—3428.36
1983	3848.01	3325.91	3060.41—3591.41
1984	4021.17	3496.30	3230.80—3761.80
1985	4202.13	3674.37	3408.87—3939.87

§6.2.5　可化为线性的回归

现实的社会经济现象之间并不都呈线性关系，更多的是非线性的，如图 6.3 所示的单机成本与产量的关系。这些非线性的统计关系，往往也可以配合适宜的曲线模型，但非线性回归不能进行上述的检验和推断，因为那是建立在线性统计模型基础上的。许多非线性回归模型经过适当变换，可以转化为线性回归模型的形式。通常采用的能化为线性回归的曲线模型有

(1)幂函数曲线

$$\hat{Y} = aX^b$$

(2)双曲线

$$\hat{Y} = a + \frac{b}{X}$$

或

$$\frac{1}{\hat{Y}} = a + \frac{b}{X}$$

(3)指数曲线

$$\hat{Y} = ae^{bX}$$

或

$$\hat{Y} = ae^{\frac{b}{x}}$$

(4)对数曲线

$$\hat{Y} = a + b\ln X$$

(5)S 曲线

$$\hat{Y} = \frac{1}{a + be^{-x}}$$

表 6.4 是上述曲线模型的函数变换表。根馅变换以后的线性模型，可以采用最小二乘法确定回归曲线中相应的参数值。

回归曲线可以根据专业知识(理论上的推导或积累的实际经验)选择，也可以将有关样本数据点绘制成相关散点图，根据其分布形状和特点选择。为了判断回归曲线的选择是否合适，可以将变换后的因变量和自变量数据绘制成散点图，如果所有的点集中在一条直线附近，表明回归曲线的选择是恰当的。否则，应调整回归曲线模型。

例 6.7　分析某企业电视机单机成本与月产量之间的回归关系，有关数据如表 6.5。

解：根据生产实际考虑，一般单位成本与产量之间成反比例关系。将表 6.5 的数据绘制成散点图，如图 6.3。由图 6.3 可以看出，电视机单机成本与月产量的关系大致呈双曲线形式(双曲线的一支)，因而可以考虑建立回归曲线

$$\hat{Y} = a + \frac{b}{X}$$

根据表 6.4 提供的变换，可以得到新方程

$$Y' = a + bX'$$

式中：$X' = \frac{1}{X}$。变换后的数据绘制的散点图如图 6.8。由图 6.8 可以看出，Y'与 X'基本呈线性关系，表明建立双曲线回归模型合适。用最小二乘法估计参数，得到月产量与单机成本之间的回归方程 $\hat{Y} = 250.7848 + 355457.05\frac{1}{X}$

表 6.4　　常见曲线的函数变换表

原方程式	自变量和因变量变换函数		新方程式	参数变换公式	
	$Y'=f_1(Y)$	$X'f_2(X)$			
$\hat{Y}=aX^b$	$Y'=\ln Y$	$X'=\ln X$	$Y'=a+bX'$	$a=a'$	$b=b'$
$\hat{Y}=a+\frac{b}{X}$	$Y'=Y$	$X'=\frac{1}{X}$	$Y'=a+bX'$	$a=a'$	$b=b'$
$\frac{1}{\hat{Y}}=a+\frac{b}{X}$	$Y'=\frac{1}{Y}$	$X'=\frac{1}{X}$	$Y'=a+bX'$	$a=a'$	$b=b'$
$\hat{Y}=ab^{bx}$	$Y'=\ln Y$	$X'=X$	$Y'=\ln a+bX'$	$a=a'$	$b=b'$
$\hat{Y}=ae^{\frac{b}{X}}$	$Y'=\ln Y$	$X'=\frac{1}{X}$	$Y'=+\ln abX'$	$a=e^{a'}$	$b=b'$
$\hat{Y}=a+b\ln X$	$Y'=Y$	$X'=\ln X$	$Y'=+abX'$	$a=a'$	$b=b'$
$\hat{Y}=\frac{1}{a+be^{-X}}$	$Y'=\frac{1}{Y}$	$X'=e^{-x}$	$Y'=+abX'$	$a=a'$	$b=b'$

表 6.5　　　　　　　　　　　　　　　　单机成本与月产量

时　间	单机成本(元/台)	月产量(台)	时　间	单机成本(元/台)	月产量(台)
1987 年 1 月	346.23	4300	1987 年 9 月	310.82	6024
2 月	343.34	4004	10 月	306.83	6194
3 月	327.46	4300	11 月	305.11	7558
4 月	313.27	5016	12 月	300.71	7381
5 月	310.75	5511	1988 年 1 月	306.84	6950
6 月	307.61	5648	2 月	303.44	6471
7 月	314.56	5876	3 月	298.03	6354
8 月	305.72	6651	4 月	296.21	8000

图 6.8　变换后的单机成本与月产量散点图

§6.3　多元线性回归

一元线性回归将影响因变量的自变量限制为一个，这在现实的大量社会经济现象中并不易做到。因而，实际应用回归分析法时，常需要有更一般的模型，把两个或更多个解释变量的影响分别估计在内。这就是多元回归亦称多重回归。当影响因素与因变量之间是线性关系时，所进行的回归分析就是多元线性回归。

§6.3.1　多元线性回归的数学模型

当影响变量 Y 的主要因素有 k 个时，可以建立起的总体回归模型为

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\varepsilon \qquad (6.17)$$

这是 Y 对 X_1，X_2，…，X_k 的多元回归，也称多重回归或复回归。$\beta_1,\beta_2,\cdots,\beta_k$ 称为偏回归系数。

模型的基本假设大致与一元线性回归模型相同，只是自变量 $X_1,X_2,\cdots,X_k$ 之间不能有较强的线性关系。

利用变量 Y 与 X 的 n 组样本数据，依照一定准则，可以得到回归系数 $\beta_0,\beta_1,\cdots,\beta_k$ 的估计值 $b_0,b_1,\cdots,b_k$；建立起样本回归模型

$$Y=b_0+b_1X+b_2X_2+\cdots+b_kX_k+e \qquad (6.18)$$

相应的多元线性回归方程为

$$\hat{Y}=b_0+b_1X_1+b_2X_2+\cdots+b_kX_k \qquad (6.19)$$

§6.3.2　参数的最小二乘估计

(6.17)式中参数 $\beta_0,\beta_1,\cdots,\beta_k$ 可以利用变量 Y 与 X_1，X_2，…X_k 的 n 组样本数据，依照最小二乘准则得到，它们可以采用求解正规方程组

计算得到。

式中：$Q=\Sigma(y_i-\hat{y}_i)^2$；$y_i$ 是因变量的观测值，$\hat{y}_i$ 是(6.19)式中的回归值。

当自变量的数目 k 较大时，求解正规方程组(6.20)式很复杂，需应用电子计算机，因而通常将多元线性回归模型表述为矩阵形式。令

$$Y=\begin{bmatrix}y_1\\y_2\\M\\y_n\end{bmatrix}\qquad B=\begin{bmatrix}b_0\\b_1\\M\\b_k\end{bmatrix}\qquad e=\begin{bmatrix}e_1\\e_2\\M\\e_n\end{bmatrix}$$

$$X=\begin{bmatrix}1 & x_{11} & x_{12} & \cdots & x_{1k}\\1 & x_{21} & x_{22} & \cdots & x_{2k}\\M & M & M & \cdots & M\\1 & x_{n1} & x_{n2} & \cdots & x_{nk}\end{bmatrix}$$

则(6.18)式可以写成矩阵形式

$$Y=XB+e$$

依据最小二乘准则，回归系数阵 B 为

$$B=(X'X)^{-1}X'Y \qquad (6.22)$$ [①]

---

① 由于 e2=e'e=(Y-XB)'(Y-XB)=(Y'-B'X')(Y-XB)=Y'Y-B'X'Y-Y'XB+B'X'XB=Y'Y-2B'X'Y+B'XB　　推导过程中，由于 B’X’Y 与 Y’XB 均为秩是 1×1 的矩阵，且有(B’X’Y)’=Y’XB,所以 Y’XB=B’X’Y 由此得　整理得到

式中：X’是矩阵 X 的转置阵；$(X'X)^{-1}$是(X’X)阵的逆阵。

由最小二乘准则得到的回归系数b_j(j=1,2,…,k)表明在其他自变量保持不变的情况下，自变量X_j(j=1,2,…,k)变动一个单位所引起的因变量 Y 的平均变动量。

§6.3.3 模型的检验

多元线性回归模型与一元线性回归模型一样，得到参数的最小二乘估计值后，对模型是否满足基本假设条件也需要进行检验。

1.回归系数的显著性检验。多元线性回归中，需要对每个回归系数的显著性进行检验，以便使模型中只保留那些对因变量有显著影响的因素。其步骤同一元线性回归，只是查 t 分布表时，自由度应是 n-k-1。

回归参数的 t 检验通不过，可能是与这个系数相应的自变量对因变量的影响不显著所致，也可能是自变量之间有共线性所致②。若自变量不是影响因变量的显著因素，应从回归模型中剔除；若自变量间有共线性，应设法消除共线性。

2.回归方程的显著性检验。在一元线性回归中，回归系数的显著性检验(t 检验)与回归方程的显著性检验(F 检验)是等价的，但在多元线性回归中，这个等价不再成立。t 检验是分别检验回归模型中各个系数的显著性，而 F 检验则检验整个回归关系的显著性，即

原假设 H_0：$\beta_1=\beta_2=\cdots=\beta_k=0$

对立假设 H_1：β_j不同时为 0 j=1,2,…,k

计算 F 统计量

$$F=\frac{\Sigma(\hat{y}-\bar{y})^2/k}{\Sigma(y-\hat{y})^2/n-k-1} \tag{6.23}$$

根据给定的显著性水平 a，自由度$df_1=k$,$df_2=n-k-1$，查 F 分布表，得到相应的临界值F_a。若

$$F>F_a$$

拒绝H_0，可以认为回归方程有显著的意义，回归方程回归效果显著。若

$$F\le F_a$$

则不能拒绝H_0，回归方程无显著意义，回归方程回归效果不显著。

3。拟合程度的测定。与一元线性回归中可决系数r^2相对应，多无线性回归中有多重可决系数R^2。它计量了在因变量的变动中，由回归关系解释的变动所占的比重。R^2与r^2一样，定义为

$$R^2=\frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y-\bar{y})^2}=1-\frac{\Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2} \tag{6.24}$$

R^2是样本回归线对样本数据点拟合程度的测度，但它受回归方程中引进自变量数目多少的影响，与因变量有关的因素引进越多，R^2越接近于 1。但这样做，不仅加大计算工作量，还会引起自变量间的共线性。为消除自变量数目对R^2的影响，常采用调整的(修正的)R^2，调整的R^2被定义为

X’XB=X’Y

② 自变量之间有较强的线性关系，如自变量 X1 与 X2 之间可表述为：X1=0.25X2，则称自变量有共线性。

$$\overline{R}^2 = 1 - \frac{\Sigma(y-\hat{y})^2 / n-k-1}{\Sigma(y-\overline{y})^2 / n-1} \qquad (6.25)$$

$\overline{R}^2$与R^2有如下的关系

$$\overline{R}^2 = 1-(1-R^2)\frac{n-1}{n-k-1} \qquad (6.26)$$

多无线性回归中也应进行 D。W 检验并考察回归的估计标准误差，这与一元线性回归中所讲述的没有差别，放不再赘述。

§6.3.4　应用举例

例 6.8　利用表 6.6 提供的数据，考察火柴销售量与各影响因素之间的回归关系。

解：令人柴销售量为因变量 Y，煤气、液化气户效，卷烟销量，蚊香销量，打火石销量分别为自变量 X_1,X_2,X_3,X_4。根据表 6.6 的数字，通过有关统计软件，采用最小二乘法估计参数。得到关于火柴销售量的四元线性回归方程

$$\hat{Y} = 17.3973 + 0.0503X_1 + 0.2551X_2 - 0.0040X_3 - 0.2432X_4$$
$$(2.4002)(11.9004)(-0.1162)(-19.5454)$$

式中：每个回归系数下面括号中的数值是与其相应的 t 值。可以看出，其中 $|tb_3|=0.1162<2$，根据经验[①]可知，回归系数 b_3 的 t 检验未通过，且 b_3 的符号与其实际经济意义相反。蚊香销量增加，火柴销量相应地应该增加，b_3 应为正数。变量调 X_3 在这种情况下引人回归方程是不合适的。剔除 X_3。用其余三个变量建立 Y 的回归方程得

对模型进行各种检验。

1.回归系数的显著性检验。根据显著性水平 a=0.05，df=15-3-1=11，查 t 分布表，得 $t_{0.05/2}=2.2010$，通过在电子计算机上运用统计软件处理得到

$$tb_1=3.5164>t_{0.05/2}=2.2010$$
$$tb_2=13.2278>t_{0.05/2}=2.2010$$
$$|tb_3|=20.5262>t_{0.05/2}=2.2010$$

表明三个回归系数的 t 检验均通过，所选择的自变量是影响火柴销售量的主要因素。

2.回归方程的显著性检验。通过电子计算机处理得 F＝645.4824，根据显著性水平 a=0.05，$df_1=3$，$df_2=11$，查 F 分布表得 $F_{0.05}(3，11)=3.59$，因为

$$F=645.4824>F_{0.05}=3.59$$

所以，F 检验通过，表明回归方程的回归效果显著。

3.D.W 检验。计算残差序列 d 统计量得 d＝2.0661，根据显著性水平 a=0.05，样本数据个数 n＝15，自变量个数 k＝3，查 D.W 表得 $d_l=0.82$，$d_u=1.75$。由于 $d_u=1.75<d=2.0661<4-d_u=2.25$，D.W 检验通过，表明残差序列无自相关。

[①] 经验规则，一般取 t0.05/2=2 来判断。

4.拟合程度测定。计算得到 R^2＝0.9944，$\overline{R}^2$＝0.928，接近于 1，表明回归线对样本数据点的拟合程度很高。

5.回归标准误差。计算得到的回归估计标准误差 S_y=0.4178，表明估计标准误差很小。

结论：回归方程通过了模型的所有统计检验，表明以煤气、液化气户数，卷烟销售量，打火石销售量来解释说明火柴销售量的变化是适宜的，所建立的回归方程表述了这种回归关系。

习 题

1.某市电子工业公司有 14 个所属企业,各企业年设备能力与年劳动生产率统计数据如表 1。试分析企业年设备能力与年劳动生产率的关系。若该公司计划新建一个设备能力为 9.2 千瓦/人的企业，估计劳动生产率将为多少。

表 1

企　业	设备能力(千瓦)	劳动生产率(千元/人)	企　业	设备能力(千瓦/人)	劳动生产率(千元/人)
1	2.8	6.7	8	4.8	9.8
2	2.8	6.9	9	4.9	10.6
3	3.0	7.2	10	5.2	10.7
4	2.9	7.3	11	5.4	11.1
5	3.4	8.4	12	5.5	11.8
6	3.9	8.8	13	6.2	12.1
7	4.0	9.1	14	7.0	12.4

2.对某市的百货商店进行抽佯调查,其中被抽查的 10 家商店职工月平均销售额和利润率的数字如表 2。试分析两个变量间存在的关系，并建立利润率对销售额的回归方程。

表 2

商店	人均月销售额(千元)	利润率(%)	商店	人均月销售额(千元)	利润率(%)
1	6	12.6	6	7	16.3
2	5	10.4	7	6	12.4
3	8	18.5	8	3	6.2
4	1	3.0	9	3	6.5
5	4	8.1	10	7	16.8

3.某广告公司对购买该公司广告劳动的日用化工厂作随机调查。调查内容一是作广告后一年内销售额比这以前 12 个月销售额的增长率,二是广告作出后第 3 个月末顾客中知道广告商品的人数比率(商品知悉率)。8 家被调查厂家的有关资料如表 3。试确定商品知悉率为 20%时，销售额增长率平均水平的 95%置信区间。

表 3

工厂	销售额增长率 (%)	商品知悉率 (%)	工厂	销售额增长率 (%)	商品知悉率 (%)
1	11	13	5	21	18
2	48	22	6	35	15
3	72	29	7	82	22
4	95	30	8	62	24

4.某公司的决策者通过反复调查分析认为，影响该公司总销售额的因素主要有：人均生活费收入、商品价格、投资额和广告费用。根据每半年一期的统计，有关数据如表 4。试分析评价总销售额对四个变量的回归方程。

5.对经营同一类产品出口业务的公司进行抽样调查，被抽查的 13 家公司，其出口换汇成本与商品流转费用率资料如表 5。试分析两个变量之间的关系，并估计某家公司商品流转费用率是 6.50%的出口换汇成本。

第 7 章　时间序列与指数

§7.1　时间序列的分析指标

反映社会经济现象的数据大部分是按时间顺序记录下来的，这些各期记录的观察值的序列就叫做时间序列。如 1983 年 1 月—1988 年 12 月各月的工业总产值，

1978 年—1989 年每年的粮食产量等。对时间序列进行分析的目的是描述时间序列的过去行为，分析这种行为，从而进一步预计未来的情况。对时间序列过去行为的描述可以采用一系列动态分析指标，如平均发展水平、平均发展速度。

§7.1.1　发展水平和平均发展水平

发展水平是时间序列中原有的统计指标数值，它通常用符号 a 表示。$a_0, a_1, \cdots, a_n$ 是序列各个时期(或时点)的发展水平，其中 a_0 是最初水平，a_n 是最末水平，中间各项是中间各时期(或各时点)的水平。若有 1983 年 1 月到 1988 年 12 月我国工业总产值的时间序列，那么，1983 年 1 月工业总产值是序列的最初水平，1988 年 12 月的工业总产值是序列的最末水平。在研究某一时期的发展水平时，常把研究的那个时间的发展水平称作报告期水平或计算期水平，用来作为比较基础的时间的发展水平称为基期水平。如对比 1988 年 12 月与 1983 年 12 月的工业总产值,则 1988 年 12 月的工业总产值是报告期水平，1983 年 12 月为基期水平。基期和报舍期是随对比的时间而确定。

平均发展水平是把不同时间的发展水平加以平均所得到的平均数，统计上称为序时平均数。它将社会经济现象在不同时间上的数量差异抽象化，从动态上反映现象在一段时间的一般发展水平。序时平均数在时间序列的动态分析中，可以用来修匀序列，消除现象在短时间内的波动，使序列能更明显地反映出现象的发展变化趋势。序时平均数还广泛用来对比不同单位、不同地区、不同部门以至不同国家在某一时间内现象发展的一般水平。

序时平均数因时间序列指标性质的不同而有几种计算公式。

若时间序列反映的是一个时期指标在各个时期的发展水平，那么计算序时平均数可以采用下面公式：

$$\bar{a} = \frac{a_1 + a_2 + \cdots + a_n}{n} = \frac{\Sigma a_i}{n} \tag{7.1}$$

式中：$\bar{a}$ 是序时平均数；a_i ($i=1,2,\cdots,n$)是各个时期的发展水平；n 是时期数目。

用(7.1)式可以计算 1983 年至 1988 年的年平均工业总产值，1989 年 1 月至 12 月的月平均工资等。若序列表现的是某一时点指标在不同时期的发展水平，则当整个研究期的各个时点数据齐备时，也可以采用上式。如某工厂利用 1989 年 10 月份各日的工人人数，求 10 月份的平均工人数；某商店利用 8 月份每日的商品库存额，计算该月商品平均库存额等。

若时间序列反映某一时点指标在不同时期的发展水平，而时点数据资料不齐备，因而不连续，但时点的时间间隔相等时，可以采用下列公式计算序时平均数。

$$\bar{a} \frac{\frac{1}{2}a_1 + a_2 + \cdots + a_{n-1} + \frac{1}{2}a_n}{n-1} \qquad (7.2)$$

式中各符号意义同(7.1)式。

例 7.1 某工厂职工人数资料如表 7.1，试计算第二季度的平均职工人数。

表 7.1 第二季度职工人数 (单位：人)

	4月1日	5月1日	6月1日	7月1日
职工人数	2040	2035	2045	2058

解：假设职工人数在两个时点之间的变动是均匀的(因为表 7.1 提供的数据可认为时点间的间隔是相等的)，故采用(7.2)式计算第二季度平均职工人数：

$$\bar{a} = \frac{\frac{1}{2} \times 2040 + 2035 + 2045 + \frac{1}{2} \times 2058}{4-1} = 2043(\text{人})$$

该厂第二季度平均职工人数为 2043 人。

若时间序列反映某一时点指标在不同时期的发展水平，而掌握的时点数据资料的时间间隔不相等，则需以时间间隔长度 f 为权数，采用下列公式计算序时平均数。

$$\bar{a} = \frac{(\frac{a_1 + a_2}{2}) \times f_1 + (\frac{a_2 + a_3}{2}) \times f_2 + \cdots + (\frac{a_{n-1} + a_n}{2}) \times f_{n-1}}{\sum_{i=1}^{n-1} f_i} \qquad (7.3)$$

例 7.2 某商店 1988 年商品库存额资料如表 7.2。试计算全年平均库存额。

表 7.2 商品库存额数据 （单位：万元）

时 间	库存额	时 间	库存额
1月1日	5.2	9月30日	4.2
3月31日	3.6	12月31日	5.6
5月31日	3.0		

解：假定商品库存额在两个时点之间的变动是均匀的，由于时点数据资料的时间间隔不等，故采用(7.3)式求年平均库存额。

$$\bar{a} = \frac{(\frac{5.2+3.6}{2}) \times 3 + (\frac{3.6+3.0}{2}) \times 2 + (\frac{3.0+4.2}{2}) \times 4 + (\frac{4.2+5.6}{2}) \times 3}{12}$$
$$= 4.075$$

该商店 1988 年商品平均库存额为 4.075 万元。

§7.1.2 发展速度和平均发展速度

发展速度是时间序列中两个时期发展水平的比，即

发展速度＝报告期水平/基期水平

它是用来研究社会经济现象发展程度的相对指标，说明报告期水平已发展到基期水平的若干倍或百分之几。由于计算发展速度时采用的基期不同，发展速度可分为定基与环比两种。

定基发展速度是以各个报告期水平同某一固定基期发展水平之比。若以a_0表示固定基期，则定基发展速度为

$$\frac{a_1}{a_0},\frac{a_2}{a_0},\cdots,\frac{a_{n-1}}{a_0},\frac{a_n}{a_0} \tag{7.4}$$

定基发展速度用来说明被研究现象在一定时期内总的发展情况。

环比发展速度是用各报告期水平同前一期水平相比。若时间序列是：$a_0,a_1,a_2,\cdots,a_n$,那么，环比发展速度为

$$\frac{a_1}{a_0},\frac{a_2}{a_1},\frac{a_3}{a_2},\cdots,\frac{a_{n-1}}{a_{n-2}},\frac{a_n}{a_{n-1}} \tag{7.5}$$

环比发展速度用来说明被研究现象逐期发展变化的情况。

表 7.3 是某企业 1984 年到 1988 年工业生产的发展情况。从(7.4)式和(7.5)式以及表 7.3 提供的数据可以看出,定基发展速度和环比发展速度虽然各说明不同的问题，但它们之间存在一定的数量关系，即

表 7.3　　某企业工业生产发展情况　　(单位：万元)

	1984 年	1985 年	1986 年	1987 年	1988 年
工业总产值 (1970 年不变价格)	677	732	757	779	819
定基发展速度(%)	—	108.12	118.82	115.07	120.97
环比发展速度(%)	—	108.12	103.42	102.91	105.13

$$\frac{a_n}{a_0}=\frac{a_1}{a_0}\times\frac{a_2}{a_1}\times\frac{a_3}{a_2}\times\cdots\times\frac{a_n}{a_{n-1}} \tag{7.6}$$

(7.6)式表明定基发展速度等于相应的各环比发展速度的连乘积。利用(7.6)式二者之间的关系，可以进行相互间的推算。

发展速度不仅表明社会经济现象发展的程度，还表明其发展的方向。若发展速度大于 1 即大于 100%，说明现象是上升的发展趋势；着小于 1 即小于 100%，说明现象是下降的发展趋势。

平均发展速度是某一段时间内，各时期环比发展速度的平均数，用以说明现象在这段时间内逐年平均发展变化的程度。由于社会经济现象在各个时期所处的条件及影响其变化的因素不同，因而各时期的发展速度有差别，平均发展速度通过对各个时期发展速度的平均，消除了差别，便于对不同时期社会经济现象的发展变化情况进行对比。它是编制计划的依据，也常是进行各种推算和预测的依据。

平均发展速度依据速度指标的特性采用不同的方法计算，常采用的有几何平均法和方程法两种。

1.几何平均法即水平法。若以x_1，x_2，…，x_n分别表示各期的环比发展

速度，则这段时间年的平均发展速度$\bar{x}$为

$$\bar{x}=\sqrt[n]{x_1\cdot x_2\cdots x_n}=\sqrt[n]{\Pi x} \tag{7.7}$$

根据(7.6)式可知，平均发展速度$\bar{x}$也可为

$$\bar{x}=\sqrt[n]{a_n/a_0} \tag{7.8}$$

(7.8)式表明，几何平均法的平均发展速度，实际上只与时间序列的最初水平 a_0 和最末水平 a_n 有关。当最初水平 a_0 作为比较的基期时，平均发展速度的大小只取决于最末水平 a_n，而与时间序列中间各时期的水平无关。因此，它无法反映序列中间各时期水平的发展变化。

利用表 7.3 提供的统计数字，可以计算 1984 年到 1988 年间某企业工业总产值的平均发展速度。根据(7.7)式计算

$$\begin{aligned}\bar{x}&=\sqrt[n]{a_n/a_0}\\&=\sqrt[4]{108.12\%\times103.42\%\times102.91\%\times105.13\%}\\&=104.88\%\end{aligned}$$

根据(7.8)式计算

$$\bar{x}=\sqrt[n]{a_n/a_0}=\sqrt[4]{819/677}=104.88\%$$

2.方程法即累计法。时间序列的各期发展水平为 a_0，a_1，…，a_n，环比发展速度为 x_1，x_2，…，x_n，平均发展速度为$\bar{x}$，则从最初水平 a_0 出发，每期按固定的平均发展速度发展，则有

$$a_0\bar{x}+a_0\bar{x}^2+\cdots+a_0\bar{x}_n=a_1+a_2+\cdots+a_n$$

即

$$\bar{x}+\bar{x}^2+\bar{x}^3+\cdots+\bar{x}^n=\frac{\sum_{i=1}^{n}a_i}{a_0} \tag{7.9}$$

解这个高次方程，得到的$\bar{x}$的正根就是所求的平均发展速度。直接解高次方程相当复杂，通常利用编好的累计法平均增长速度查对表计算。

对时间序列的同一段时间内计算的累计法平均发展速度和水平法平均发展速度不一致。水平法主要考虑最后一年所达到的水平，侧重点在最末发展水平；累计法则主要考虑整个时期中各期发展水平的总和，侧重点在各期发展水平的累计总和。实际应用时要依据研究现象的性质、研究的目的和要求确定采用哪一种方法。一般来说，主要关心一段时期最后达到的水平，适宜采用水平法，如各种产品的产量，国民收入，物价指数等等。若主要关心一段时期内完成的总量，适宜采用累计法，如固定资产的投资，培养的大学生人数，住宅的完工面积等。

§7.1.3　增长量、增长速度和平均增长速度

增长量是两个时期发展水平的差值，即

增长量＝报告期发展水平－基期发展水平

差值大于 0 是增长量，小于 0 是减少量。它反映报告期比基期增减的绝对数量。因对比的基期不同，分为逐期增长量和累积增长量。

逐期增长量是以前一期为基期计算的相邻两个时期的增减绝对量。若以 a_0，a_1，…，a_n 表示时间序列各期发展水平，则逐期增长量为

$$a_1-a_0, a_2-a_1, \cdots, a_n-a_{n-1}$$

累积增长量是以某一固定时期为基期计算的各报告期与某一固定时期的增减绝对量，即

$$a_1-a_0, a_2-a_0, \cdots, a_n-a_0$$

它说明在某一段时期内总的增减量。累积增长量等于相应各个时期逐期增长量之和，即有

$$a_n-a_0=(a_1-a_0)+(a_2-a_1)+\cdots+(a_n-a_{n-1})$$

增长速度是增长量与基期水平的比值，即

$$\text{增长速度}=\frac{\text{增长量}}{\text{基期水平}}=\frac{\text{报告期水平}-\text{基期水平}}{\text{基期水平}}=\text{发展速度}-1$$

它用以说明社会经济现象的增长程度，当发展速度大于 1(或 100%)时，增长速度大于 0，表明现象增长的程度；发展速度小于 1，增长速度小于 0，表明现象下降的程度。增长速度与发展速度一样，也分为定基和环比两种。

定基增长速度是用累积增长量与固定基期水平相比，即

$$\text{定基增长速度}=\frac{\text{累积增长量}}{\text{固定基期水平}}=\text{定基发展速度}-1$$

它用以反映社会经济现象在一段较长的时间内总的增长程度。

环比增长速度是用逐期增长量与前期水平相比，即

$$\text{环比增长速度}=\frac{\text{逐期增长量}}{\text{前期水平}}=\text{环比发展速度}-1$$

它用以反映社会经济现象逐期的增长程度。

平均增长速度是环比增长速度的平均数，但不能根据各个环比增长速度直接求得，而是根据平均发展速度计算，即

$$\text{平均增长速度}=\text{平均发展速度}-1$$

§7.2 传统时间序列分析

§7.2.1 时间序列的分解

传统时间序列分析是将时间序列分解为 4 种因素：

长期趋势(T，Trend)

季节变动(S，Seasonal variation)

循环变动(C，Cyclical variation)

不规则变动(I，Irregular variation)

长期趋势变动是指时间序列在一段长时期的变动，若将其用图形表现，可得一长期趋势线。若趋势线是直线，则称为直线趋势；若趋势线是曲线，则根据其曲线形式称为某种曲线趋势，如二次曲线趋势，指数曲线趋势等。趋势线表示着时间序列长期的动态,例如我国近 40 年的国民收入增长趋势是 40 年间的动态，人口近 30 年的增长趋势是 30 年间的动态，等等。这种长期趋势，通常可认为是由各种固定的因素作用于同一方向而形成的。这些固定的因素，既有随时间的推移按直线变化的，也有表现为各种曲线形状的。

季节变动是时间序列由季节性原因而引起的周期性变动，许多经济领域的时间序列都受这种变动的支配。当反映时间序列的数据是按周、月或季的间隔记录时，季节变动很明显。如汗衫、背心的月零售量，糕点的季销售额，平板玻璃的月产量等。虽然不同时间序列季节变动的幅度不同，但它们的周期却是固定的，一般均为一年。在按年记录的时间序列中，季节波动不会出现。

循环变动是以年度记录的时间序列所表现出的某种周期性变动。它与季节变动不同，循环的幅度和周期都可以不很规则。如某些经济活动序列表现出的以 8 年或 9 年为一个周期的循环，这种循环通常称为商业循环。

不规则变动是时间序列除去长期趋势、季节变动和循环变动(若存在此种变动的话)之后余留下来的变动。这种变动细分为两种类型，一是严格的随机变动，它是由许多细小的原因综合引起的，以一种纯随机的方式使序列产生上下波动；二是偶然性变动，它是不经常出现的某些孤立的或不规则的、但却是强有力的突发性变动。如政治动荡，战争爆发，大的自然灾害产生的影响。在传统时间序列分析中，一般须把它与那些经常起作用的更为随机的变动区别开来。

§7.2.2 传统时间序列分析的基本假设

传统时间序列分析的直接目的是把序列分解为长期趋势、季节变动、循环变动和不规则变动几个主要构成部分。也就是把序列的变动看做是由这些因素复合而成，并试图分别揭示出各个因素数量的大小，表明各构成部分的变动如何在一起形成这个序列的变动。由于采用的研究方法是历来传统的方法, 与 1940 年前后开创的将时间序列用概率过程进行分析的方法不同，因而称作传统时间序列分析。

传统时间序列分析方法主要取决于对序列的各个构成部分如何结合及相互作用的假设。最简单的假设是，假定各构成部分影响的数量值可加且是相互独立的，这样时间序列的 4 因素表现为

$$Y_t=T_t+S_t+C_t+I_t \tag{7.10}$$

式中：Y_t 是时间序列在 t 时刻的数值；T_t、S_t、C_t、I_t，分别为同一时刻序列的趋势值、季节变差、循环变差、不规则变差。

显然，若是年度的时间序列，季节变动不会出现，时间序列就分解为

$$Y_t=T_t+C_t+I_t$$

(7.10)式是假设季节、循环、不规则变动是围绕一个正常趋势变动的一种波动，这里 S_t、C_t、I_t 都是正值或负值，它们分别表示由于季节、循环、不规则变动的影响，在序列 t 时刻的趋势值上增加或减少若干单位。(7.10)式反映了时间序列的发展变化是由 4 种因素的作用迭加而成。

相对于上述关系的另一假设是

$$Y_t=T_t \cdot S_t \cdot C_t \cdot I_t \tag{7.11}$$

这里，S_t、C_t、I_t 是在 T 上下波动的数值，通常称为指数，它们分别表示由于季节、循环、不规则变动的影响，在序列 t 时刻的趋势值增加或减少的百分比。(7.11)式反映了时间序列的发展变化是由 4 种因素的作用交乘而成。由于(7.11)式在两边取对数后，也成为(7.10)式的形式，因此可以理解这两种假设在原则上没有区别，都是假设时间序列诸因素是可加的。

§7.2.3　趋势的测量

估计时间序列趋势的直接目的是描述这个序列的总的最基本的变动。寻求趋势线的方法很多，常用的有以下几种。

1.随手描绘法。将时间序列的数据描绘在平面坐标图上，对这些数据点随手画一条光滑的曲线(或直线)，以其表示序列的趋势。这种方法简单却很粗糙，它可以大体上区分出序列的趋势是直线还是曲线。

2.移动平均法，将时间序列的数据逐项移动，依次计算包含一定期数的序时平均数，形成一个新的序时平均数序列，这就是移动平均法。采用移动平均法分析时间序列的趋势变动，关键在于平均期数或称移动步长的选择。移动步长为奇数时，移动平均数就是平均期中间一期的“修匀”值；移动步长为偶数时，要进行二次平均。表 7.4 和表 7.5 列出了移动步长为 3 和 4 的计算。

表 7.4　　　　移动步长为 3 的计算式

时期	序列数据(Y_t)	移动平均数
1	Y_1	—
2	Y_2	$\frac{1}{3}(Y_1+Y_2+Y_3)$
3	Y_3	$\frac{1}{3}(Y_2+Y_3+Y_4)$
4	Y_4	$\frac{1}{3}(Y_3+Y_4+Y_5)$
5	Y_5	

应用移动平均法时应注意,若序列有围绕趋势的周期性变动(季节的或循环的)，移动步长应与周期相同，以消除这些变动；若序列循环周期和(或)幅度是变动的，移动平均法将无法消除全部循环成分。

3.最小二乘法。如果时间序列随着时间的推移表现出某种趋势，可以采用最小二乘法为其配合一适当的趋势线。第 6 章中介绍过的直线及可以线性

化的曲线模型，将自变量调换成时间 t， 即 $Y_t=a+bt$ 等都可以用来描述时间序列的长期趋势。

例 7.3 利用表 7.6 的数据为我国 1969—1980 年钢材消费量配合一条趋势线。

表 7.5 移动步长为 4 的计算式

时期 (t)	序列数据 (Y_t)	4 年移动平均数 ($Y_t^{'}$)	$Y_t^{'}$ 的 2 年移动平均数
1	Y_1		
2	Y_2	$Y_1^{'}=\frac{1}{4}(Y_1+Y_2+Y_3+Y_4)$	$\frac{1}{2}(Y_1^{'}+Y_2^{`})$
3	Y_3	$Y_2^{'}=\frac{1}{4}(Y_2+Y_3+Y_4+Y_5)$	$\frac{1}{2}(Y_2^{'}+Y_3^{`})$
4	Y_4	$Y_3^{'}=\frac{1}{4}(Y_3+Y_4+Y_5+Y_6)$	$\frac{1}{2}(Y_3^{'}+Y_4^{`})$
5	Y_5	$Y_4^{'}=\frac{1}{4}(Y_4+Y_5+Y_6+Y_7)$	$\frac{1}{2}(Y_4^{'}+Y_5^{`})$
6	Y_6	$Y_5^{'}=\frac{1}{4}(Y_5+Y_6+Y_7+Y_8)$	$\frac{1}{2}(Y_5^{'}+Y_6^{`})$
7	Y_7	$Y_6^{'}=\frac{1}{4}(Y_6+Y_7+Y_8+Y_9)$	$\frac{1}{2}(Y_6^{'}+Y_7^{`})$
8	Y_8	$Y_7^{'}=\frac{1}{4}(Y_7+Y_8+Y_9+Y_{10})$	
9	Y_9		
10	Y_{10}		

表 7.6 我国 1969 — 1980 年钢材消费量 (单位 :万吨)

年份	时序	钢材消费量	年份	时序	钢材消费量
1969	1	1025	1975	7	1960
1970	2	1316	1976	8	1902
1971	3	1539	1977	9	2013
1972	4	1561	1978	10	2446
1973	5	1765	1979	11	2736
1974	6	1726	1980	12	2825

解：将表 7.6 数据绘制成时间序列曲线图，如图 7.1。由图可知，钢材消费量的变动趋势类似直线。采用最小二乘法估计参数，得到趋势线为

$$\hat{Y}_t = 950.3939 + 146.2727t$$

系数 b＝146.2727 表明 1969—1980 年钢材消费量的趋势值的逐年增长额。图 7.1 上的直线是这条趋势线。可以看出，它基本上反映了该期间我国钢材消费量的平均变动趋势。

图 7.1　1969—1980 年我国钢材消费量曲线

时间序列的增长变动有极限时，可以用一类有增长上限的数学曲线描述其长期趋势。常用的曲线有

修正指数曲线　$\hat{Y}_t = L + ae^{bt}$

龚珀兹曲线　$\hat{Y}_t = Le^{-ae^{-bt}}$　(a>0,b>0)

或　$\hat{Y}_t = La^{bt}$

皮尔曲线　$\hat{Y}_t = \dfrac{L}{1 + ae^{f(t)}}$

式中：f(t)是关于 t 的多项式函数。

当极限值 L 可以事先确定时，能够采用最小二乘法估计参数 a、b；当极限值 L 无法预先确定时，可运用三和值法对三个参数 L、a、b 进行粗略的估计。

例 7.4　试利用表 7.7 数据，为某地收音机销售量配合一条适宜的趋势线。

表7.7　某地区收音机销售量　(单位：万台)

年份	序号（t）	销量（Yt）	销量的倒数($\frac{1}{Y_t}$)
1975	1	11.02	0.09074
1976	2	17.80	0.05618
1977	3	27.85	0.03591
1978	4	42.08	0.02376
			第一段和：S_1=0.20659
1979	5	56.07	0.01783
1980	6	69.45	0.07440
1981	7	80.05	0.01249
1982	8	89.98	0.01111
			第二段和：S_2=0.05583
1983	9	94.06	0.01063
1984	10	98.13	0.01019
1985	11	98.72	0.01013
1986	12	101.08	0.00989
			第三段和：S_3=0.04084

解：将表 7.7 数据绘制成收音机销量 1975—1986 年的时序曲线图如图 7.2。图中实线部分代表收音机销售量实际值；虚线部分代表趋势线的估计

值。从图 7.2 中可以看出，该地区收音机销售量在经过 1975—1982 年的迅速增长期后已进入缓慢增长阶段，可以用皮尔曲线进行拟合。趋势线为

图 7.2　某地区 1975—1986 年收音机销售量曲线

$$\hat{Y}_t = \frac{L}{1 + ae^{-bt}}$$

式中：t 是时间，1975 年是 l；L 是增长上限。现采用三和值法计算参数 a、b 及 L。将序列分为三段，每段 4 期，即 n＝4。每段的和 S_1、S_2、S_3，见表 7.7 的第四列，根据

$$D_1 = S_1 - S_2 = 0.15076$$

$$D_2 = S_2 - S_3 = 0.01499$$

$$b = \frac{1}{n}(\ln D_1 - \ln D_2) = 0.5771$$

$$L = \frac{n}{S_1 - \dfrac{D_1^2}{D_1 - D_2}} = 102.093$$

$$a = \frac{L}{C} \cdot \frac{D_1^2}{D_1 - D_2} = 14.8197$$

式中　$C = \dfrac{e^{-b}(1 - e^{-nb})}{1 - e^{-b}} = 1.15329$

可以得到趋势线为 $\hat{Y}_t = \dfrac{102.093}{1 + 14.8197 e^{-0.5771 t}}$

将趋势线绘制在图 7.2 上，可以看出它较好地反映了该地区收音机销售量在 1975—1986 年期间的变化趋势。

§7.2.4　季节变动的测量

在测定时间序列的季节变动时，依据时间序列是否同时具有长期趋势而采用不同的方法。

1.按月(或季)于均法。按月(或季)平均法是根据序列月度(或季度)数据，先求得若干年内同月(或季)平均数，再求出若千年内总的月(或季)平均数，然后将二者相比，求得各月(或季)的季节比率。季节比率 f_t，也称为季节指数，计算公式为

$$f_t = \frac{\text{同月(季)平均数}}{\text{总平均数}} \tag{7.12}$$

它反映序列在某月(或季)内由于受季节变动影响高于或低于总平均数的百分比。这种方法仅适用于没有明显趋势变动只有季节变动的时间序列，一般至少有三年分月(或季)的数据资料。

例 7.5　测定某市汗衫背心零售量的季节变动。有关数据如表 7.8。

解：为观察序列是吝具有趋势变动，绘制时间序列曲线图，如图 7.3。从图 7.3 中可以看出，在那几年中，汗衫背心零售量无

明显的趋势变动，但季节变动明显，每年夏季销量多，冬季销量少，现采用按月平均法测定季节变动。计算 1979 年至 1982 年同月的平均数，列入表 7.8 中，再计算 1979 年至 1982 年各月的平均数，也列入表 7.8 中，然后计算这几年总的月平均数。最后，根据(7.12)式求得季节指数，列入表 7.8

的最后一列。为了更清楚地看出序列的季节变动，可以将各月的季节指数绘制成图，如图 7.4。

季节指数可以简明地表示季节变动的数量。如例 7.5 中，一月份季节指数为 19.10%，表示一月份汗衫背心零售量为年平均

图 7.3　某市汗衫背心零售量曲线

月销售量的 19.10%。各月的季节指数说明各月销量比年总平均水平高或低的程度，反映了季节变动的一般规律。例 7.5 中一、二月份为销售的淡季，三月开始上升，五、六、七月是销售旺季，八月开始下降。这就是某市汗衫背心零售量的季节变动规律。

图 7.4　季节指数图

2.趋势剔除法。时间序列的趋势变动和季节变动同时存在，应先将序列的趋势剔除，再来测定季节变动。趋势剔除法就是解决这一问题的方法。其计算步骤如下：

(1)根据时间序列的数据求出各期趋势值 V_t。

(2)将时间序列的实际观测值 Y_t 除以趋势值 V_t，得到各个月的季节指数，由于它是利用样本数据直接求得，故也称为样本季节指数。样本季节指数中不可避免地包含不规则变动，它只能反映在特定的某月(或季)由于季节变动和不规则变动所造成的高于或低于趋势值的比率，即

$$Y=T\cdot S\cdot I$$

$$Y/T=S\cdot I$$

为了反映季节变动所造成的影响，应消除不规则变动。

(3)将计算得到的各年同月样本季节指数进行简单算术平均即可得到理论季节指数。它已经不包含不规则变动，而反映序列季节变动的一般规律。

上面的步骤也可以用下面的式子表示

$$\frac{Y_t}{V_t}=f_t \qquad (7.13)$$

$$f_i=\frac{f_i+f_{i+T}+\cdots+f_{(m-1)T+i}}{m} \qquad i=1,2,\cdots,T \qquad (7.14)$$

式中：f_t 是样本季节指数，f_i 是理论季节指数，T 是季节周期的长度，m 是时间序列数据所包含的年数。

利用趋势剔除法测定季节变动时，可以用移动平均法得到各期的趋势值 V_1。由于季节周期通常为 4 或 12，故移动步长为偶数，需进行两次移动平均才可得到各期趋势值。此外，确定各期趋势值也可以用最小二乘法。理论季节指数的平均值应为 1。若遇到均值不为 1 的情况，应对其进行调整，记 f_i 的均值为 $\bar{f}$，T 为季节周期，则

$$\bar{f}=\frac{1}{T}(f_1+f_2+\cdots+f_r)$$

正规化(调整)后的理论季节指数 f_i' 为

$$f_i'=\frac{1}{\bar{f}}\cdot f_i \qquad i=1,2,\cdots,T \qquad (7.15)$$

例 7.6　试测定我国 1984 年 1 月到 1988 年 12 月工业总产值的季节变动。有关数据见表 7.9。

解：将表 7.9 数据绘制成图，如图 7.5，从图 7.5 中可以看出，工业总产值序列随着时间的推移呈现上升趋势。为测定其季节变动，应先剔除趋势。采用移动平均法求各期趋势值。因为季节周期为 12 个月，故取移动步长为 12。先计算 12 个月的移动平均数，再依次对它们计算两个月的移动平均数，以把它们置于适当的中心，与某一时期相对应。最好的做法是先计算 12 个月的移动总

图 7.5　工业总产值序列曲线

和，然后把相邻的总数相加，除以 24[①]，这些移动平均数将代表季节和不规则变动被消除后的序列，即序列的趋势值。

将序列的实际观测值与移动平均值相比，得到各期季节指数，见表 7.9。然后将每年同月的季节指数加以平均，得到理论季节指数，见表 7.10。

表 7.10　　工业总产值序列的季节指数

月份	季节指数(%)	月份	季节指数(%)
1	95.37	7	96.84
2	83.74	8	97.03
3	102.23	9	101.29
4	102.99	10	99.68
5	104.38	11	102.12
6	108.31	12	105.33

§7.2.5　循环变动和经济周期

当时间序列具有季节变动时，采用移动步长为季节周期长度进行移动平均,所得到的移动平均值实际上既包含了趋势变动也包含了循环变动(序列若存在循环变动)。因此，有必要将长期趋势因素与循环变动因素分离开来，可以确定一种能够最好地描述序列的长期趋势类型，以得到相应的长期趋势值 T，然后将移动平均数 MA 序列，除以长期趋势值 T，即可得一组循环变动值 C

$$\frac{MA}{T}=\frac{T\times C}{T}=C$$

确定长期趋势类型，常用最小二乘法建立各种趋势线来解决。

[①] 假设 Tl 代表序列头 12 个数的总和，T2 代表其次的 12 个数的总和，由表 7.9 知，

§7.3　现代时间序列分析

现代时间序列分析是将时间序列看作一个随机过程，通过分析时间序列的特性来认识序列变化的一般规律，又通过为序列配合一个按照一定数学理论建立的综合模型来描述序列的一般变化规律，并在获得最佳模型的基础上预测序列未来的状况。现代时间序列分析需要大量数据并借助电子计算机进行复杂的运算。

§7.3.1　自相关

现代时间序列分析运用的主要工具之一是时间序列的自相关。自相关是指时间序列诸项之间的单重相关，如 Y_i 与 Y_{t-1} 之间或 Y_i 与 Y_{t-2} 之间的相关等。时间序列自相关的程度用自相关系数 r_k 度量。k 是时滞，即时间序列两项之间的时间间隔。共计算公式为

$$r_k = \frac{\sum_{t=1}^{n-k}(Y_t - Y_t)(Y_{t+k} - Y_t)^2}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2} \qquad (7.16)$$

式中：n 是时间序列的观测值数目；Y_t 是 n 个样本数据的平均值；Y_t 是时间序列在 t 时刻的值；Y_{t+k} 是时间序列与 t 时刻相隔 k 期的值。r_k 的取值范围是[－1，＋1]，它的意义与简单相关系数 r 实质上一样。

总体自相关系数 ρ_k 根据不同的 k 值组成，亦称为自相关函数。自相关函数通常从样本中取得估计值，即由公式(7.16)求得的 r_k。序列不同时滞 k 的自相关系数可以绘制成自相关图，图 7.6 是某时间序列的自相关图。

图 7.6　某序列的自相关图

从理论上讲，由随机数字构成的序列，其自相关系数应该是 0。事实上，这只对样本容量无限大成为可能。从一个总体中随机抽取的不同样本，计算得到的样本自相关系数，围绕总体自相关系数 ρ_k 构成一种分布。当序列诸项之间没有相关时，样本自相关系数的抽样分布近似于以 0 为均值，以 $1/\sqrt{n}$ 为标准差的正态分布。这样，可以建立序列自相关系数的随机区间 $\pm t\frac{1}{\sqrt{n}}$，t 为一定置信水平的概率度。如样本数据 n＝60，置信水平为 95%则 t＝1.96，随机区间为±1.96x0.129。在序列自相关图上标出陡机区间就是自相关分析图，它可以用来分析时间序列的特性。

§7.3.2　偏自相关

时间序列 Y_t 与 Y_{t-k} 之间的相关是与中间各项 Y_{t-1}，Y_{t-2}，…，Y_{t-k+1} 的相关结合在一起的，为了排除中间诸项因素的影响，只观察 Y_t 与 Y_{t-k}，之间的相关，需要计算偏自相关，在时间序列中，偏自相关是在给定了 Y_{t-1},Y_{t-2},…,Y_{t-k} 的条件下，Y_t 与 Y_{t-k+1} 之间的条件相关。设{Y_t}为平稳、零均值序列，考虑由 Y_{t-1},Y_{t-2},……,Y_{t-k} 对 Y_t 的线性最小方差估计时，即选择系数 ϕ_{k1},ϕ_{k2},…,ϕ_{kk},使得

$$\begin{aligned}\delta &= E(Y_t - \sum_{j=1}^{k}\phi_{kj}Y_{t-j})^2 \\ &= r_0 - Z\sum_{j=1}^{k}\phi_{kj}\gamma_j + \sum_{j=1}^{k}\phi_{kj}\phi_{ki}\gamma_{j-i}\end{aligned}$$

达到极小。为此目的，对δ的右端分别求偏导数 $\frac{\partial\delta}{\partial\phi_{kj}}$ **(j＝1，2，…，k)，并令其为零，使得到** ϕ_{kj}**，应该满足的线性方程组**

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ & \cdots & & \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ M \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ M \\ \rho_k \end{bmatrix}$$①

由上述线性方程组求得的序列 ϕ_{kk} **(k=1,2,…)称为序列{Yt}的偏自相关函数。实际求解方程组时，都是利用样本数据，因而** ρ_k**，常用样本自相关函数** r_k **代替，从而得到样本偏自相关系数。原则上，** ϕ_{kj} **可由上述线性方程组直接求得，但有时直接计算比较复杂，通常采用下面的公式求得**

$$\phi_{kk} = \begin{cases} r_1 & k=1 \\ \dfrac{T_k - \sum_{j=1}^{k-1}\phi_{k-1;j}\cdot T_{k-j}}{1-\sum_{j=1}^{k-1}\phi_{k-1;j}\cdot r_j} & k=2,3,\cdots \end{cases} \qquad (7.17)$$

式中 $\phi_{kij}=\phi_{k-1ij}-\phi_{kk}\cdot\phi_{k-1ik-j}$ **j=1,2,…,k-1**

例 7.7 某时间序列的样本自相关系数 r_k **如表 7.1。试求偏自相关系数** ϕ_{33}**。**

表 7.11 序列的自相关系数

k	1	2	3	4	5	6	7	8
r_k	0.841	0.683	0.584	0.515	0.457	0.427	0.405	0.386

解：根据公式(7.17)可知

第 7 **期为中心的** 12 **个月移动平均数。**

① (k **为自协方差函数，即有**(k=E(Yt,Yt-k)。

$$\phi_{11} = r_1 = 0.841$$

$$\phi_{22} = \frac{r_2 - \phi_{11} \cdot r_1}{1 - \phi_{11} \cdot r_1} = \frac{0.683 - (0.841)^2}{1 - (0.841)^2}$$

$$= -0.083$$

$$\phi_{21} = \phi_{11} - \phi_{22} \cdot \phi_{11} = 0.841 - (-0.083)(0.841)$$

$$= 0.911$$

于是

$$\phi_{33} = \frac{r_3 - \phi_{21} \cdot r_2 - \phi_{22} \cdot r_1}{1 - \phi_{21} \cdot r_1 - \phi_{22} \cdot r_2}$$

$$= \frac{0.584 - (0.911)(0.683) + (0.083)(0.841)}{1 - (0.911)(0.841) - (0.083)(0.683)}$$

$$= 0.111$$

偏自相关系数的取值范围为[－1，＋1]，$|\phi_{kk}|$接近于 1，表明序列 Y_t 与 Y_{t-k} 之间的联系程度强，$|\phi_{kk}|$越接近于 0，表明序列 Y_t 与 Y_{t-k} 间的联系程度弱。如例 7.7 中，ϕ_{11}＝0.841，ϕ_{22}=-0.083,表明序列 Y_t 与 Y_{t-1},Y_t 与 Y_{t-2} 之间的自相关系数虽然相差不多，但它们之间的联系强弱程度却相差甚远。Y_t 与 Y_{t-1} 之间的联系大大地强于 Y_t 与 Y_{t-2} 之间的联系。

在现代时间序列分析方法(B-J 方法)中，已把偏自相关与自相关结合起来，分析判断时间序列适用于哪一类模型。

§7.3.3　时序特性的分析

时间序列的特性是指序列的平稳性、季节性和随机性。

1.平稳性。时间序列的统计特性不随时间的推移而发生改变，称其为平稳的时间序列，或说时间序列具有平稳性。平稳时序有两个特点：一是对于任意的时间 t，其均值恒为常数，即

$$EY_t=\mu$$

二是序列的自相关系数只与时间间隔有关，而与时间的起始点无关，即有

$$\gamma_k=E(Y_t-\mu)(Y_{t+k}-\mu)$$

$=E(Y_{t+\tau}-\mu)(Y_{t+\tau+k}-\mu)$上述两条中有一条不满足，时间序列就不具有平稳性[①]，而被称为非平稳的时间序列。

时间序列的平稳性可以通过自相关分析图得到检验。当时滞 k＝2(或 3)以后，序列的自相关系数迅速趋于 0，则序列具有平稳性。若在 k＝2(或 3)以后，序列的自相关系数还有较多的落在随机区间外，则序列不具有平稳性，为非平稳时序。

一般来说，非平稳的时间序列可以通过逐期差分的办法将其平稳化。经过差分处理能够平稳的时间序列被称为齐次非平稳序列。若记∇为差分算子，那么一阶逐期差分为

$$\nabla Y_t=Y_t-Y_{t-1}$$

二阶逐期差分为

[②] (k 为总体自相关函数，自机关函数与自协方差函数之间有关系式

[①] k　为自协方差函数。在时间序列中，　与(k 一样刻划了在不同时刻 Yt 与 Yt+k 间的相关程度。

$$\nabla^2 Y_t=\nabla(\Delta Y_t)=\nabla(Y_t-Y_{t-1})$$

$$=\nabla Y_t-\nabla Y_{t-1}=Y_{t-2}Y_{t-1}+Y_{t-2}$$

类推可以得到 d 阶逐期差分序列 Z_t 为

$$Z_t=\nabla^d Y_t \qquad t>d$$

若引进后移算 B，$BY_t=Y_{t-1}$，则 d 阶逐期序列 Z_t 与原序列 Y_t 可以写为

$$Z_t=\nabla^d Y_t=(1-B)^d Y_t \qquad t>d$$

序列经过差分后是否平稳，可以通过对差分后序列的自相关分析得到检验。

2.季节性。按月或按季记录的时间序列是否具有季节性，可以通过序列自相关系数的变动规律得到识别。图 7.7 是某市汗衫背心零售量的自相关图。从图 7.7 可以看出，在 $k=12$ 时，$r_{12}=0.7358$；当 $k=24$ 时，$r_{24}=0.4717$。这两个数值，显著地不为 0 且比附近时滞的自相关系数要大，表明序列存在周期为 12 个月的季节变动。图 7.7 显示的自相关系数呈现一种类似正弦波的变动，这很容易看出序列存在周期性变动，而无明显的趋势变动。若序列的趋势性很强，其是否具有季节性无法从序列的自相关图上立即看出，需要经过逐期差分处理，基本消除趋势性后。观察差分后序列的自相关系数才可判断。

时间序列的季节性能够通过季节差分(长期差分)加以消除。若季节周期为 12 个月，则一阶季节差分为

$$\nabla_{12} Y_t=Y_t-Y_{t-12}$$

二阶季节差分为

$$\begin{aligned}\nabla_{12}^2 Y_t &= \nabla_{12}(\nabla_{12} Y_t)\\ &= \nabla_{12}(Y_t - Y_{t-12})\\ &= \nabla_{12} Y_t - \nabla_{12} Y_{t-12}\\ &= Y_t - 2Y_{t-12} + Y_{t-24}\end{aligned}$$

可以类推得到季节周期长度为 s，经过 D 阶季节差分后的序列 W_t 与原序列 Y_t 的关系可表述为

$$W_t=\nabla_s^D Y_t \qquad t>Ds$$

采用后移算子 B^s，即 $B^s Y_t=Y_{t-s}$ 则上式可写为

$$W_t=\nabla_s^D Y_t=(1-B^s)^D Y_t$$

3.随机性。若时间序列完全是由数据随机构成的，则序列中的诸项之间是相互独立的，各时滞的自相关系数都接近于 0。这样的时间序列称之为完全的随机序列，或称时间序列具有随机性。利用序列的自相关分析图可以直观地判断序列是否具有随机性。若序列的自相关系数全部落入随机区间，并且没有某种变动的规律，则序列具有随机性。图 7.8 是某序列的自相关图，由图 7.8 可知，该序列具有随机性。

§7.3.4 ARMA 模型

ARMA(Autoregressive/moving average)模型即自回归移动平均模型，是一种用以描述时间序列的随机时序模型，适用于平稳的时间序列。它包括三类模型：自回归模型(AR)、移动平均模型(MA)、自回归一移动平均混合模型(ARMA)。

1.自回归模型。其模型形式为

$$Y_t=\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+\cdots+\phi_p Y_{t-p}+e_t \quad (7.18)$$

式中：$\phi_1,\phi_2,\cdots,\phi_p$是自回归系数；$e_t$是随机项或称误差项；P 是自回归阶数。若 P＝1，则 AR(1)模型或 ARMA(1，0)模型为

$$Y_t=\phi_1 Y_{t-1}+e_t$$

2.移动平均模型。其模型形式为

$$Y_t=e_t-\theta_1 e_{t-1}-\theta_q e_{t-q} \quad (7.19)$$

式中：是移动平均系数，q 是移动平均阶数。若 q=1，则 MA(1)或 ARMA(0,1)模型为

$$Y_t=e_t-\theta_1 e_{t-1}$$

3.自回归一移动平均混合模型。其模型形式为

$$\begin{aligned}Y_t=&\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+\cdots+\phi_p Y_{t-p}+e_t\\ &-\theta_1 e_{t-1}-\theta_2 e_{t-2}-\cdots-\theta_q e_{t-q}\end{aligned} \quad (7.20)$$

显然，(7.20)式是(7.18)与(7.19)式的合并。若 p=1,q=1,则 ARMA(1,1)模型为

$$Y_t=\phi_1 Y_{t-1}+e_t-\theta_1 e_{t-1}$$

上式可以写成

$$Y_t-\phi_1 Y_{t-1}=e_t-\theta_1 e_{t-1}$$

(7.20)式也可以写成

$$\begin{aligned}&Y_t-\phi_1 Y_{t-1}-\phi_2 Y_{t-2}-\cdots-\phi_p Y_{t-p}\\ &=e_t-\theta_1 e_{t-1}-\theta_2 e_{t-2}-\cdots-\theta_q e_{t-q}\end{aligned}$$

若 B 表示 k 步后移算子，并令

$$\phi(B)=1-\phi_1 B-\phi_2 B^2-\cdots-\phi_D B^D$$

$$\theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$$

那么，(7.20)式可以简写为

$$\phi(B)Y_t=\theta(B)e_t \quad (7.21)$$

4.改进的 ARMA 模型。实际的社会经济序列往往是不平稳的，有些还含有季节变动，无法直接用 ARMA 模型描述，但通过差分处理后，大多可以平稳。对这样的时间序列，可以用改进的 ARMA 模型加以描述。

对于无季节性变动的非平稳时序可应用 ARIMA(p，d，q)模型。其一般形式为

$$\phi(B)(1-B)^d Y_t=\theta(B)e_t \quad (7.22)$$

或

$$\phi(B)\nabla^d Y_t=\theta(B)e_t$$

若 p＝1，d＝1，q＝1，那么模型为

$$(1-\phi_1 B)(1-B)Y_t=(1-\theta_1 B)e_t$$

含有季节性的非平稳序列，除需进行逐期差分外，还需进行季节差分，这类序列可以用 ARIMA(p，d，q)(P，D，Q)^{s}模型，其一般形式为

$$\phi_p(B)\Phi_P(B^s)(1-B)Y_t=(1-\theta_1 B)e_t$$

式 中 ： P 是 季 节 自 回 归 阶 数 ， Q 是 季 节 移 动 平 均 阶 数 ，

$\Phi_p(B^s)=1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots-\Phi_p B^{ps}, \Theta_Q(B^s)=1-\Theta_1(B^s)-\Theta_2(B^{2s})$
$-\cdots-\Theta_Q(B^{QS})$分别为季节自回归算子和季节移动平均算子。

若 p=1,q=1,P=1,Q=1,d=1,D=1,S=4，模型为

$$(1-\phi_1 B)(1-\Phi_1 B^4)(1-B)(1-B^4)Y_t$$
$$=(1-\theta_1 B)(1-\Theta_1 B^4)e_t$$

§7.3.5　模型的识别

要为时间序列配合一个适合的模型，首先应进行模型的识别，也就是要先确定各个阶数 p、q、p、Q、d 和 D 的数值。

1.确定 d、D。时序若平稳且无季节变动，则 d＝0.D＝0。若需经过 d 阶逐期差分，D 阶季节差分，序列可以消除趋势和季节变动，那么，这两个阶数就是模型的 d、D。通常，d 和 D 的数值为 0、1、2。

2.选择 p，q。利用平稳序列的自相关和偏自相关函数的特征，可以识别模型的 p、q。

如图 7.9 所示，如果平稳(包括差分后平稳)序列的自相关函数呈现拖尾形式，即呈负指数衰减或正弦衰减波形，而偏自相关函数在 P 步后截尾，即当 k＞p 时，φ_{kk}≡0，那么序列可以建立 AR 模型，自回归阶数就是 p。如果平稳序列的偏自相关函数拖尾，呈负指数衰减或正弦衰减波形，而自相关函数在 q 步后截

图 7.9　自相关和偏自相关函数示意图

尾，即当 k＞q 时，r_k≡0，那么序列可以建立 MA 模型，q 为移动平均的阶数，如果序列的自相关和偏自相关函数都呈现拖尾形式，呈负指数衰减或正弦衰减波形，那么序列只能建立 ARMA 模型。其阶数的选择较为复杂，一般可以采用这样的准则：显著不为 0 的自相关系数的个数为移动平均阶数 q 的数值，自相关的摆动周期或具有统计有效性的偏自相关系数的个数为自回归阶数 P 的数值。实际应用中，往往会选出几组不同的 p,q 值。最后确定，还需待模型建立并进行检验之后完成。

3.选择 P、Q。含有季节性的序列，尽管差分后季节性基本消除，仍需考虑所建立的模型包含季节自回归和季节移动平均部分，故应选择 p,Q。其选择方法基本同 p、q，但在观察自相关和偏自相关函数时，只分析 k＝12，24，…或 k＝4，8，…时的情况。

§7.3.6　模型的检验

通过识别初选的模型在参数估计后，还需进行检验。随机时间序列分析要求残差序列 e，为自噪声。若经检验，残差序列确为白噪声，表明所建模型合理，用它来描述时间序列是适宜的；若残差序列不是白噪声，则所建模型不合理，它没有完全描述出时间序列的变化规律，应加以调整。模型检验通常可采用下述两种方法。

1.宜观判断。计算模型的残差序列的自相关系数，并绘制出自相关分析图。若自相关系数均落入随机区间，表明模型的残差序列相互之间独立，是白噪声序列；若自相关系数有较多的落人随机区间外，表明模型的残差序列还可能包含有某种模式，其不是白噪声序列。

2.x^2 检验。计算模型残差序列的 Q 统计量

$$Q = n\sum_{k=1}^{m} r_k^2(e) \tag{7.24}$$

式中：$r_k(e)$是残差序列的自相关系数，m 是残差序列自相关的最大时滞；n 是残差序列的数据个数。根据给定的显著性水平 a(a＝0.05)，自由度 m，(参数估计时采用最小二乘准则)或 m-P-q(采用极大似然准则)查找 x^2 分布表，得到标准值 x_a^2，若

$$Q \le x_a^2$$

则 x^2 检验通过，残差序列为白噪声；若

$$Q > x_a^2$$

x^2 检验未通过，不能认为残差序列是白噪声。

例 7.8　对我国工业总产值进行分析预测。

解：查找《中国统计月报》，可以搜集到 1983 年 1 月至 1988 年 12 月各月的工业总产值数值，采用 B—J 方法进行分析预测。

(1)模型识别。设 1983 年 1 月至 1988 年 12 月工业总产值数值为序列 Y，绘制其自相关图，如图 7.10。由图 7.10 可以看出，随着时滞 k 的增大，自相关系数逐渐减小，由正变负，在图中呈现为一条自右向左的斜线。这表明，序列随着时间间隔的加大，相关程度逐渐减弱，然后由正相关变为负相关。序列自相关系数的这种变化规律说明序列具有很强的趋势性。对序列进行一阶逐期差分后，观察差分后序列的自相关系数，在时滞 k＝12、24、36 时，自相关系数显著不为 0，故序列还存在季节性，季节周期为 12 个月。对一阶逐期差分后的序列再进行一阶季节差分，得到序列 W_u，绘制其自相关系数图，如图 7.11。从图 7.11 看，序列的趋势性和季节性基本消除。故可知，应确定 d＝1，D＝1。

绘制序列 W_t 的偏自相关图，如图 7.12。根据图 7.11 和图 7.12，选择非季节自回归阶数 p 和非季节移动平均阶数 q。从图中可以看出，显著不为 0 的偏自相关个数为 2 或 3，有统计有效性的自相关数目也为 2 或 3，故可以取 P=2 或 p=3，q＝2 或 q=3。

从图 7.11 和图 7.12 的自相关与偏自相关图中还可以看到当 k＝12 时，r_{12}＝－0.539，ϕ_{12}＝－0.262 均与 0 显著不同，故应考虑选取 P＝1，Q＝1。

经过初步识别，可以建立以下几个模型

Ⅰ:$ARIMA(2,1,2)(1,1,1)^{12}$

Ⅱ:$ARIMA(2,1,3)(1,1,1)^{12}$

Ⅲ:$ARIMA(3,1,2)(1,1,1)^{12}$

Ⅳ:$ARIMA(3,1,3)(1,1,1)^{12}$

(2)模型参数估计。运用电子计算机完成对上述四个模型的参数估计，根据结果可知，模型Ⅳ中有三个参数的 t 检验未通过。其余三个模型中，均只有一个参数的 t 检验未通过，三个模型的参数估计值分别为

Ⅰ	Ⅱ	Ⅲ
ϕ_1=0.7776	ϕ_1=0.5927	φ_1=0.6379
ϕ_2=-0.1323	ϕ_2=0.1565	φ_1=-0.6989
θ_1=1.4885	θ_1=1.4604	φ_1=-0.1392
θ_2=-0.9624	θ_2=-0.8240	θ_1=1.0728
Φ_1=-0.4692	θ_3=-0.2803	θ_2=-1.1506

Θ_1=0.6344　　　　Φ_1=-0.4438　　　　Φ_1=-0.4864

Θ_1=0.6355　　　　Θ_1=0.6578

在模型Ⅰ中，ϕ_2的 t 检验未通过，但 t 值很接近临界值；模型皿中的θ_3、模型III中的ϕ_3t 检验未通过，其 t 值均较小。

(3)模型检验。绘制三个模型的残差序列自相关图，只有模型Ⅰ的残差序列自相关系数全部落入随机区间，如图 7.13。

分别计算三个模型的 Q 统计量为

Ⅰ：Q_1=28.47　　II:Q_2=36.09　　III:Q_3=36.44

从计算结果可知，模型Ⅰ的 Q 统计量数值最小，这与自相关图显示的结果一致。以显著性水平 a＝0.05，查 x^2 分布表可知，三个模型的 x^2 检验均已通过。这种情况下，可选择模型Ⅰ进行预测。

(4)预测。当无法确定模型II、III是否应舍弃的时候，可以用三个模型分别进行试预测，通过对预测结果的分析比较，决定取舍。

利用上述三个模型分别对 1989 年 1 月至 1990 年 12 月各月工业总产值进行预测。从结果发现，模型II、III的偏差较大，不适宜

图 7.10　工业总产值序列的自相关分析图

图 7.11　序列 W^t 的自相关分析图

图 7.12　序列 W_t 的偏自相关图

图 7.13　模型Ⅰ残差序列的自相关分析图

作为预测结果，故舍弃这两个模型，只以模型Ⅰ的预测值作为预测结果，如表 7.12(表中附有 95%的置信区间)。

7.4 指 数

指数是用以测定一群变量在时间或空间上综合变动的一种相对数。例如，在人们的生活中，每天都要接触到许多商品的价格。这些价格中，有些是上升的，有些是下降的。我们想用一个数值，来概括地反映这些商品价格变动的幅度，这时，我们就要计算价格指数。当然，利用指数这种方法，不仅可以测定价格的变动，也可以测定生产发展、工资变动、劳动生产率提高等。但由于价格指数的编制方法具有一般方法论的特点，我们在此主要讨论价格指数的编制方法。

§7.4.1 简单指数

简单指数就是不加权的指数。一般地说，它可以有六种计算方法，即除了计算简单综合指数外，对应于算术平均数、倒数平均数，几何平均数、众数和中位数，还可计算 5 种平均数指数。在这里，我们讨论三种简单指数。

1.简单综合指数。它的公式为

$$I=\frac{\Sigma p_1}{\Sigma p_0}$$

设有如下资料(见表 7.13)。

表 7.13　　某地粮食价格情况　　（单位：元/斤）

粮食品种	1977 年	1978 年
小麦	0.20	0.23
玉米	0.20	0.24
高梁	0.15	0.16
合计	0.60	0.68

简单综合指数的计算方法,是从 1987 年全部商品的单价合计数除以 1977 年的单价合计数。即

$$I=\frac{\Sigma p_{87}}{\Sigma p_{77}}=\frac{0.68}{0.60}=105\%$$

从某一种商品看，有的商品涨价了，有的商品降价了。但从整体上看，价格水平提高了，1987 年比 1977 年提高了 5%。

简单综合指数有它的弊病。上例中，如果说高粱的单位不用斤，而用公斤，那么我们就可能得到不同的计算结果

$$I=\frac{0.23+0.24+0.32}{0.20+0.25+0.30}=1.053$$

仅仅改换了一种商品的计量单位，计算结果就发生了变化，而计量单位本身是可以任意规定的。这样，用同一个资料，就可能计算出无数个指数数值。这种弊病，只有通过计算加权指数才能克服。

2.平均数指数。简单算术平均数指数的公式为

$$I=\frac{\Sigma\frac{p_1}{p_0}}{n}$$

利用表 7.13 的资料可以计算出

$$I=\frac{\Sigma\frac{p_{87}}{p_{77}}}{3}=\frac{\frac{0.23}{0.20}+\frac{0.24}{0.25}+\frac{0.16}{0.15}}{3}$$

$$=\frac{1.15+0.96+1.07}{3}=1.06$$

简单几何平均数指数公式为

$$I\sqrt[n]{\Pi\frac{p_1}{p_2}}$$

利用表 7.13 的资料可以计算出

$$I=\sqrt[n]{\prod_{i=1}^{3}\frac{p_{87}}{p_{77}}}=\sqrt[3]{1.15\times0.96\times1.07}=1.057$$

这两种方法比简单综合法进了一步，不再受商品计量单位的影响，但仍有不足之处，就是把三种商品的重要性同等看待了。这也要靠加权指数来克服。目前世界各国使用较多的简单指数是简单几何平均数指数。

§7.4.2　加权指数

1.加权综合指数。某种商品对于消费者来说，它的重要性究竟有多大，这可以根据该种商品的购买数量或付出的金额的大小来确定。假设某地的粮食销售量资料如表 7.14。

表 7.14　　　　　某地 1977 年和 1978 年粮食销售情况

	1977 年		1987 年	
	价格(元/斤)	销售量(万斤)	价格(元/斤)	销售量(万斤)
小麦	0.20	500	0.23	550
大米	0.25	200	0.24	150
高粱	0.15	50	0.16	60

显然，在这个地区三种商品对于人们生活的重要性是不同的。小麦最为重要，大米次之，高粱又次之。而计算价格指数，就必须把这个因素考虑进去，用各种商品的价格分别乘上它们的数量，然后加以汇总，进行比较。这里，商品的销售量起着权衡轻重的作用，所以把它们称为权数。

为了说明各种商品价格的变动，权数必须使用同一时期的，也就是说，两个时期的总销售额是按同一个时期的销售量计算的，即

$$I=\frac{\Sigma p_1 q}{\Sigma p_0 q}$$

式中：下标 1 代表计算期，下标 0 代表基期；p 代表价格；q 代表销售量。

价格变了，但使用的销售量不变，这样，价格指数所反映的，就只是商品价格的变动了。那么，作为权数的销售量应该固定在什么时期呢？一般地说，可以有两种作法。

一种作法是把权数固定在基期，即

$$I=\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

这个公式称为拉斯佩雷斯公式(简称为拉氏公式)。用这个公式计算的价

格指数为

$$I=\frac{\Sigma p_1q_0}{\Sigma p_0q_0}=\frac{0.23\times500+0.24\times200+0.16\times50}{0.20\times500+0.25\times200+0.15\times50}$$

$$=\frac{115+48+8}{110+50+7.5}=\frac{171}{157.5}=108.57\%$$

这说明三种商品的价格比基期提高了 8.57%。

另一种作法是把权数固定在计算期，指数公式为

$$I=\frac{\Sigma p_1q_1}{\Sigma p_0q_1}$$

这个公式被称为派煦公式(简称派氏公式)。用这个公式计算价格指数为

$$I=\frac{\Sigma p_1q_1}{\Sigma p_0q_1}=\frac{0.23\times550+0.24\times150+0.16\times50}{0.20\times550+0.25\times150+0.15\times50}$$

$$=\frac{126.5+36+8}{110+37.5+7.5}=\frac{170.5}{155}=110\%$$

这说明三种产品价格比基期提高了 10%。

以上两种计算方法所得的结果略有差别。这两个公式的提倡者各有自己的道理，但在实际应用中，拉氏公式用得更普遍一些。因为在编制长期连续性的价格指数统计数列时，拉氏公式比派氏公式方便。拉氏公式的权数固定在基期，从而可以更好地综合反映长期的连续性的价格变动。另外，由于拉氏公式可以利用过去的资料，计算指数的时效性也比较强。

2.算术加权平均数指数。加权指数也可以用平均数的方法来编制，使用最多的平均数指数公式是基期加权算术平均数指数公式，即

$$I=\frac{\Sigma\frac{p_1}{p_0}\times p_0q_0}{\Sigma p_0q_0}$$

从形式上看，这个公式在约分化简之后就成了拉氏公式，即

$$I=\frac{\Sigma\frac{p_1}{p_0}\times p_0q_0}{\Sigma p_0q_0}=\frac{\Sigma p_1q_0}{\Sigma p_0q_0}$$

从指数公式的历史发展来看，它确实也是从拉氏公式变形而来的。但是，在用于编制具体的经济指数时，这二者却有重要区别。作为价格指数的拉氏公式，物量项目和价格项目是相适应的，有三种商品的价格，也相应地要有(也只能有)三种商品的物量。然而，作为价格指数的基期加权算术平均数指数公式，其中的 $\frac{p_1}{p_0}$ 只为某一种代表商品的个体价格(如某号大米)指数，而 p_0q_0 则为代表商品所代表的那个组(如粮食组)的全部销售额。那么为什么要用代表商品的个体指数来代表粮食组的指数呢？这是因为，粮食类商品的品种、种类很多，要得到每一种商品的价格和物量的资料费时费力，有时甚至根本办不到，其他商品的情况就更是这样了。因此，人们常常从成千上万种商品中选出几十种成几百种代表商品，计算这些商品的个体指数，用它们的价格变动来反映整个组的价格变动，然后再利用各组的销售额对它们加权平均，得出总的价格指数。实践证明，只要代表商品选得好，权数确定得当，用这种方法编制的指数能很好地反映商品价格的变动。目前。世界各国编制大范

围的价格指数(其他指数也是如此)，大都采用这种方法。

§7.4.3　指数体系

社会经济现象间常常是相互联系的，而这种联系有时可以表示为相乘关系，如

商品销售额＝价格×销售量

产品成本＝单位成本×产量

在这种情况下，就可以用指数的方法在联系中研究它们的动态变化，如

商品销售额指数＝价格指数×销售量指数

而这三个有联系的指数，就形成了一个指数体系。这里的销售额指数，其实是一个一般的发展速度，但为了研究的方便，也把它作为指数处理。

显然，由于作为权数的因素所选择的时期不同，可以有以下两套指数体系

1.按派氏公式计算的指数体系。

$$\frac{\Sigma p_1 q_1}{\Sigma p_1 q_1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0}$$

销售额　价格指数　销售量指数
指数　(派氏公式) (拉氏公式)

2.按拉氏公式计算的指数体系。

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0}$$

销售额　价格指数　销售量指数
指数　(拉氏公式) (派氏公式)

在上面两个体系中，已知三个指数中的两个指数，就可以推知第三个指数。值得注意的是，如果价格指数采用拉氏公式，销售量指数就要采用派氏公式，这样，指数体系在数学上才能保持完整。在某些情况下，价格指数(或物量指数)不是采用综合指 数公式计算的，在这种情况下，仍可以根据同样的原理，从已知 的两个指数来近似地估算第三个指数。

习　题

1.某市 1981 年社会商品零售额为 68 亿元，比 1978 年增长 54.6%，其中用的商品零售额占社会商品零售额的比重由 28.2% 上升到 32.2%，试分析 1978—1981 年间用的商品零售额年平均递 增多少。

2.某外贸冷库 1982—1985 年间备月冷冻兔肉产量如表 1。试 分析研究这一时期冷冻兔肉产量的变化规律。

3.我国 1975—1985 年水泥产量如表 2，试分析其变化趋势并建立一合适的方程。

4.某市几种主要副食品调整价格前后资料如表 3。(1)试接拉氏公式和派氏公式分别计算价格指数；(2)试根据指数体系的管理推算销售量指数，并说明它所采用的公式。

5.某市副食品调价前后肉禽蛋价格的权数资料如农 4。计算肉禽蛋类零售物价指数。

第 8 章　非参数统计

§8.1　引 言

§8.1.1　什么是非参数统计

第 4 章曾介绍了参数的估计与假设检验，其内容是在已知总体分布的条件下对一些主要参数(如均值、方差)进行估计和假设检验。在估计和检验时一般都假设总体是服从正态分布，方差是相等的等等。但是在统计分析中许多实际问题要求解决的并不都是这一类问题。例如，从总体中获取一个样本，想了解其总体是否服从某种分布，或者某种现象的出现是否是随机的，或者两种及两种以上的现象之间是否有联系、联系的紧密程度如何等等。有时虽然也是要求检验有关总体参数，但是客观情况并不符合参数假设检验的条件。还有一些客观的现象是用比较低的计量水准，如用列名或顺序的计量使得无法用通常的参数方法来进行。因此，对于上述不是参数的估计和检验问题以及不是建立在总体分布服从一定假设基础上的有关方法，习惯上都称作非参数统计，其内容很多，应用也十分广泛。

§8.1.2　非参数统计方法的优缺点

非参数统计方法与传统的参数估计与检验相比，有以下优点：

(1)由于非参数方法要求的假设条件比较少，因而它的适用范围比较广泛；

(2)有些非参数方法的运算比较简单，可以较快地取得结果，因此比较节省时间，尤其是当要迅速取得结果和手头又没有高功能计算工具的情况下比较适用；

(3)这些方法在直观上比较容易理解，并不需要太多的数学和统计理论；

(4)适用一些比较低的计量水准。

但是非参数方法也有以下缺点：

(1)由于方法简单，用的计景水准比较低，因此，如果能与参数检验的方法同时使用时，其方法不如参数方法敏感，检验的功效要差一些，这是由于使用较低的计量水准而失去了一部分信息的缘故。

(2)虽然不需要用较深的数学理论，但是有些算术运算也是相当麻烦的。

§8.2　c^2(卡方)检验

在非参数统计中应用最广泛的方法之一是利用χ^2分布进行独立性、一致性和吻合性的检验。

§8.2.1　c^2分布的数学性质

我们在第 4 章抽样分布一节中曾介绍过χ^2分布，它也是一个随机变量的理论分布。当随机变量 X_i，服从正态分布，其均值为μ,标准差为σ时，经过标准化$Z_i = \frac{X_i - \mu}{\sigma}$，则 Z_i 为标准化正态分布，即 Z_i～N(0,1)。于是$Z_i^2 = (\frac{X_i - \mu}{\sigma})^2$就服从$\chi^2$分布。如果从正态分布的随机变量中独立随机地抽取两个值 X_1，和 X_2，分别进行标准化$Z_1 = \frac{X_1 - \mu}{\sigma}$，$Z_2 = \frac{X_2 - \mu}{\sigma}$然后计算其平方和，$Z_1^2 = Z_2^2$也同样是一个$\chi^2$分布。推广到一般，n 个独立标准正态分布变量的平方和$Z_1^2 + Z_2^2 + \cdots\cdots Z_n^2$是自由度为 n 的$\chi^2$分布。由此可以看出$\chi^2$分布的值始终是正的，其分布的形状随着自由度的变化而不同。在自由度较小时是偏斜的，随着自由度的增加，根据中心极限定理的一个结果，χ^2分布也接近正态分布(见图 8.1)。

图 8.1　不同自由度的χ^2分布

χ^2分布的均值等于其自由度，它的方差二倍于χ^2分布的自由度。

在实践中经常要对一些观察值出现的实际次数与理论次数进行比较，以了解实际发生的结果与理论之间是否一致。设观察的次数为 f_0，理论的次数为 f_e，则可定义统计量χ^2

$$x^2 = \sum \frac{(f_0 - f_e)^2}{f_e} \tag{8.1}$$

数理统计证明了在大量的试验中，若实际频数和理论频数相一致时，χ^2就服从χ^2分布，用一个简单的例子来说，如果随机掷一个硬币以观察其正面向上还是反面向上，共作了 100 次试验，出现正面 46 次，反面 54 次。假如这枚硬币是均匀的，则从理论上讲其出现正面和反面的概率是相同的，应为各 50 次，因此可以计算统计量χ^2

$$x^2 = \sum \frac{(f_0 - f_e)^2}{f_e} = \frac{(46-50)^2}{50} + \frac{(54-50)^2}{50} = 0.64$$

在假设该硬币是均匀的前提下，显然 f_0-f_e 之间的差距应该是比较小的，从而χ^2值也比较小，χ^2在理论上应该是服从χ^2分布，这里自由度为 1，这是由于观察结果出现正面或反面之间是不独立的，当 100 次试验中出现正面 46 次，其反面必然是 100－46＝54 次，所以只有一个频数是自由的。如果观察的结果和理论的假设不一致，其离差就会比较大，从而χ^2值也大。当χ^2值大到按χ^2分布出现的概率很小时，就可以判断实际观察结果和理论的假设是不一致的，χ^2检验就是利用这一原理。

§8.2.2　c^2的独立性检验

在研究问题时经常会遇到要求判断两个变量之间是否有联系的问题。例

如，社会学家想知道目前受教育程度和收入高低之间是否有联系；教育工作者欲了解学生在校的考试成绩和工作后的表现之间是否有联系等。如果两个变量之间没有联系则称作是独立的。其检验的方法就是将研究的对象按两个变量分别进行分类,编制成一张交错分类的表,通常叫做列联表(contingency table)其格式如表 8.1。

表 8.1　　　　两向分类的列联表

		第二个变量分类				合计
		1	2	……	c	
第一个变量分类	1	n_{11}	n_{12}		n_{1c}	$n_{1\cdot}$
	2	n_{21}	n_{22}		n_{2c}	$n_{2\cdot}$
	M					
	r	n_{r1}	n_{r2}		n_{rc}	$n_{r\cdot}$
合　计		$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot c}$	n

其中 n 为观察对象的总数，n_{ij} 表示按第一变量属于 i 类，第二个变量属于 j 类的观察对象数，为了区别观察值和理论值，将它们分别用 O_{ij} 和 E_{ij} 表示。

假设检验的过程和前面参放假设检验一样，可建立假设

H_0：两个分类的变量之间是独立的

H_1：两个分类的变量之间是不独立的

检验的结果若是接受 H_0，就说明不能推翻两个分类的变量是独立的假设：反之，拒绝 H_0 接受 H_1 就说明它们之间是不独立的。

检验的统计量就是χ^2

$$x^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

式中，O_{ij} 就是实际的观察结果，即 $O_{ij}=n_{ij}$。关于理论值的计算就是利用概率论中概率独立的基本规则：如果两个事件是独立的，则它们的联合概率等于它们分别概率的乘积,即落入第 ij 格的概率等于落入 i 行的概率乘以落入 j 列的概率。

$$P(\text{落入ij格概率}) = (\frac{n_{i\cdot}}{n})(\frac{n_{\cdot j}}{n})$$

由于总的试验次数为 n，所以

$$E_{ij} = n(\frac{n_{i\cdot}}{n})(\frac{n_{\cdot j}}{n}) = \frac{n_{i\cdot}\, n_{\cdot j}}{n} \qquad (8.2)$$

这时χ^2 分布的自由度为(r－1)(c－1)，当取显著性水平为 a 时，就可得到 x^2_{1-a}[(r-1)(c-1)]的临界值，当 $x^2 > x^2_{1-a}$ 时就拒绝 H_0 接受 H_1。

例 8.1 若希望知道某市居民中文化水平与年收入之间是否有联系，从研究对象中抽取了 2764 人进行观察，分别按文化水平与年收入进行两向分类，文化水平分成 3 类，年收入分成 5 组，见表 8.2。要求检验这二者之间是否有联系。

表 8.2　　文化水平与收入水平列联表

		文化水平			合计
		大学及以上	中学	小学及以下	
收入水平	1500 以下	186	38	35	259
	1500-2000	227	54	45	326
	2000-2500	219	78	78	375
	2500-3000	355	112	140	607
	3000 以上	653	285	259	1197
合计		1640	567	557	2764

解：(1)建立假设

H_0：文化水平与收入水平之间是独立的

H_1：两个变量之间不独立(即有联系)

(2)计算统计量χ^2。在计算χ^2之前，首先要算出每一格的理论值 E_{ij}，利用公式(8.2)，有

$$E_{11}=\frac{(259)(1640)}{2764}=153.68$$

$$E_{12}=\frac{(259)(567)}{2764}=53.13$$，依次类推其结果见表8.3

$$\chi^2=\frac{(186-153.68)^2}{153.68}+\frac{(227-193.43)^2}{193.43}+\cdots+\frac{(259-241.22)^2}{241.22}=47.9$$

若规定显著性水平 a＝0.01，本例的自由度为 8.得 $x^2_{0.99}(8)=20.090$，x^2＞$x^2_{0.99}$。因此结论是拒绝 H_0 接受 H_1，即两个变量之间是不独立的。换句话说，文化水平与收入水平之间是有联系的。

表 8.3　　观察值与理论值比较（括号内为理论值）

	大学及以上	中学	小学及以下	合计
1500 以下	186(153.68)	38(53.13)	35(52.19)	259
1500-2000	227(193.43)	54(66.87)	45(65.70)	326
2000-2500	219(222.50)	78(76.93)	78(75.57)	375
2500-3000	355(360.16)	112(124.52)	140(122.32)	607
3000 以上	653(710.23)	285(245.55)	259(241.22)	1197
合计	1640	567	557	2746

在利用 x^2 检验时应注意：(1)要求试验的次数比较多；(2)每一格中期望的理论次数不能太小，虽然在统计学界没有统一结论，但科克伦提出了一个参考的准则，即要求 n＞20，且理论值小于 5 的格不能超过 20%。在必要时可将邻近的格加以合并。

当两个变量的分类都分成两类时，就形成 2×2 列联表，如表 8.4，这在日常研究工作中是经常遇到的，现用 a、b、c、d 分别表示其观察值。

表 8.4　　　　　　　　2 × 2 列联表

	1	2	合　计
1	a	b	a+b
2	c	d	c+d
合　计	a+c	b+d	n

这时χ^2值有一个简便的计算公式

$$\chi^2=\frac{n(ad-bc)^2}{(a+c)(b+d)(c+d)(a+b)} \tag{8.3}$$

例 8.2 有人研究抽烟与肺癌之间是否有联系。在医院门诊部搜集了 56 人的资料，整理如表 8.5。

表 8.5　　　　关于抽烟与肺癌的有关资料　　　　（单位：人）

	抽　烟	不抽烟	合　计
有 肺 癌	20	16	36
没有肺癌	6	14	20
合　计	26	30	56

试检验抽烟与肺癌之间是否有联系。

解：建立假设

H_0：抽烟与肺癌之间没有联系

H_1：抽烟与肺癌之间有联系

计算统计量　$\chi^2=\frac{56[(20)(14)-(16)(16)]^2}{(26)(30)(20)(36)}=3.376$

a=0.05 时 $x^2_{0.95}$(1)=3.841，a=0.10 时 $x^2_{0.9}$(1)=2.706

因此在 0.05 显著性水平下接受 H_0，即不能以 95%的可靠性推翻 H_0，但在 0.1 的显著性水平下拒绝 H_0。说明二者之间还是有较弱的联系。

由于χ^2分布是一个连续的变量，而在 2×2 的列联表中用的是离散的方法，因此那茨于 1934 年提出了一个修正的公式为

$$\chi_c^2=\frac{n(|ad-bc|-0.5n)^2}{(a+c)(b+d)(c+d)(a+d)} \tag{8.4}$$

根据例 8.2 的列联表计算得

$\chi_c^2=\frac{56[|(20)(14)-(16)(6)|-0.5(56)]^2}{(26)(30)(20)(36)}=2.427$，其结论仍与前面相同，近年来也有人对耶茨的修正系数提出了不同的看法。

§8.2.3　c^2 的一致性检验

在用参数方法检验两个比例是否相等时，曾讲过用正态分布的近似方法来检验$H_0:P_1=P_2$，$H_1:P_1\neq P_2$在这里也可以将从每一个总体中抽取的样本分成两类，把结果写成 2×2 的列联表，然后用χ^2 检验，这种检验称作χ^2 的一致性检验。从表面上看来，列联表的形式以及计算χ^2值的公式与独立性的检验是相同的，但是其抽样方法和对期望次数的计算规则是不同的。在独立性的检验时，是从研究的总体中抽取一个容量为 n 的样本，然后根据样本的观察值

进行两向分类：一致性检验则是从比较的总体中分别抽取独立的随机样本，然后把抽到的单位划分为两类中的一个类别。关于理论期望频数的计算，在独立性检验时，假设按两个标准的分类是独立的，因而两个水平的联合概率是两个单独概率的乘积，而在一致性检验时，假定比较的几个总体中具有某种特征的单位数的比例相同，当假设成立时，各类的期望频数应该根据各样本的总数的比例来计算。

例 8.3　某厂生产一种新型自行车，希望了解消费者喜欢这种型号的人数比例，分别从青年人和 40 岁以上的人中各抽 100 人进行调查。调查结果青年人中有 80 人表示喜欢，40 岁以上的人中有 70 人表示喜欢，试以 0.05 的显著性水平检验这两组人喜欢这种新型车的比例是否一致。

解：

H_0：青年人与 40 岁以上的人喜爱的比例是一致的

H_1：喜爱的比例是不一致的

将以上调查结果写成列联表(见表 8.6)。

计算χ^2值：$\chi^2=\dfrac{200(80\times30-20\times70)^2}{(150)(50)(100)(100)}=2.667$

$\chi^2_{0.95}$(1)，因此接受 H_0，即用 0.05 的显著性水平不能说两组消费者对这种新型车的爱好有显著差别。

表 8.6　　两组消费者对新型自行车喜爱的人数　　（单位：人）

	喜爱	不喜爱	合计
青年人	80	20	100
40 岁以上	70	30	100
合　计	150	50	200

用下态分布近似的方法检验两个比例是否有显著差别可用 Z 检验。

$$Z=\sqrt{\frac{P_1-P_2}{\frac{\overline{P}(1-\overline{P})}{n_1}+\frac{\overline{P}(1-\overline{P})}{n_2}}}=\sqrt{\frac{0.8-0.7}{\frac{(0.75)(0.25)}{100}+\frac{(0.75)(0.25)}{100}}}=1.633$$

$Z_{1\text{-}\bar{a}\text{-}0.95}=1.96$由于$1.633<1.96$所以结论与 χ^{2}检验相同，而且$Z^2=1.633^2=2.667=\chi^2, Z^2_{1-\bar{a}}=$。$\chi^2_{1-\bar{a}}=3.841$前面讲到$\chi^2$的数学性质，从这个例子可以得到验证。

利用 χ^{2}进行一致性检验还可以推广到几个总体的比例是否一致,分类也可以推广到两类以上，若是 r 个总体，每个总体分成 c 类，就形成 r×c 列联表。

例 8.4　某市电视台调查本市市区、近郊以及远郊的居民对电视节目的意见是否一致，分别调查了 40 个居民，其结果如表 8.7。试以 0.05 的显著性水平检验市区、近郊及远郊居民对电视节目的意见是否一致。

解：

H_0：这三组居民对电视节目的意见是一致的

H_1：这三组居民对电视节目的意见不一致

计算检验统计量，首先要计算各格的理论期望值，当 H_0 为真时，各组满意的人数占总人效的比例应为 $\frac{72}{120}$，因此各组理论上满意的人数应为 40×6=24 人，比较满意的理论人数应为 40× $\frac{26}{120}$ =8.6667 人，依次类推。对电视节目的意见没有显著差别。

表 8.7　　关于某市居民对电视节目意见的调查表　　（单位：人）

	满意	比较满意	不太满意	不满意	合计
市区	31	5	2	2	40
近郊	21	10	4	5	40
远郊	20	11	7	2	40
合计	72	26	13	9	120

$$\chi^2=\frac{(31-24)^2}{24}=\frac{(5-8.6667)^2}{8.6667}+\cdots\cdots\frac{(2-3)^2}{3}$$
$$=10.391$$

$\chi_{0.95}(6)=12.592$，$\chi^2<\chi_{0.95}(6)$ 不能拒绝 H_0，即这三组居民对电视节目的意见没有显著差别。

§8.2.4　χ^2 的吻合性检验

在实际的研究工作中有时需要对变量是否遵从某一理论分布进行检验，χ^2 分布用于这一方面的检验叫做吻合性检验，也称拟合优度的检验。

这类检验要求所抽取的样本是随机样本，变量的计量水准至少是列名的。若被检验的总体共真实的分布函数为 F(χ)，但它是未知的，只能从这一分布中抽取一个随机样本，要求通过这一样本检验这一总体的分布是否与规定的理论分布 F·(χ)相一致。因此，

H_0：F(x)=F·(x)

H_1：F(x)≠F·(x)

检验的统计量 χ^2 计算方法与前面相同，其理论期望次数 $E_j=nP_j$，其中 P_j 为按理论分布计算的概率。

例 8.5　有一个提供随机数字的计算机程序连续输出了 300 个随机数字，现希望利用统计方法检验这一程序出现的数字是否是随机的。某技术员提供了一个检验方案，即将这 300 个数字按出现的顺序划分为 30 组，每组有 10 个数字，在每一组中统计其中是偶数的个数，其结果如下

偶数个数	0	1	2	3	4	5	6	7	8	9	10
组　　数	1	4	1	1	3	14	5	1	6	1	0

如果出现的数字是随机的，则每一次可以看作是一次独立试验，每次试验出现偶数的概率应为 0.5，则每一组中出现偶数的次数理应服从二项式分布，因此这个问题就转化为在各组中出现的偶数是否服从二项式分布。

H_0：出现的数字是随机的(各组偶数服从二项式分布)

H_1：出现的数字不是随机的(各组偶数不服从二项式分布)

各组中出现偶数 j 个的概率为 P_j

$$P_j = \binom{10}{j}\left(\frac{1}{2}\right)^{10}$$

所以在这 30 个数字组中出现 j 个偶数的期望数为：E_j =30×(P_j)，将实际的观察值与理论期望比较如下

	0	1	2	3	4	5
观察值 O_j	0	4	1	1	3	14
理论值 E_j	0.03	0.291	1.32	3.516	6.153	7.380
	6	7	8	9	10	
观察值 O_j	5	1	0	1	0	
理论值 E_j	6.153	3.516	1.32	0.291	0.03	

根据前述 χ^2 检验的限制条件，即理论值小于 5 的格不能超过 20%，需要将一些邻近的格适当合并

	j≤3	j=4	j=5	j=6	j≥7
0_j	6	3	14	5	2
E_j	5.157	6.153	7.380	6.153	5.157

然后计算 χ^2 值

$$\chi^2 = \frac{(6-5.157)^2}{5.157} + \frac{(3-6.153)^2}{6.153} + \frac{(14-7.380)^2}{7.380} + \frac{(5-6.153)^2}{6.153} + \frac{(2-5.157)^2}{5.157} = 9.84$$

这里自由度为 4 对照 χ^2 的临界值显著性水平 α＝0.04 左右，当 α=0.05 时，$\chi^2 > \chi^2_{0.95}(4)$。看来这计算机程序出现的数字并不是随机的。

拟合优度检验时，有时理论分布的参数是未知的，这时也要从样本来估计总体的参数值。在这种情况下每计算一个参数值就要减少一个自由度。因此若估计了 k 个参数，其自由度为 c-1-k。拟合优度的检验也广泛应用于科学研究中对一些理论的检验，例如孟德尔在遗传学研究的一次试验中根据豌豆杂交的理论得出杂交豌豆中圆形黄色、圆形绿色、尖形黄色和尖形绿色的比例应为 9：3：3：1，实际观察结果是，在 556 粒豌豆中分别为 315，108，101 和 32.根据上述理论比例计算应分别为 312.75，104.25，104.25 和 34.75，计算 χ^2 值为 0.47 小于预期的 χ_{1-a} 值。因此说明试验结果与理论相一致。

§8.3 正负符号检验

在前面曾经提到，有些参数的假设检验需要有一定的假设条件，当这些条件不能满足时就要用不需要严格假设的方法。正负符号检验是一种比较简单的方法，正负符号检验既可用于单一样本，也可用于两个样本的比较检验：既可用于独立样本的检验，也可用于两个有联系的样本的检验。现分别加以介绍。

§8.3.1 单一样本中位数的符号检验

反映一个总体分布位置的参数主要有平均数和中位数。平均数反映的是分布重心的位置，而中位数是在该数上下出现的概率均为 1/2 处。当分布为对称时这二者是一致的，分布为不对称时二者就有差别。在很多场合需对中位数的位置进行检验，可采用正负符号检验。

例 8.6 某钢铁厂生产的钢材在正常情况下中位数的长度为 10 米。现随机地从生产线上抽取 10 根，测得长度(单位：米)如下

9.8,10.1,9.7,9.9,10,10,9.8,9.7,9.8,9.9

试问生产过程中对长度的控制是否需要适当调整。

解：这里中位数是需要研究的参数，如果在生产过程中钢材的长度在 10 米上下各占一半就不需要调整生产过程。反之，若多数过长或多数过短均需加以调整。

H_0：中位数(M)＝10　　　H_1：中位数(M)≠10

在进行正负符号检验时可以将样本中每根的长度减去中位数，如果大于中位数就为正号(+)，小于中位数为负号(-)，现计算如表 8.8。从表 8.8 可以看出 10 个样本单位中除有 2 个与中位数相同外，余下的 8 个中有 1 个是正号，7 个是负号。如果进一步用精确的测量器进行测量，则与中位数相同的 2 个单位也可以区分为正号或负号，现假定为一个正号和一个负号，这样 10 个样本中就有 2 个正号和 8 个负号。如果总体的中位数为 10，那么理论上出现正号和负号的应该各占一半。现在问出现 2 个及 2 个以下正号的概率有多少，可以用二项式分布 P ＝ 0.5 来计算

表 8.8　　　　中位数符号检验计算表

长度 x	x-10	符　号
9.8	-0.2	-
10.1	0.1	+
9.7	-0.3	-
9.9	-0.1	-
10	0	
10	0	
9.8	-0.2	-
9.7	-0.3	-
9.8	-0.2	-
9.9	-0.1	-

$$P(X \le 2)=\sum_{x=0}^{2}\binom{10}{X}0.5^{10}=0.0547$$

由于 H_1 是一个双尾检验，因此也应包括负号在 2 个及 2 个以下的概率，因此 P＝2×(0.0547)＝0.1094 这就是说当中位数为 10 时，出现上述结果的概率为 0.1094，当 α=0.05 时尚不能拒绝 H_0。决策人员可以根据这一情况，结合其他因素作出是否需要调整生产过程的决策。

在大样本的情况下，用二项式分布计算概率比较麻烦，也可以用正态近似计算

$$Z_+=\frac{S_+ - 0.5n - 0.5}{0.5\sqrt{n}}，Z- = \frac{S_- - 0.5n - 0.5}{0.5\sqrt{n}} \qquad (8.5)$$

其中 S+代表正号的数目，0.5n 表示在 P=0.5 条件下正号或负号的平均数目(理论数目)，0.5 称作连续的修正项，分母 $0.5\sqrt{n}$ 为 P＝0.5，样本容量为 n 时的标准差。当 $|Z| \ge Z_{1-aZ2}$ 时拒绝 H0。

§8.3.2　两个独立样本的符号检验

假设 X_1，$X_2 \cdots X_{n1}$ 和 Y_1，Y_2，$\cdots Y_{n2}$，为抽自两个总体的独立样本，X 和 Y 的计量水准至少是顺序的，研究的变量为一连续的变量，要求检验这两个总体的中位数是否相同，假设的形式是

H_0：两个总体的中位数相同

H_1：两个总体的中位数不同(X 的总体大于 Y 的总体或者相反)

检验的方法是将两个样本统一按照顺序排列，找出中位数。从直观看，如果 H_0 为真，则在样本 X 和 Y 中，超过中位数和低于中位数的数目应该接近相等，也就是说其正、负符号各占一半左右。因此可以把两个样本分别按照超过中位数的数目和低于中位数的数目整理成列联表的形式，应用 χ^2 检验或正态近似检验。

例 8.7　现有两个独立样本的观察值如下

样本 1：10,10,10,12,15,17,17,19,20,22,25,26

样本 2：6,7,8,8,12,16,19,22

要求检验这两个总体的中位数是否相同。

解：先将两组样本的观察值统一按顺序排列，找出其中位数为 16，然后将每个观察值和它比较，大于中位数用正号表示，小于中位数用负号表示，我们得到

样本 1：-　-　-　-　-　+　+　+　+　+　+　+

样本 2：-　-　-　-　-　-　+　+　+

以上资料可以写成 2×2 列联表的形式，如表 8.9。

表 8.9　　　　2 × 2 列联表

	+	—	
样本 1	7	5	12
样本 2	3	6	9
合　计	10	11	21

计算 $\chi^2=\frac{21(7\times6-15)^2}{10\times11\times9\times12}=1.29$, 若 $a=0.05, \chi_{0.95}(1)=3.84 \chi^2<\chi^2_{1-a}$，因此接受

H_0，即不能说两个总体的中位数有显著差别。

§8.3.3 两个有联系样本的符号检验

关于两个有联系的样本的数据有各种各样的情况，最常见的是对样本中同一个单位经过某种处理前后的比较，例如，同一样本的工人在实行奖金制度前后产量的比较，同一样本的下蛋鸡在使用某种饲料前后下蛋数量的比较等等；另一种情况是对成对的单位，其中一个给以某种处理，而另外一个则给以另一种处理或不加处理。在进行这种比较时，配对的单位应尽可能一致，例如在研究某种药物的效果时，配对试验的人必须在年龄、性别、体质等方面尽可能一致，不至于有其他因素影响试验的结果。像在工业品的试验中通常将同一材料分成两半，随机地用其中一半进行某种处理。在两个有关样本分析中所研究的变量是这些成对样本两个观察值之差。如果差别比较大，说明这种处理具有效应，反之就是没有效应。

两个有关样本的符号检验是要求有 n 对成对样本观察值$(X_1，Y_1)(X_2，Y_2)\cdots(X_n，Y_n)$，这 n 对观察值是独立的，观察值的计量水准在每对之间至少是顺序的，即可以比较其大小。研究的变量一般是连续的。其假设为

H_0：两个变量之差$(X_i-Y_i)=D_i$ 这一总体的中位数为 0；

H_1：两个变量之差$(X_i-Y_i)=D_i$ 这一总体的中位数不为 0；这是双尾检验，当然也可以是单尾检验。在双尾检验时，如果 H_0 为真，则 D_i 的正负号数目应该比较接近，当正负号的数目差别达到一定界限时就应拒绝 H_0，所以以上假设也可以写成可以看出两个有关样本的符号检验也是二项式检验的特殊情况。

例 8.8 某城市为了克服噪音污染，在全市开展了宣传活动，现从全市中抽取 18 个路口测试了宣传前后的噪音(单位：分贝)记录如下

宣传前 X_i	55	74	68	80	77	69	57	72	63
宣传后 Y_i	41	64	61	70	75	60	53	59	61
宣传前 X_i	52	85	66	71	48	83	78	51	67
宣传后 Y_i	48	79	65	70	47	81	69	50	62

能否以 0.01 的显著性水平说明经过宣传以后,城市的噪音已下降了 5 分贝？

解：H_0: $X_i-Y_i\leq 5$ H_1: $X_i-Y_i>5$

计算观察值之差为 14，10，7，10，2，9，4，13，2，4，6，1，1，1，2，9，1，5 分别减去 5 得 S+=8，S-=9，根据二项式检验，要使 $P(X_i-Y_i)>5$ 的拒绝域概率小于 0.01，则要求 S+在 13 个以上，现只有 8 个正号，因此接受 H0，即不能得出已下降了 5 分贝的结论。

§8.4 建立在等级基础上两个样本的检验

上一节正负符号检验只是利用样本观察值之间的大小，用符号来表示，这样就没有充分利用数据所提供的信息，从而也就损失了部分统计功效。建立在等级基础上的非参数方法弥补了这方面的不足。

§8.4.1 威尔科克森的配对带正负号的等级检验

这是由弗兰克·威尔科克森于 1945 年提出的一种非参数检验。它适用于两个有联系样本的比较，其计量水准至少是按顺序尺度计量的。因为符号检验时只考虑成对样本之间的差别方向，未考虑其差别的大小。而这一检验既考虑差别方向又考虑差别的大小，因此就比正负符号检验敏感，其检验的方法和顺序和以前的假设检验一样，H_0：两个总体是相同的，然后将 n 对成对样本观察值$(X_1，Y_1)(X_2，Y_2)\cdots(X_n，Y_n)$分别计算共差 $d_i=X_i-Y_i$，将 $d_i=0$ 的样本除去，先不考虑其正负符号，根据其绝对值的大小，用等级 1 代表最小的 d_i，等级 2 代表其次小的 d_i，依次类推。当其差的绝对值相同时就用这些值相应等级的平均数代替。然后恢复其原来的正负号，分别将正负符号的等级相加，用 T_+代表正的等级和，T_-代表负的等级和。选择其中较小的一个等级和 T 作为检验的统计量，使之与一定显著性水平的临界值加以比较。若统计量 T 小于临界值就拒绝 H_0。这一点和通常的检验统计量恰好相反。因为从直观上可以想象，当两个总体相同时，T_+和 T_-应该比较接近，当两个总体差别较大时 T_+和 T_-的差别也就大，而当 n 确定时 T_+和 T_-之和是一个常数，所以当两个总体差别比较大时，取较小的一个就应该比较小。T 在各种显著性水平的临界值表中可以查到，见书后附录。

例 8.9 试对例 8.6 的数据用威尔科克森配对带正负号的等级检验作出决策。

解：例 8.6 中是单一样本的数据，而威尔科克森配时带正负符号的检验则是两个有关系的成对样本检验,但这里可以把总体中位数 10 作为每一个配对样本值。这样就可以把这个单一样本观察值作为一个特殊的成对样本的比较。(即 Y_i 用 M_0 代替)可以算出 d_i 和有关正负号的等级，如表 8.10。其计算步骤如下，第一步是列出 X_i 和 Y_i 的观察值，本例中 Y_i 的观察值用总体假设中位数 10 代替，(见表 8.10 第 1 列和第 2 列)。第二步是计算 $d_i=X_i-Y_i$，(见表 8.10 第 3 列)，暂时不考虑其符号，将绝对值列入第 4 列，根据绝对值的大小由小到大编上等级，见第 5 列。第三步是把等级恢复原正负号，分别将正负号的等级相加，得 $T_+=4$，$T_-=51$(见表 8.10 第 6 列)。以较小的 $T_+=4$ 作为检验的统计量。查威尔科克森 T 值的临界值表，当 $N=10$ $a=0.05$ 时，双尾检验的临界值 $T_a=8$，由于 $T<T_a$，因此拒绝 H_0，也就是说钢材长度的中位数并不是 10。把这一结论与例 8.6 的结论相比较可以看出，在正负符号检验 $\alpha=0.05$ 时尚不能推翻 H_0，而利用等级检验时就比较敏感。

在大样本的情况下，统计量 T 也近似服从正态分布，因此就可以应用正态近似的检验方法。

$$T的均值=\frac{n(n+1)}{4}$$

$$T的方差=\frac{n(n+1)(2n+1)}{24}$$

所以
$$Z_T=\frac{T_+-\frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \tag{8.6}$$

由于正态近似的检验方法一再应用，这里就不再赘述。但是这里需要补充说明的是，当有 d_i 数值相等时需要作适当的修正。例如 d_i 为 2，5，5，9，14，14，14，16，其中有 2 个 5，其原有等级应为 2 与 3，现分别用 $\frac{2+3}{2}$ =2.5 代替。同样有 3 个 14，其相应等级为 5，6，7，现分别用 $\frac{5+6+7}{3}$ =6 代替。有了相同的等级以后，T 的方差应作相应的修正，其修正后的方差公式为

$$\frac{n(n+1)(2n+1)}{24}-\frac{\Sigma t_i^3-\Sigma t_i}{48} \tag{8.7}$$

式中：t_i 表示相同等级的样本数，如本例中 $t_1=2, t_2=3$。

例 8.10　某心理学家为弄清经常观看凶杀、暴力等电视节目是否对孩子有影响这一问题进行了一组试验，他选了 16 对孩子，每一对孩子的家庭坏境、遗传基因、智力等等尽可能一致。他让一组孩子经常观看暴力等录像，另一组则经常观看儿童卡通片等。经过一段时期后，确定一些指标分别进行测试，指标的分值高反映粗暴、敢作敢为的倾向大。测试结果如表 8.11。试检验观看暴力录像对孩子是否有影响。

解：

H_0：两组总体相同(即没有影响)

H_1:观看暴力录相组的指标值大于未观看暴力录像组的指标值

通过对表 8.9 的计算可得

$$T_-=1.5+4.5+7+4.5=17.5$$

$$T_+=\frac{16(17)}{2}-17.5=118.5$$

因为 n＝16，接近大样本，故用正态近似计算

$$E(T)=\frac{n(n+1)}{4}=68$$

因为 d_i 中有 2 个 1，分别用 1.5 代替；有 4 个 2，分别用 4.5 代替；2 个 9 分别用 12.5 代替。因此需要修正方差

$$
\begin{aligned}
&\frac{n(n+1)(2n+1)}{24}-\frac{\Sigma t_i^3-\Sigma t_i}{4}\\
&=\frac{16(17)(33)}{24}-\frac{(2^3+4^3+2^3)-(2+4+2)}{48}\\
&=374-1.5=372.5
\end{aligned}
$$

$$Z_+=\frac{T_+-E(T)}{\sqrt{\sigma(T)}}=\frac{118.5-68}{\sqrt{372.5}}=2.63$$

当 a=0.05 时，$Z_{0.95}$=1.645(单尾) $Z+>Z_{0.95}$ 拒绝 H_0，由此看来，经常观看暴力电视节目对儿童是有影响的。

§8.4.2 曼一惠特尼一威尔科克森检验

这种方法也是建立在等级和的基础上的检验，它适用于两个独立样本的比较。假设一个随机样本 X_1，X_2，…，X_{n1} 抽自总体 1，另一个随机样本 Y_1，Y_2，…，Y_{n2} 抽自总体 2，变量为连续的随机变量，使用的计量水准至少是顺序的，要进行检验的是

H_0：这两个总体具有相同的分布(中位数位置相同)

H_1：这两个总体具有不同的分布(中位数位置不同)

以上是双尾检验，通过改变 H_1，也可以是单尾检验。检验的方法是将这两个样本加以合并，按顺序由小到大排列，和以前一样按观察值的大小编上等级，如果遇到观察值相同时，也用它们相应等级的平均数代替。然后分别计算两个样本的等级和 T_1、T_2。从直观上看，若两个总体的分布相同，T_1 和 T_2，在样本容量相同时应该比较接近。在样本容量较小时就可以直接用 T_1 和 T_2 进行检验。例如当 n_1＝3，n_2=5 时，T_1 的等级和是 1，2，…，8 任意 3 个数的组合之和。根据排列组合一共有$\binom{8}{3}$＝56 种可能的组合。用最小的等级加以组合

1+2+3＝6
1+2+4＝7
1+3+4＝8
1+2+5＝8

即$P(T_1\le 8)=\frac{4}{56}=0.0714, P(T_1\le 7)=\frac{2}{56}=0.0357,$因此可以通过 T_i 的大小来判断两个总体是否相同，但是 T_1 的大小受 n_1 大小的影响，在 n_1+n_2 比较大时也不可能一一列举所有可能的样本。曼—惠特尼在这一基础上提出了统计量 U，其中

$$U_1=T_1-\frac{n_1(n_1+1)}{2}, U_2=T_2-\frac{n_2(n_2+1)}{2} \qquad (8.8)$$

U 的各种显著性水平的临界值已编制成表，当 U＜U_a 时就拒绝 H_0。

例 8.11　某工厂欲测定在装配线上男工和女工的机械技能有没有差别，随机选择了 9 个男工与 5 个女工，每一个人都给以机械技能的测试并给以评分，取得结果如表 8.12。

表 8.12　　　　　　　　**男女工人的机械技能测试分数**

男		女	
原始分数	等　级	原始分数	等　级
1500	13	1400	12
1600	14	1200	9
670	2	780	3
800	4.5	1350	11
1100	7	890	6
800	4.5		
1320	10		
1150	8		
600	1		

$T_1=64$　　　　　　$T_2=41$

试以 0.05 的显著性水平检验男工和女工的机械技能是否相同。

解：建立假设

H_0：男女性别之间的机械技能没有差别

H_1：男女性别之间的机械技能存在差别

根据 H_1，这是一个双尾检验。根据分数的顺序编上等级，见表 8.12，计算 T_1 ＝64，T_2＝41，因此

$$U_1 = T_1 - \frac{n_1(n_1+1)}{2} = 64 - \frac{9(10)}{2} = 19$$

$$U_2 = T_2 - \frac{n_2(n_2+1)}{2} = 41 - \frac{5(6)}{2} = 26$$

当显著性水平 $\alpha=0.05$ 时，双尾检验 $n_1=9$，$n_2=5$ 可查表得 $U_a＝7$，因为 $U＞U_a$，接受 H_0，就是说在装配线的男工与女工在机械技能方面并没有差别。

当n_1和n_2增大时，U的抽样分布也接近正态分布，U的均值为$\frac{n_1 n_2}{2}$，U的方差为$\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$ 所以

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \tag{8.9}$$

如果存在相同等级时，则U的方差需加以修正，修正的方差为

$$\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 n_2 (\Sigma t^3 - \Sigma t)}{12(n_1 + n_2)(n_1 + n_2 + 1)}$$

因此，

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12} - \frac{n_1 n_2 (\Sigma t^3 - \Sigma t)}{12(n_1 + n_2)(n_1 + n_2 + 1)}}} \tag{8.10}$$

§8.5　多个样本的检验

§8.5.1　克罗斯考尔—瓦里斯利用等级的单因素方差分析

在第 5 章曾介绍过的方差分析，是通过几个样本来检验这些总体的平均数是否相等。但方差分析要求总体是正态分布的，而且要假设各个总体的方差相等。若这些条件不满足其结论就会受到影响。而克罗斯考尔—瓦里斯单因素方差分析并不依赖于以上严格的假设，它是建立在等级基础上的，如果只考虑两个样本，就相当于上一节的曼—惠特尼 U 检验。

设共有 K 个随机样本，抽自 K 个总体，其样本容量分别为 n_1，$n_2 \cdots n_k$，各个样本是独立的，样本内的各个观察值也是独立的，研究的变量是连续的，计量的水准至少是顺序的。

H_0：k 个总体的中位数是相同的

H_1：k 个总体的中位数并不完全相同

当 H_0 为真时，从直观上看各个样本的等级之和(用样本容量调整)应该比较接近，若等级和的差别较大，将导致拒绝 H_0，克罗斯考尔—瓦里斯检验的统计量就是各样本的等级和与期望的等级之间离差的加权平方和，以样本容量的倒数作为权数，用 H 表示

$$H = \frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{1}{n_i}[R_i - \frac{n_i(n+1)}{2}]^2 \qquad (8.11.a)$$

式中：R_i 是第 i 个样本的等级之和，$\frac{n_i(n+1)}{2}$ 是期望的等级和。这一公式也可以改写为一个比较容易的计算公式

$$H = \frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(n+1) \qquad (8.11.b)$$

H 近似于自由度为 k-1 的χ^2分布，因此就可以利用χ^2分布进行检验。

例 8.12　设对 A、B 和 C 三种牌号的自行车进行评价，将自行车的性能进行分解后评分，最后加成总分进行比较，其评分结果如表 8.13。

表 8.13　　三种自行车的评分结果

A		B		C	
评　分	等　级	评　分	等　级	评　分	等　级
78	12	68	6	82	14
95	20	77	11	65	5
85	16	84	15	50	1
87	17	61	3	93	19
75	10	62	4	70	7
90	18	72	8	60	2
80	13			73	9

要求检验这三种自行车的评分是否一致。

解：

H_0：三种自行车的总评分一致

H_1：不完全一致

将三种评分接顺序排列并编上等级见表 8.13，其中 $n_1=7$，$n_2=6$，$n_3=7$，$n=\sum_{i=1}^{k} n_i=20$，$R_1=106$，$R_2=47$，$R_3=57$。计算克罗斯考尔—瓦里斯统计量

$$H=\frac{12}{20(20+1)}\left(\frac{106^2}{7}+\frac{47^2}{6}+\frac{57^2}{7}\right)-3(20+1)=6.641$$

若显著性水平 α=0.05，则 $\chi^2_{0.95}$ 以拒绝 H_0，因此这三种自行车并不完全一致。

如果在等级中有相同的等级，用它们相应的等级的平均数来代替时，则 H 要进行修正，H 的修正公式为

$$H_c=\frac{H}{1-\left[\frac{\Sigma(t_i^3-t_i)}{(n^3-n)}\right]} \tag{8.12}$$

式中：t_i 是一群相同等级中相同等级的观察值数。

§8.5.2　弗利德曼利用等级的方差分析

这种方法是弗利德曼与 20 世纪 30 年代提出来的，它适用于 k 个有关系的样本，要求计量的水准至少是顺序的。但弗利德曼确定等级的方法和克罗斯考尔—瓦里斯检验中确定等级的方法不同，弗利德曼检验是抽取 n 个样本单位，每个单位给予 k 种不同的处理形成 nk 个观察值如表 8.14。

现对每一个样本单位中的不同处理的观察值给以等级，即每个样本单位由小到大有 k 个等级。如果所有的处理具有同样的效果，那么每一列中的等级应该是随机分布的，若不是随机的，就说明各种处理有差别。弗利德曼检验的统计量就是建立在各列等级和的基础上，其公式为

$$\chi_r^2=\frac{12}{nk(k+1)}\sum_{j=1}^{k}T_j^2-3n(k+1) \tag{8.13}$$

当 H_0(各种处理的效果相同)成立时，χ_r^2 就服从自由度为 k-1 的 χ^2 分布，因此可以利用 χ^2 进行检验。

表 8.14　　弗利德曼双向分析观察值

		处	理		
		1	2	……	k
样	1	X_{11}	X_{12}		X_{1k}
	2	X_{21}	X_{22}		X_{2k}
	M				
本	n	X_{n1}	X_{n2}		X_{nk}

例 8.13　假设有一个新的训练规划分成 4 个单元，每个单元采用一种不同的方法，参加这一训练的有 14 个随机抽选的工人，在每个单元结束时对这 14 人进行考试并给予成绩，见表 8.15。

表 8.15　　14 名工人的考试成绩

样本	方法 1		方法 2		方法 3		方法 4	
	成绩	等级	成绩	等级	成绩	等级	成绩	等级
1	20	(4)	6	(1)	9	(2)	15	(3)
2	5	(1)	12	(3)	19	(4)	10	(2)
3	11	(2)	21	(4)	8	(1)	16	(3)
4	21	(3)	18	(2)	30	(4)	15	(1)
5	8	(1)	12	(2)	20	(4)	16	(3)
6	9	(2)	7	(1)	10	(3)	12	(4)
7	21	(4)	20	(3)	16	(2)	10	(1)
8	18	(3)	27	(4)	9	(1)	12	(2)
9	30	(4)	16	(1)	22	(3)	21	(2)
10	22	(3)	27	(4)	19	(2)	18	(1)

续前表

样本	方法 1		方法 2		方法 3		方法 4	
	成绩	等级	成绩	等级	成绩	等级	成绩	等级
11	10	(3)	8	(2)	4	(1)	12	(4)
12	6	(1)	12	(3)	7	(2)	14	(4)
13	10	(1)	12	(2)	21	(4)	20	(3)
14	14	(2)	11	(1)	23	(3)	27	(4)

要求检验这 4 种方法总的效果之间有无显著差别($\alpha=0.05$)。

解：

H_0：这 4 种方法总的效果之间并没有差别

H_1：这 4 种方法总的效果之间有差别

这里是 4 个有关系的样本，因为这是同样 14 人的考试成绩，成绩可以按顺序加以排列，因此适合用弗利德曼检验。在用这种方法时先要对每一个工人的 4 种考试成绩进行顺序排列，编上等级(见表 8.14 括号内的数字)。然后计算统计量

$$\chi_r^2=\frac{12}{nk(k+1)}\Sigma T_i^2-[3n(k+1)]$$

$$=\frac{12}{14(4)(5)}[34^2+33^2+36^2+37^2]$$

$$-[3(14)(4+1)]=0.43$$

当 $\alpha=0.05$ 时 $\chi_{0.05}^2(3)=7.81$　$0.43<7.81$,因此接受 H_0，即 4 种方法总的效果没有显著差别。

§8.6 其他非参数统计方法

§8.6.1 关于联系的非参数计量

关于研究两组变量之间相互关系的密切程度这一问题在第 6 章曾经介绍过相关系数，但相关系数是反映两组变量之间线性相关的程度，若两组变量之间存在非线性的联系，相关系数就受到限制；另外相关系数要求变量必须是用数字来计量的，而实际资料有时又缺乏数字的计量，因而也要用非参数的方法。

1.斯皮尔曼等级相关系数。斯皮尔曼等级相关系数是反映两组变量之间联系的密切程度，它和相关系数 r 一样，取值在-1 到+1 之间，所不同的是它是建立在等级的基础上计算的。现结合一个例子来加以说明，某工厂对工人的业务进行了一次考试，欲研究考试成绩与每月产量之间是否有联系，若随机抽选了一个样本，其考试成绩和产量数字如表 8.16。

表 8.16　　6 个工人的考试成绩和产量

工　人	考试成绩	产　量	等　级	
			成　绩	产　量
1	50	500	6	6
2	60	510	5	5
3	70	530	4	4
4	80	580	3	3
5	90	560	2	2
6	95	1000	1	1

从表中的数字可以看出，工人的考试成绩愈高其产量也愈高，二者之间的联系程度是很一致的，但是相关系数 r＝0.676 并不算太高，这是由于它们之间的关系并不是线性的，如果分别按考试成绩和产量高低变换成等级(见表 8.16 第 3、4 列)，则可以计算它们之间的等级相关系数为 1。计算等级相关系数可以将数据变换成等级以后用原有的相关系数公式计算，也可以将算出每一对样本的等级之差 d_i，然后用下列公式计算

$$\rho = 1 - \frac{6\Sigma d_i^2}{n^3 - n} \tag{8.14}$$

在前面所举的例子中由于等级完全一致，所有的 d_i＝0，所以ρ=1。

等级相关系数和通常的相关系数一样，它与样本的容量有关，尤其是在样本容量比较小的情况下，其变异程度较大，等级相关系数的显著性检验与普通的相关系数的显著性检验相同。

2.肯达尔一致性系数。肯达尔一致性系数的性质和斯皮尔曼等级相关系数一样，只不过它是用来计量许多个顺序等级变量之间一致性程度的，例如抽取 10 个学生分别按其不同方面的能力确定了等级如表 8.17，欲分析学生在各个方面的等级是否一致。

表 8.17　　10 个学生各种能力的等级排序

学生	动手实验	艺术	文学	音乐	数学	办事	T_i
1	4	5	7	6	5	1	28
2	6	2	1	5	7	4	25
3	1	8	9	2	2	9	31
4	2	6	5	10	1	7	31
5	8	1	2	8	9	5	33
6	10	3	4	3	8	3	31
7	9	7	6	4	10	2	38
8	3	4	3	1	4	8	23
9	5	9	8	7	6	10	45
10	7	10	10	9	3	6	45

其计算公式为

$$W = \frac{12\sum_{i=1}^{k} T_i^2}{K^2 n(n^2-1)} - \frac{3(n+1)}{n-1} \qquad (8.15)$$

式中：T_i 是第 i 个学生各个方面的等级之和；n 是样本中的学生数(本例为 10)；k 是测试比较的方面(本例为 6)。将有关数字代入公式得

$$W = \frac{12(28^2+25^2+\cdots\cdots+45^2)}{6^2(10)(10^2-1)} - \frac{3(10+1)}{10-1} = 0.173$$

这说明这些学生在以上各方面的等级是不太一致的。

以上公式看起来比较复杂，但可以作直观的解释。当各个方面等级排列一致时，各个 T_i 之离差平方达到最大，因为某个学生在以上各个方面都是等级 1，故 $T_1=6$。某个学生在各方面都是等级 10，则 $T_{10}=6(10)=60$，因此其离差平方和在本例中应为

$$(6-33)^2+(12-33)^2+\cdots\cdots+(60-33)^2=2970$$

当各个方面等级排列不一致时，每个学生的 T_i 之间的离差平方和就小，在本例中为

$$(28-33)^2+\cdots\cdots+(45-33)^2=514$$

两者之比为 $\frac{514}{2970}=0.173$，即为肯达尔一致性系数 W。

§8.6.2　关于随机性的检验

在统计分析中常常要求随机抽取样本，但在一般研究工作中却常会遇到一组数据，需要检验是否是随机的。游程检验就是检验一组数据是否随机的方法。为了说明游程检验，首先需要介绍什么是游程。假如我们从妇产医院的记录中找出出生婴儿的性别，按照出生先后的顺序排列为

男，男，女，女，女，男，女，女，男，男，男，男

这里首先是 2 个男的连在一起，接着是 3 个女的连在一起，这种同样性质的单位连在一起就叫做一个游程。以上样本的数列就共有 5 个游程。因此当一个样本容量确定以后这种游程的多少也应该是随机的，因为在随机的情况下就是指没有预先指定的模式。如果样本出现的序列为

女女女女女男男男男男男男

这种可能性是比较少的，出现这种情况，人们就可能怀疑，这是记录人员有意识排列的，而不是随机的。如果出现的顺序为

男女男女男女男女男女男女

人们也可能有同样的怀疑。在前一种情况下只有 2 个游程，后一种情况下共有 12 个游程。因此过多的游程和过少的游程都可能被认为不是随机的。游程检验的原理就是根据样本的容量和游程的多少来判别这一序列是否为随机的。当样本容量为 n，其中一类为 n_1，另一类为 n_2，游程的临界值已经编制成表，可以直接查表。

例 8.14　有一批容器，其重量有些差异。连续抽查了 15 个容器，共重量分别为

3.6,3.9,4.1,3.6,3.8,3.7,3.4,4,3.8,4.1,3.9,4,3.8,4.2,4.1

能否认为其重量的变动是随机的（$\alpha=0.05$）？

解：以上数据的中位数为 3.9，如果是随机的变动，应该是在中位数的上下波动，现将上述数据在 3.9 以下的用“-”号表示；3.9 以上用“+”号表示；正好为 3.9 的除去。则得到以下序列

\- + - - - - - + - + + + - + +

其中：$n_1=6$；$n_2=7$；其游程数为 V=8，查游程检验表，若 H_0(序列是随机的)为真，$P(V\leq 8)=0.733$ 并不是小概率落入接受域，因此可以认为这一序列是随机的。

当 n_1 和 n_2 均增大时，游程 v 的抽样分布也接近正态分布，其均值为

$$\mu_V=\frac{2n_1n_2}{n_1+n_2}+1 \tag{8.16}$$

方差为

$$\sigma_V^2=\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)} \tag{8.17}$$

因而就可以用正态检验

$$Z=\frac{V-\mu_V}{\sigma_V} \tag{8.18}$$

习 题

1.某香皂厂考虑包装的颜色可能对销售有影响，曾作一试验，采用了 4 种不同颜色的包装(红、白、蓝、绿)，在某商店的销售范围内抽选了 200 个家庭主妇，每人送 4 块不同颜色的香皂，告诉她们系采用不同的配方(其实配方是相同的)一个月试用后，让每个主妇自己挑选 4 种颜色中的一种，其挑选结果如下：挑选红色的 50 人，挑选白色的 75 人，挑选蓝色的 30 人，挑选绿色的 45 人。试说明顾客对颜色的喜爱是否有显著差别。($\alpha=0.05$)

2.棉织厂质量检验部门抽验了 50 匹布，每匹布上的疵点数如下

2	1	0	1	1	2	0	5	1	1
3	0	1	1	2	0	1	1	0	0
1	3	4	0	0	1	1	4	1	2
5	2	6	2	4	1	5	1	1	2
0	1	1	0	3	2	0	2	3	3

(1)检验布匹上的疵点是否服从 $\lambda=1.5$ 的普哇松分布。

(2)从数据中估计 λ 并作检验。($\alpha=0.05$)

3.某电视机厂的显像管原来需要进口，现有一国产厂可提供，价格比较便宜。但电视机厂要求显像管寿命的中位数至少应在 500 小时以上才能改用国产，现从国产提供的 1000 个显像管中随机抽取 25 个作寿命测试，为了减少损失凡显像管测试达到 500 小时就停止。这 25 个显像管中有 22 个到 500 小时并没有损坏。请你帮助该厂作一统计分析，看看该厂是否应改用国产显像管。

4.有 A 和 B 二种万能胶，现欲比较其粘合强度，在 10 种不同的材料上加以试验，结果如下(假设数字大表示强度大)

材料 1 2 3 4 5 6 7 8 9 10

A 10 9 20 40 14 30 26 60 30 42 B 12 10 23 45 12 31 20 65 32 39

试说明哪一种万能胶比较有效。

5.一农业试验站测试小猪的公或母对饲养增重是否有影响，在一栏中饲养了 8 头小公猪，一栏中饲养了 8 头小母猪，用同样的条件饲养，一段时间以后其增重如下

小母猪：9.31,9.57,10.21,8.86,8.52,10.53,9.21

小公猪：9.14,9.987,8.46,8.92,10.14,10.17,11.04,9.43

因一头小母猪在饲养过程中死去，所以只有 7 个观察值。试用适当的方法进行一单侧检验。

6.从城市住户调查的样本户中随机抽取 9 户，令 X 表示年总收入，Y 表示食物支出占总收入的%，其数据如下

X:	1200	1300	2100	2250	2500	2600	2870	3700	5000
Y:	42	36	33	25	27	28	29	25	20

计算等级相关系数，并加以说明。

7.有一种植物发生了病虫害，随机抽取一行逐棵检查，发现有病虫害的用 D 表示，无病虫害的用 H 表示，其结果为

HHHHHDDDDHHHDHHHH

试检验病虫害的分布是否为随机的。

8.从一个大城市的中学里随机选出一部分男学生，问他们毕业后的打算，他们的回答是

	找工作	上大学	未定
一年级	50	50	300
二年级	45	105	200
三年级	80	150	80

用适当的检验说明各年级学生的打算是否有显著差别。(α=0.05、

第 9 章　抽样调查

§9.1　引　言

俗话说，“巧妇难做无米之炊”。统计分析也是一样，它离不开统计数据。即使有了统计数据，如果这些数据不是准确可靠的，也可能会得出错误的结论。因此如何根据研究的需要去获得统计数据是统计分析的一个前提。在第 2 章我们已经介绍了统计资料搜集的一般问题，这一章将专门介绍如何利用抽样调查这一重要手段去获取统计资料，并利用参数估计的方法对这些资料进行科学的分析和推断。

§9.1.1　什么是抽样调查

抽样调查是从研究的总体中抽取部分单位组成样本进行调查(或观察)，并利用样本的信息来推断总体特征，它是一种非全面调查。抽样调查分为概率抽样与非概率抽样两大类。本章只讨论概率抽样调查，它是建立在概率论和数理统计基础上的一种抽样方法，有以下特点：

1.一个要进行抽样调查的总体，能够被抽到的各个不同的样本所组成的集合，是能够加以确定的，每一次抽样相当于一个样本点，全部可能样本组成样本空间。

2.每个可能抽到的样本都被规定一个已知的被抽中概率 πi,在抽样时要根据随机原则排除主观有意识的选择样本单位。

3.以样本的数据从数量上推断全体，按照概率论的原理可以对调查结果的抽样误差在一定可靠性条件下作出推断。

§9.1.2　抽样调查的作用

1.有些客观现象不能作全面调查，而又需要了解总体的情况，这就只能抽取一部分进行调查。例如破坏性的产品检验，导线的拉力强度，灯泡、显像管的寿命，炮弹的射程以及农业中的种子发芽试验等等。

2.一些大规模的调查如市场需求情况，家庭收支状况，电视观众对电视台节目的意见以及民意测验等等，虽然在理论上可以进行全面调查，但实际上很困难，甚至几乎是不可能的。

3.抽样调查可以节约调查的人力和费用。统计工作搜集资料也像其他工作一样有一个经济效益问题，即在满足需要的前提下，如何能节省人力物力。

4.速度快。当有些资料具有很强的时间性时，以有限的人力和物力采用全面调查会使时间拖延，只能获得陈旧的信息；而抽样调查能及时获取信息。

5.由于抽样调查的调查单位少，调查队伍可以专门训练，因此可以适当增加调查内容并提高调查的质量。

由于抽样调查是一种非全面调查，人们往往对调查结果的可靠性持怀疑态度，总认为不如全面调查准确，因此有必要分析一下如何看待抽样调查的准确性问题。

第一，人们对客观事物准确性的要求是相对的，尽管对任何一个调查来说，人们总是希望调查的准确性越高越好。但是对不同的调查内容而言，人们对准确性的要求是大不相同的，正如对航天飞机助推火箭主轴的加工精度的要求和制造一根普通钢锭所要求的精度不能相提并论一样，在统计工作中检验测量精密仪器时所要求的精确性和调查某个电视节目收视率所要求的精确性显然也是不一样的。这些例子说明我们对调查准确度的要求不是绝对

的，而是相对的，即要求调查达到我们能够接受的精度即可。

第二，在实际调查中，准确度常常是和调查费用相联系的。一般说来，数字精确度要求越蒿，费用开支也就越大。但这二者之间的关系不是线性的，因为在一般情况下费用与样本容量之间有线性关系，而样本容量与调查精度之间不是线性关系。当样本容量比较小时，每增加一个样本单位对提高精度的影响比较大；随着样本容量的增大，每增加一个样本单位的影响就逐渐减小,这一关系可用图 9.1 来说明。假设用 100%的费用可以达到 100%的精度，而费用支出提供的精度是递减的。如图 9.1 中, 以 20%的费用可得到 50%的精度；以 50%的费用可以达到 90%的精度，如果有 90%的精度基本上可以满足需要、就没有必要以另外的 50%的费用来换取余下 10%的精度。由此可见，我们的目标是使用尽可能少的费用支出，来取得准确性合乎要求的数据，而不必求绝对精确。

图 9.1　费用与精度的关系示意图

第三，抽样调查有时会比全面调查更准确。因为在调查中有两类误差：一大是抽样误差，即样本对总体代表性引起的误差，这是抽样调查所特有的；另一类是非抽样误差，这是所有的调查所共有的。如调查中有关项目的定义不清楚，调查单位有遗漏或重复，不准确的计算方法，问卷设计不合理，被调查人有意错报，调查人员的粗心大意等等，有时这些误差远远超过抽样误差以致不能反映真实情况。因此尽管全面调查不存在抽样误差，但是由于调查面广，调查人员水平不一等原因，会使非抽样误差比较大。抽样调查由于调查单位少，调查人员可以进行培训使之比较精干，而且在调查组织工作方面人力集中，调查的过程可以更仔细地监督检查，所以调查质量可以提高，从而非抽样误差可以大大缩小。这样倒有可能取得更准确的结果。当然，抽样调查也不能完全代替全面调查，二者各有不同的作用。

§9.1.3　抽样调查的几个基本概念

1，总体与样本。这是抽样调查中最基本的概念，尽管我们还没有正式介绍，但是在前面已经多次提到。总体是指统计研究对象的全体，它是由许多元素(通常叫总体单位)所构成的。总体的特点是各个总体单位都有某种相同的性质把它们联系在一起，但是各个个体的某个标志之间却又存在着差异，这种统一和差异就构成了一个总体分布。抽样调查的目的往往是对总体分布的一些特征作出估计。因此在进行一项抽样调查时首先遇到的是确定总体的问题。总体的划分有时比较容易，如要估计一批灯泡的寿命，总体就是这批待检验的灯泡，似乎没有什么问题。但有时候却并不容易，如要在全国的电视观众中进行对电视节目意见的调查，就必须对电视观众下一个定义，一些难以划分的情况就会发生，这需要根据研究的目的和进行调查的可能来划分。

总体根据总体单位的数目可以分为有限总体和无限总体两类。有限总体是指总体单位数是有限的，在理论上是可以进行全面调查的，但由于上述种种原因而往往采用抽样调查方法，社会经济调查中多数是这种情况。无限总体是指总体单位数是无限的，例如在自然科学中的试验，它可以无限次地进行下去，因此也只能是通过抽样来取得数据。抽样调查时从总体中抽取的那部分单位所组成的集合体，就称作样本，样本是总体的一个缩影。抽样调查是从样本中获取信息来对总体的有关特征作出估计。由于抽样方式不同，抽

样单位可以是总体单位，也可以不是总体单位。样本中包括的单位数称作样本容量，但当抽样单位和总体单位不一致时样本容量是指抽样单位数还是总体单位数应加以说明。

2.抽样框。它是指包含有全部总体单位及其主要标志特征的一个框架。抽样框是从中抽选样本的基础资料，故应把所研究总体的全部单位都包括在内而不要遗漏或重复。但有时候要取得与研究的总体相一致的抽样框比较困难，只能用一个近似的抽样框来代替。在这种情况下，用样本推断总体时就要考虑抽样的总体与目的总体之间的差别。抽样框的主要形式有两种，一种是一览表，表中列出总体的全部单位，有时还应该包括一些辅助资料。例如某工业局所属企业一览表，某市国营零售商业网点一览表；某校学生花名册等等。另一类抽样框的形式是地图，在地图上将调查单位分布在各个地区，然后根据地图进行抽样。

3.抽样平均误差。前面曾经提到过抽样误差，它是指样本指标值与所要估计的总体真值之间数量上的差别，这种差别纯粹是由抽样引起的，所以称作抽样误差。但是总体的真值是不知道的，因此每一次抽样的结果其确切的抽样误差也无法知道。但是从数理统计的理论可知，如果总体的方差已知，在样本的容量确定以后，所有样本的平均抽样误差是可以计算出来的。平均抽样误差也叫做抽样标准误，它的平方叫做抽样方差。抽样平均误差既可以用来反映一个抽样设计的精确程度，也可以对总体的真值作区间估计。

4.偏误。这是抽样误差以外的一种偏差。如果在抽样调查中把所有可能样本的标志值加以平均，其平均数等于总体真值，通常称为无偏估计；如果不等于总体真值就是有偏估计，其平均数与总体真值之差就称为偏差，因此这种误差是一种方向性的系统偏差。抽样误差与偏差可以形象地用一个射击打靶的例子来说明，见图 9.2。

图 9.2　方差和偏差示意图

我们把每一个样本的结果比喻为一次打靶的着弹点，1 号靶是密集于靶心周围，说明每个样本虽有抽样误差，但平均起来中心也在靶的中心，因而没有偏误；2 号靶虽然密集程度差些，也即抽样平均误差较大，但也没有偏差；3 号靶虽然平均误差小但有偏差，4 号靶则平均误差较大且又有偏差。

§9.2 简单随机抽样

§9.2.1 简单随机抽样及其抽样方法

简单随机抽样又称纯随机抽样，是一种最基本的抽样方式。设总体的大小为 N，从中随机抽取容量为 n 的样本，每一个样本都有同样的机会被抽中，这种抽样的方法称为简单随机抽样，所抽到的样本为简单随机样本。

简单随机样本的抽取可以有多种方法，抽签、摸球就是最原始的方法。具体做法是每一个被抽选的总体单位都用一个签或球来代表，然后把它们搅均匀，从中随机摸取，直到抽够所需的 n 个样本单位数目为止，抽中者即为样本单位。简单随机抽样又可以分为重复抽样和不重复抽样两种。重复抽样也称作有放回抽样，即在抽取下一个样本单位时，把上一个抽中的单位放回去，因此一个单位有被重复抽中的可能；不重复抽样也称为无放回抽样，即抽中的单位不再放回去，因此一个单位只有一次抽中的机会。在社会经济的抽样调查中一般是不重复抽样。简单随机抽样最常用的是用随讥数字表抽取样本，这种表是由计算机或其他随机方法制成的，即 0，1，…9 这 10 个数字出现的机会是等概率的，但排列的顺序是随机的。使用时首先将总体单位编上号码(1—N)，然后根据总体单位的数目决定从表上抽取数字的位数，假定 N＝638，是三位数，就需要在表上相应地取三位数，取数的方法可以用随机的起点，按预先规定取数的方向抽取三位数，凡抽到的数字在 638 以下，相应的单位就为抽中单位；若抽中的单位大子 638，则予以舍去；继续按规则抽取，在不重复抽样时遇到重复的号码也同样舍去，直到抽够 n 个样本单位为止。

§9.2.2 总体均值的估计

在抽样调查对总体的有关参数进行估计时，通常最关心的是对总体均值、总体总值以及总体的比例进行估计。为了以后叙述方恒，现将有关符号说明如下：总体单位的标志值本章用 Y_i 表示(Y_1，Y_2，…Y_n)；样本中各单位的标志值用小写 y_i 表示，平均数在符号上面加一“ — ”号，如$\overline{Y},\bar{y}$ Y 等，估计值在原总体符号上加一“ ^ ”号，如 $\hat{\overline{Y}}=\bar{y}$。现将本章符号列举如表 9.1。

表 9.1 简单随机抽样符号

	总　体	样　本
均值	$\overline{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i$	$\bar{y}=\frac{1}{n}\sum_{i}^{n}y_i$
总值	$Y=\sum_{i=1}^{N}Y_i$	$\sum_{i}^{n}y_i$
比例	$p=\frac{A}{N}$	$p=\frac{a}{n}$
方差	$\sigma^2=\frac{1}{N}\Sigma(Yi-\overline{Y})^2$ $S^2=\frac{1}{N-1}\Sigma(Y_i-\overline{Y})^2$	$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$

在简单随机抽样中样本均值为总体的无偏估计，而且随着样本容量 n 的增大，样本的均值趋向于正态分布，因此总体均值的估计公式为

$$\bar{\bar{Y}} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (9.1)$$

$$V(\bar{y}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \text{或} \frac{S^2}{n} \cdot \frac{N-n}{N} \qquad (9.2)$$

(9.1)和(9.2)式中σ^2与 S^2 都是总体方差，只是两种不同的表示方法，其差别在表 9.1 中已有说明。当用 S^2 代替σ^2 时，均值方差公式后面的系数变为$\frac{N-n}{N}$，也可以写成 1-f，f=$\frac{n}{N}$称作抽样比，它为公式的推导带来很多方便。但是在抽样时总体方差常常是未知的，也需要从样本中获取信息，样本方差 s^2 是 S^2 的无偏估计，而且当大样本时 s^2 趋近于 S^2，因此也常用 s^2 代替 S^2 来估计均值的方差，方差的估计量用 $s^2(\bar{y})$表示。这时理论上应该用 t 分布进行区间估计，但是在大样本情况下 t 分布与正态分布之间的差别也就不大了。

例 9.1　某工厂欲了解工人由于停工待料及机器故障所造成的每个工人平均每周的工时损失数。全厂共有 750 个工人，从中抽取 50 个工人为样本进行调查，得到每个工人平均每周损失工时数为 10.31 小时，且 s^2=2.25。试估计推断全厂每个工人平均每周工时损失数，并以 95%的可靠性作出区间估计。

解：可以用样本平均数来估计总体平均数，根据公式(9.1)

$\bar{\bar{Y}} = \bar{y}$＝10.31小时，

在作区间估计时，需要知道 S^2，因为是大样本可以用样本方差来代替。根据公式(9.2)的估计量

$$s^2(\bar{y}) = \frac{s^2}{n}(1-f) = \frac{2.25}{50}(1-\frac{50}{750}) = 0.042$$

抽样标准误为到 $s(\bar{y})$=0.205，平均每人每周由于停工待料及机器故障的工时损失约为

$$\bar{y} \pm 2s(\bar{y}) = 10.31 \pm 0.41$$

即 9.9—10.72 小时，估计的可靠性约为 95%。

例 9.2　某乡共有 484 户，现欲了解该乡家庭副业的平均月收入。随机抽取了 9 个家庭为样本，这 9 户的收入分别为 33.5，32，52，43，40，41，45，42.5，39。假设该乡的家庭副业收入近似正态分布，试估计该乡每户的月平均副业收入，并以 95%的置信系数对全乡这一总体平均数作出区间估计。

解：首先计算样本平均数作为总体平均数的估计值，根据公式(9.1)

$$\bar{\bar{Y}} = \bar{y} = \frac{1}{n}\Sigma y_i = \frac{368}{9} = 40.86$$

为了确定置信区间，需要知道总体方差，而总体方差未知，只能用样本方差来估计。

$$s^2 = \frac{1}{n-1}\Sigma(y_i - \bar{y})^2 = \frac{1}{8}[15332.9 - 9(40.89)^2]$$
$$= 35.67$$

根据公式(9.2)的估计量有

$$s(\bar{y})=\sqrt{\frac{s^2}{n}(1-f)}=\sqrt{\frac{35.67}{9}(1-\frac{9}{484})}$$
$$=1.9722$$

由于用样本方差代替总体方差，而且样本的容量比较小，所以应该用 t 分布，当 α=0.05，自由度为 n-1=8 时，t=2.306。故该乡每户平均副业收入的区间估计是 40.89±2.306(1.9722)，大约在 40.89±4.55 之间。

§9.2.3　总体总值的估计

一些抽样调查的目的是要估计总体总值，例如估计农作物总产量等。估计总体总值的公式比较简单，即在求平均数的基础上乘以总体单位数 N 即可，其总体总值及其方差的公式为

$$\hat{Y}=N\bar{y}=\frac{N}{n}\sum_{i=1}^{n}y_i \tag{9.3}$$

$$V(\hat{Y})=N^2V(\bar{y})=\frac{N^2S^2}{n}(1-f) \tag{9.4}$$

当方差公式中的 S^2 用样本方差 s^2 估计时 $V(\hat{Y})$ 就用 $s^2(\hat{Y})$ 表示。

例 9.3　根据例 9.1 的资料估计全厂由于停工待料及机器故障造成的工时损失效。（α＝0.05）

解：例 9.1 已经计算了平均每个工人的工时损失数，又知全厂共有 750 个工人，根据公式(9.3)

$$\hat{Y}=N\bar{y}=750(10.31)=7732.5\text{工时}$$

根据公式(9.4)的估计量计算方差

$$s2(\hat{Y})=N^2S^2(\bar{y})=(750)^2(0.042)=23625$$

全厂工时损失数的置信区间为 7732.5±2$\sqrt{23625}$ 即 7425.1-8032.5 工时，置信系数为 95%。

§9.2.4　总体比例的估计

抽样调查中还经常调查总体中具有某种特征的单位数所占的比例。例如工业产品中有缺陷的产品所占的比例，居民户中具有彩色电视机的户数所占的比例等。这种抽样可以看作是概率论中的贝努里试验，其每一个单位的观察结果只能分成两种情况，或是属于这一类或是不属于这一类。若今总体中具有某种特征的单位 Y_i＝1，不具有某种特征的单位 Y_i＝0，那么当样本容量为 n 时，具有某种特征单位的总数为 $\sum_{i=1}^{n}y_i$，其所占的比例为 p，p＝$\frac{1}{n}\sum_{i=1}^{n}y_i$，p 也可以看作是样本中 1 和 0 的平均数。总体比例及其方差的估计公式为

$$\hat{P}=\frac{1}{n}\Sigma y_i \quad (y_i\text{取}1\text{或}0) \tag{9.5}$$

$$V(\hat{P})=\frac{PQ}{n}(\frac{N-n}{N-1}) \quad (\text{其中}Q=1-P) \tag{9.6}$$

但总体比例 P 通常是未知的，所以也要用样本比例 p 来估计，V(P)的估计量为 S^2(P)=$\frac{pq}{n-1}(1-f)$

例 9.4　某市郊农村共有 300 户，欲调查拥有彩电的户数所占的比例，

现随机调查了 100 户，结果有 35 户拥有彩电，试估计全村拥有彩电户数占的比例(置信系数为 0.95)。

解：已知$N=300$，$n=100$，$\Sigma y_i=35$，根据公式(9.5)

$$p=\frac{1}{n}\Sigma y_i=\frac{35}{100}=0.35(\text{或}35\%)$$

其方差估计量为 $s^2(p)=\frac{pq}{n-1}(1-f)$

$$=\frac{(0.35)(0.65)}{99}(1-\frac{1}{3})=0.001532$$

$\bar{p}\pm 2s(p)=0.35\pm 2(0.039)$，即在$27.2\%$—$42.8\%$之间。

§9.2.5 样本容量的确定

在进行抽样调查的方案设计时，一个重要的问题是确定样本容量，因为样本太大就要耗费过多的人力物力；样本太小，又可能出现较大的误差，使调查得到的资料难以满足要求，所以样本的大小要确定得适当。

在简单随机抽样的情况下，样本容量的大小取决于总体方差的大小、允许误差的范围以及估计推断结论的可靠性要求。

设总体各单位之间的方差为σ^2，在社会经济的抽样调查中可靠性一般要求为 95%,那么在正态分布条件下其概率度 t=1.96,为了方便通常用 2 代替，平均数估计量允许误差常用Δ表示。确定样本容量的公式为

$$n=\frac{Nt^2s^2}{N\Delta^2+t^2s^2} \tag{9.7.a}$$

若估计的是总体比例，则其公式为

$$n=\frac{Nt^2pq}{N\Delta^2+t^2pq} \tag{9.7.b}$$

式中：P 是预计的总体比例，但这正是抽样目的所要估计的，若事先没有任何信息可以利用时，可以采用保守的估计，令 P＝Q＝0.5，因为这时 PQ 的乘积为最大。

例 9.5 现欲调查某市郊区的平均每户年收入，已知该郊区共有 1100 户，特别高和特别低的户约有 100 户，余下的 1000 户收入大致呈正态分布，高与低的户之间大约相差 400 元。若要估计有正常收入的 1000 户的户均收入，以 95%的置信系数保证估计误差不超过 20 元，则应抽多少户作样本？

解：由于总体的方差未知，但已知收入服从正态分布而且高与低之间大绚相差为 400 元，根据全距与标准差的关系，大约 95%的户在$\pm$2s 之间，因此 4s＝400，S＝100，S^2＝10000。根据公式(9.7.a)

$$n=\frac{Nt^2s^2}{N\Delta^2+t^2s^2}=\frac{1000(2)^2(100)^2}{1000(20)^2+(2)^2(100)^2}$$
$$=90.91=91$$

即大约要抽 91 户。

例 9.6 某大学的学生会想在全校 2000 个学生中进行一项民意测验，了解赞成某一倡议的人占多大比例，要求误差不超过 0.05，应抽多少学生为样本？

解：由于这是一项新的倡议，没有过去的信息可以利用，因此可以令 p＝0.5，若置信系数为 95%则 t=2，根据公式(9.7.b)

$$n=\frac{Nt^2pq}{N\Delta^2+t^2pq}=\frac{2000(2)^2(0.5)^2}{2000(0.05)^2+2^2(0.5)^2}$$

$$=\frac{2000}{6}=334$$

即应抽 334 个学生进行调查，才能达到要求的精确度。

§9.3 分层随机抽样

§9.3.1 分层抽样及其作用

分层抽样也称分类抽样或类型抽样，这种抽样方法是在抽样之前将总体的N个单位划分为互不交叉重叠的若干层(类)，设为L层，每一层所包含的单位数分别为 N_1，N_2，$\cdots N_L$。且 $\sum_{i=1}^{L} M_L = N$，然后再从各层中独立地抽取一定数量的总体单位，组成样本。例如，对工业企业进行调查时可以把企业划分为大型企业、中型企业和小型企业三个层；对某种农作物进行调查时可以分为山区、丘陵和平原等。

分层抽样的作用可分为以下几个方面：

1.利用已知的信息提高抽样调查的精确度，或者在一定精度下减少样本的单位数以节约调查费用。因为在简单随机抽样的情况下，抽样误差的大小主要取决于总体内部的差异大小和抽取样本容量这两个因素。而在实际的抽样调查工作中总体的差异是客观存在的，要达到减少误差的目的就要增大样本的容量，从而增加调查费用。为了解决这个矛盾，分层抽样是一种理想的方法。如果人们事先对研究的总体有一定的了解，就可以把总体中性质相同的单位，即研究的标志值比较接近的单位归并在一起，形成若干层，这样各层内的差异就可以大大缩小，各层能以较小的样本容量达到预期精确度的要求。从整个样本来说，由于这些样本单位对各层均有较高的代表性，且从整体来看分层后抽取的样本单位在总体中散布得更均匀，所以由它们构成的样本对整个总体也就有较高的代表性。根据方差分析的原理，对总体进行分层后，总体方差可以分解成两部分，一部分是层内方差，一部分是层间方差。在分层抽样时，抽样误差只和层内方差有关，而与层间方差无关，因此在分层时只要能扩大层间方差缩小层内方差，就可以提高抽样效率。但这里也说明了分层抽样应有一定的条件，即在抽样之前人们对客观的总体要有一定的了解，另外还需要知道各层的总体单位数，这是利用主观认识提高抽样效率的一种手段。

2.分层抽样有时是为了工作的方便和研究目的的需要；如果抽样调查既要了解总体的有关信息，又要了解一些子总体的信息，这种情况下就可以将子总体分层。例如按行政隶属的系统分层，按地理的区划来分层等等。但需要指出，这种分层有时与提高抽样效率的目的并不完全一致，但对工作会带来很大方便。因此如何分层需要根据研究目的来确定。

§9.3.2 总体均值和总值的估计

1.有关分层抽样的符号请见表9.2。

2.总体平均数的估计。分层抽样总体平均数及其方差的估计公式为

$$\hat{\bar{Y}} = \bar{y}_{st} = \frac{1}{N}\Sigma N_i \bar{y}_i$$

或 $$= \Sigma W_i \bar{y}_i \text{其中} W_i = \frac{N_i}{N} \tag{9.8}$$

$$V(\hat{\bar{Y}}) = \frac{1}{N^2}\Sigma N_i (N_i - n_i)\frac{S_i^2}{n_i} \tag{9.9}$$

或 $$= \Sigma W_i = \frac{S_i^2}{n_i}(1 - f_i) \text{其中} f_i = \frac{n_i}{N_i}$$

和前面一样，当用样本方差代替总体方差时，$V(\bar{Y})$就写成 $S^2(\bar{Y})$。

表 9.2　　分层抽样的符号

总　　体	样　　本
N：总体单位数	n：样本容量
N_i：第 i 层的总体单位数	n_i：第 i 层样本容量
$N=\Sigma N_i$	$n=\Sigma n_i$
$S_i^2 = \frac{\sum_{j=1}^{N}(Y_{ij} - \bar{Y}_i)^2}{N_i - 1}$ i层方差	$S_i^2 = \frac{\sum_{j=1}^{n}(\bar{y}_{ij} - \bar{y}_i)^2}{n_i - 1}$ i层样本方差
Y_{ij} 第 i 层第 j 个单位标志值	y_{ij} 样本中第 i 层第 j 个单位标志值
$\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} Y_{ij}$	$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$
$\bar{Y} = \frac{1}{N}\sum_{i=1}^{L} N_i \cdot \bar{Y}_i$	$\bar{y}_{st} = \frac{1}{n}\Sigma N_i \bar{y}_i$

由于分层抽样是分层以后在各层内进行随机抽样，因此从简单随机抽样的估计推断中可知，各悬的样本平均数是各层总体平均数的无偏估计；而总平均数则是各层平均数的加权平均数，这里应注意的是其权数应该是各层总体单位数，而不是各层样本单位数，即 $E(\bar{y}_{st}) = E(\frac{1}{N}\Sigma N_i \bar{y}_i) = \frac{1}{N}\Sigma N_i \bar{Y}_i = \bar{Y}$。显然当分层抽样时，各层抽取样本的比例与总体中各层的比例相同，则二者加权的结果也是相同的，否则二者的结果是不同的。同样道理，也应以总体权数的平方来加权各层抽样的抽样方差。

例 9.7　对某市 600 个个体户的月零售额进行抽样调查，现按申报资金分为大、中、小三类，根据调查结果的数据整理如表 9.3。

试估计该市每个个体户的平均月零售额，并以 95%的可靠性作出区间估计。

解：因为$f_1 = \frac{n_1}{N_1} = \frac{30}{60} = 0.5$；$f_2 = \frac{n_2}{N_2} = \frac{40}{240} = 0.17$；

表 9.3 某市个体户月零售额 （单位：千元）

层　别	N_i	n_i	$\bar{y}_i$	S_i^2
大	60	30	20	16
中	240	40	8	4
小	300	40	1	0.5

$$f_3=\frac{n_3}{N_3}=\frac{40}{300}=0.13$$

这是一个不等比例的抽样，这种抽样方法在实际调查工作中是常用的，因为大商店虽然数量少但零售额比重大，对整个个体户的零售额有比较重要的作用，因此多调查几个大户，使这一层的精确度提高，就能提高整个调查的精确度，所以根据公式(9.7)及(9.8)

$$\hat{\bar{Y}}=\bar{y}_{st}=\frac{1}{N}\Sigma N_i\bar{y}_i$$

$$=\frac{1}{600}[(60)(20)+(240)(8)+(300)(1)]=5.7$$

$$S^2(\hat{\bar{Y}})=\frac{1}{N^2}\Sigma N_i(N_i-n_i)\frac{S_i^2}{n_i}=0.0187$$

$$S(\hat{\bar{Y}})=0.1367$$

由于样本容量比较大，可以按正态分布作出区间估计，每个个体户每月零售额平均为 5.7±2(0.1367)千元。

3.总体总俏的估计。总体均值乘以总体单位数即为总体总值。因此分层抽样总体息值及其方差的估计公式为

$$\hat{Y}=N\bar{y}_{st}=\Sigma N_i\bar{y}_i \tag{9.10}$$

$$V(\hat{Y})=\Sigma N_i(N_i=n_i)\frac{S_i^2}{n_i} \tag{9.11.a}$$

$$S^2(\hat{Y})=\Sigma N_i(N_i-n_i)\frac{S_i^2}{n_i} \tag{9.11.b}$$

例 9.8　根据例 9.7 的资料，估计全市个体户的月零售额。解：根据公式(9.10)及例 9.7 的计算结果

$$\hat{Y}=N\bar{y}_{st}=600(5.7)=3420\text{千元}$$

$$S^2(\hat{Y})=\Sigma N_i(N_i-n_i)\frac{S_i^2}{n_i}=6735 \qquad S(\hat{Y})=82$$

因此全市个体户月零售额约在 3420±2(82)千元之间。

§9.3.3　样本数目在各层间的分配

当分层抽样总的样本容量 n 确定以后，如何分配备层的样本数目，这也是分层抽样需要研究的一个问题。

一种常用的分配方法是等比例分层抽样，即按照总体备层的单位数多少分配每层应抽的样本单位数，第 i 层的样本单位数可以表示为

$$n_i = n\frac{N_i}{N} \tag{9.12}$$

将这一关系代人分层抽样的方差公式经过简化可得

$$V(\bar{Y}) = \frac{1-f}{n}\Sigma W_i S_i^2 \tag{9.13.a}$$

若为重复抽佯，修正系数可忽略不计

$$V(\bar{Y}) = \frac{1}{n}\Sigma W_i S_i^2 \tag{9.13.b}$$

可以看出，分层抽样的方差是取决于加权的层内方差，而层内方差一般总是小于总体方差，为此通过分层可以取得效益。

但是等比例抽样只考虑各层总体单位数来分配样本单位数，而未考虑各层的方差大小，由于抽样标准误与总体的方差成正比，因此在分配样本时也应考虑各层的方差大小，如果层内方差大的层能多抽一些单位就可以使该层的抽样误差降低；而在层内方差小的层可以少抽一些单位，该层的抽样误差也不至于太大，这样就有可能使总的抽样误差降低。统计学家奈曼于 1934 年提出了根据各层总体单位数以及各层标准差结合起来分配样本的方法，可以使在同样样本容量的条件下方差达到最小，称作奈曼分配，其分配公式为

$$n_i = n\frac{N_i S_i}{\Sigma N_i S_i} \text{或} n = \frac{W_i S_i}{\Sigma W_i S_i} \tag{9.14}$$

例 9.9　某广告公司欲调查该公司所在市居民户每周看电视的小时数，根据居民的类型分成三类：市区、近郑和远郊，已知市区的居民户占 50%，近郊的居民占 20%，远郊的居民占 30%。从看电视的小时数看，近郊居民户的方差较大，约为市区居民户方差的 6 倍，远郊居民户为市区居民户的 2 倍。现欲抽取 200 户作为样本，试分别用等比例分层和奈曼分配确定各层的样本

解：根据已知数据列表并计算如表 9.4。

表 9.4　　等比例分层和奈曼分配计算表

层(i)	总体单位比例(W_i)	总体方差 S_i^2	S_i	W_iS_i
市区	0.5	S_1^2	S_1	$0.5S_1$
近郊	0.2	$6S_1^2$	$\sqrt{6}\,S_1$	$0.2(\sqrt{6}\,S_1)$
远郊	0.3	$2S_1^2$	$\sqrt{2}\,S_1$	$0.3(\sqrt{2}\,S_1)$

等比例分层抽样的样本分配为

$$n_1=nW_1=200(0.5)=100$$

$$n_2=nW_2=299(0.2)=40$$

$$n_3=nW_3=200(0.3)=60$$

奈曼分配为

$$n_1 = n\frac{W_1S_1}{\Sigma W_iS_i}$$

$$= 200 \cdot \frac{0.5S_1}{0.5S_1 + 0.2(\sqrt{6}S_1) + 0.3(\sqrt{2}S_1)}$$

$$= 200(0.3536) = 71$$

$$n_2 = n\frac{W_2S_2}{\Sigma W_iS_i} = 200 \cdot \frac{0.2(\sqrt{6}S_1)}{1.4142S_1} = 200(0.3464) = 69$$

$$n_3 = n\frac{W_3S_3}{\Sigma W_iS_i} = 200 \cdot \frac{0.3(\sqrt{2}S_1)}{1.4142S_1} = 200(0.3) = 60$$

§9.3.4　总体比例的估计

和简单随机抽样中的情况一样，分层抽样有时也要研究总体中具有某种特征的单位数所占的比例。例如电视台需要了解某个节目的收视率(收视该节目的户占总户数的比例)。采用分层抽样一方面可以将不同类型的居民分成不同的层，这样既可以了解总的收视率，也可以了解各不同层居民的收视率；另一方面恰当地分层可以提高抽样效率。用分层抽样估计比例及其方差的公式为

$$\hat{p}_{st} = \frac{1}{N}\Sigma N_i \hat{p}_i \qquad (9.15)$$

$$V(p_{st}) = \frac{1}{N^2}\Sigma N_i^2 \frac{N_i - n_i}{N_i - 1} \cdot \frac{P_iQ_i}{n_i} \qquad (9.16.a)$$

$$S^2(p_{st}) = \frac{1}{N^2}\Sigma N_i(N_i - n_i)\frac{p_iq_i}{n_i} \qquad (9.16.b)$$

例 9.10 某广告公司要了解在电视上做广告的作用，拟在有关对象中调查看电视广告的比例。设对象分为三层，N_1＝155，N_2＝62，N_3＝93，样本容量为 40。采用等比例分层抽样，调查结果第一层的比例为 0.8；第二层为 0.25；第三层为 0.5。试以 95%的可靠性，估计调查对象中收看电视广告比例的置信区间。

解：根据公式(9.15)

$$p_{st} = \frac{1}{N}\Sigma N_i P_i$$

$$= \frac{1}{310}[(155)(0.8) + (62)(0.25) + (93)(0.5)] = 0.6$$

由于样本容量为 40，采用等比例分层，其抽样比例 $\frac{40}{310}$，可求得 n_1＝20，n_2＝8，n_3＝12，根据公式(9.16)

$$S^2(p_{st})=\frac{1}{N^2}N_i(N_i-n_i)\frac{p_iq_i}{n_i}$$

$$=\frac{1}{310_2}[155(155-20)\frac{(0.8)(0.2)}{20}$$

$$+60(62-8)\frac{(0.25)(0.75)}{8}$$

$$+93(93-12)\frac{(0.5)^2}{12}]$$

$$=0.0045，S(p_{st})=0.067$$

$p_{st}\pm 2S(p_{st})=0.6\pm 2(0.067)$，即观看电视广告的比例约在 46%—74%之间，可靠性为 95%。

§9.3.5　样本容量的确定

在分层抽样的条件下，样本容量的确定与简单随机抽样有共同之点，这就是它取决于一定的允许误差和结论的可靠性；但不同的是分层抽样与各层的层内方差有关，同时也与样本在各层之间的分配相联系。设第 i 层样本的比例为ω_i，则样本容量为 n 时第 i 层的样本容量为 $n_i=n\omega_i$，代入分层抽样方差公式可以求得

$$n=\frac{\dfrac{\Sigma N_i^2S_i^2}{\omega_i}}{N^2V(\bar{y}_{st})+\Sigma N_iS_i^2} \tag{9.17}$$

若为估计比例问题，只需用 p_iq_i 代替 S_i^2 即可。

例 9.11　仍以广告公司调查各居民户每周收看电视的小时数为例，假设分为市区、近郊和远郊，已知各层的户数和方差为

市区　$N_1=155$　$S_1^2=25$

近郊　$N_2=62$　$S_2^2=225$

远郊　$N_3=93$　$S_3^2=100$

若要求估计每周看电视的小时数误差不超过 2 小时，可靠性为 95%，试分别用等比例分配和奈曼分配确定样本容量。

解：要求误差不超过 2 小时，可靠性为 95%，这相当于估计的样本平均数的 2 个标准误等于 2 小时，即$2\sqrt{V(\bar{y}_{st})}=2$，或 $V(\bar{y}_{st})=1$

(1)在等比例分配时

$$\omega_1=\frac{155}{310}, \ \omega_2=\frac{62}{310}, \ \omega_3=\frac{93}{310}$$

因此 $$\Sigma\frac{N_i^2S_i^2}{\omega_i}=\frac{(155)^2(25)}{\frac{155}{310}}+\frac{(62)^2(225)}{\frac{62}{310}}+\frac{(93)^2(100)}{\frac{93}{310}}$$

$$=1201250+4324500+2883000=8408750$$

代入公式(9.17)

$$n=\frac{8408750}{901600+27125}=69$$

$$n_1=69\times\frac{155}{310}=34 \qquad n_2=69\times\frac{62}{310}=14$$

$$n_3=69\times\frac{93}{310}=21$$

(2)在奈曼分配时

$$\omega_1=\frac{N_iS_i}{\Sigma N_iS_i}=\frac{775}{2635}, \ \omega_2=\frac{N_2S_2}{\Sigma N_iS_i}=\frac{930}{2635}, \ \omega_3=\frac{N_3S_3}{\Sigma N_iS_i}=\frac{930}{2635}$$

因此 $$\Sigma\frac{N_i^2S_i^2}{\omega_i}=\frac{(155)^2(25)}{\frac{775}{2635}}+\frac{(62)^2(225)}{\frac{930}{2635}}+\frac{(93)^2(100)}{\frac{930}{2635}}$$

$$=6943225$$

代入公式(9.17)

$$n=\frac{6943225}{123225}=57$$

$$n_1=\frac{775}{2635}(57)=17, \ n_2=\frac{930}{2635}=(57)=20,$$

$$n_3=\frac{930}{2635}(57)=20,$$

§9.4 整群抽样

§9.4.1 整群抽样及其作用

前面介绍的几种抽样方式在抽取样本时都是直接抽取总体的基本单位(或称为元素)，也就是说抽样单位和基本单位是一致的。但是在某种情况下抽样单位与基本单位也可以不一致。例如欲调查某个大学的学生身高，其基本单位是学生，但抽样单位可以是班级或系等，实际是对抽中的班级或系的全部学生作为样本进行观察。由若干个有联系的基本单位所组成的集合称为群，抽样调查时以群为抽样单位抽取样本就称为整群抽样。

1.应用整群抽样的原因。应用整群抽样主要有以下的原因：

(1)缺乏总体单位的抽样框。前面曾经提到，在抽样调查抽取样本时需要有一个抽样框，它是包括所有总体单位的名单或地图，抽样时先要编上号码，这样才能利用随机数目表或其他方式从中抽取样本。然而有些时候总体很大，没有现成的名单，而要编制抽样框也十分困难，有时甚至是不可能的。例如，我们欲调查北京市中学生中近视眼占的比例有多大，就需要全北京市中学生的名单，这是不容易办到的。如果我们以中学作为抽样单位，那么从教育局去抄一张全市中学的名单就要方便得多。

(2)为了工作方便和节约费用,有时即使具备了抽样框也不采用简单随机抽样的方法，这是由于总体包括的范围很大，总体单位分布很广，若用简单随机抽样会使样本的分布很分散，调查时需要花的人力和物力也比较大。在上例中假如我们具有全市中学生的名单，要从数十万中学生中抽取几百人或上千人调查，其抽样的过程也相当麻烦，抽中的学生分布在全布各个中学，进行调查也很费时、费力。若能抽取几个中学，对抽中的中学全部学生进行调查，这样样本单位比较集中，调查就方便得多，费用也比较节省。

2.整群抽样的缺点。整群抽样的缺点是，由于抽取的样本单位比较集中，在一个群内各单位的标志值往往具有相关性，各单位之间的差异较小，而不同群之间的差异则通常较大。因此当抽取同样数量的样本单位时，抽样误差常常太于简单随机抽样，为了达到规定精确度的要求，往往需要多抽一些群。和简单随机抽样一样，群抽得愈多一般来讲就愈精确，然而群抽得太多又不符合整群抽样节约人力物力的目的，因此需要研究一些数量界限，分析在什么情况下应用整群抽样比较有利，群的规模以多大为好等。

3.整群抽样的分群及其原则。整群抽样中的“ 群 ”大致可以分为两类，一类是根据行政、地域以及自然形成的群体，如学校、工厂等。这一类群主要是为了抽样方便和节约费用。另一类群是一个连续的总体，群的大小可以由调查者根据情况来划分，例如一大块面积可以划分成不同面积群，在这种情况下就需要研究如何分群使方差和费用达到最优。

分群的一般原则是这样的：根据方差分析的原理，当总体划分成群以后，总体方差可以分解为群间方差和群内方差两部分，这两部分相互制约，若群内方差大，则群间方差小；反之，群内方差小则群间方差大。由于整群抽样是对抽中群的所有单位进行调查，因此影响整群抽样的误差大小主要是群间方差。要使整群抽样的误差缩小，如果可能的话，分群时应使群内方差尽可能大，而使群间方差尽可能小，这和分层抽样时分层的原则恰恰相反。关于这一点也可以从直观上解释，若一个群中各单位的情况相同，在整群抽样时，虽然一个群包括许多单位，但都是重复的信息。如果群内的差异比较大，群

内各单位的分布与总体分布一样，那么只要任意抽出一个群来进行观察就可以代表总体。当然以上两种只是极端的情况，但是掌握这一原则对于考虑怎样分群是有益的。

§9.4.2　均值和总值的估计推断

1.有关整群抽样的符号请见表 9.5。

表9.5　整群抽样符号

总　　体		样　　本	
N	总体群数	n	样本群数
M_i	总体第i群的总体单位数	M_i	样本第i群的总体单位数
$\overline{M}_N$	总体中群的平均大小	$\overline{M}_n$	样本中群的平均大小
M_0	总体单位总数	$n\overline{M}_n$	样本中单位总数
$\overline{Y}_i$	第i群的标志值单位均值	$\overline{y}_i$	样本中标志值单位均值
Y_i	第i群的标志值总值	y_i	样本第i群标志值总值
Y	总体标志值总量		
$\overline{Y}$	总体标志值单位均值	$\overline{y}$	样本标志值单位均值
Y_{ij}	总体第i群第j个单位标志值	y_{ij}	样本第i群第j个单位标志值
S_b^2	$\frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_i-\overline{Y})^2$ 群间方差	s_b^2	$\frac{1}{n-1}\sum_{i=1}^{n}(\overline{y}_i-\overline{y})^2$ 样本群间方差

2.比率估计的估计量。整群抽样可以采用不同的估计量来估计总体的平均数，一个常用的估计量就是将抽中群的标志值总量相加与总体单位数相加之比

$$\overline{\overline{Y}}=\frac{\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}M_i} \tag{9.18}$$

可以看出，这一估计量的分子和分母均随着样本的不同而变化，是一个比率型的估计量。从统计理论可知，这是一个有偏的估计，但是当样本容量比较大时，其偏误可以忽略不计。另外，如果各群的大小相等，即 M_i 均等于 $\overline{M}_N$ 时

$$\overline{\overline{Y}}=\frac{\sum_{i=1}^{n}y_i}{\Sigma M_i}=\frac{1}{nM}\sum_{i=1}^{n}y_i=\frac{1}{n}\sum_{i=1}^{n}\overline{y}_i \tag{9.19}$$

这样就不成为比卒估计量了，所以偏误也同样消失而成为无偏估计量。比率估计量求平均数的近似方差公式为

$$S^2(\overline{\overline{Y}})=\frac{N-n}{Nn\overline{M}_N^2}\cdot\frac{\sum_{i=1}^{n}(y_i-M_i\overline{y})^2}{n-1}$$

$$=\frac{1-5}{n\overline{M}_N^2}\cdot\frac{\Sigma y_i^2-2\overline{y}\Sigma M_iy_i+\overline{y}^2\Sigma M_i^2}{n-1} \tag{9.20}$$

例 9.12　某社会科学研究工作者欲估计一城镇的居民户平均年收入，采

用整群抽样，以居民小组作为群，该城镇共有 415 个居民小组，从中抽取 25 个作为样本，取得资料如表 9.6。试以 95%的置信度确定该街区人均收人的置信区间。

解：根据公式(9.18)

$$\bar{\bar{Y}} = \frac{\Sigma y_i}{\Sigma M_i} = \frac{132900}{151} = 880.13\text{元}$$

为了计算方差，根据公式(9.20)需要先计算出有关的数字

$$\Sigma y_i^2 = 820390000, \Sigma M_i y_i = 840300$$

所以 $$\sum_{i=1}^{n}(y_i - M_i\bar{y}_i)^2 = 820390000 - 2(880.13)(840300)$$

$$+(88.013)^2(1047) = 152279893.3$$

表 9.6　　某城镇居民户收入抽样资料

群（i）	居民户（Mi）	总收入（yi）	群（i）	居民户（Mi）	总收入（Yi）
1	8	9600	14	10	4900
2	12	12100	15	9	5300
3	4	4200	16	3	5000
4	5	6500	17	6	3200
5	6	5200	18	5	2200
6	6	4000	19	5	4500
7	7	7500	20	4	3700
8	5	6500	21	6	5100
9	8	4500	22	8	3000
10	3	5000	23	7	3900
11	2	8500	24	3	4700
12	6	4300	25	8	4100
13	5	5400	合计	151	132900

由于 $\overline{M}_N$ 未知，用 $\overline{M}_n = \frac{\Sigma M_i}{n} = 6.04$ 来估计，

代入公式(9.20)

$$S^2(\bar{y}) = \frac{415-25}{(415)(25)(6.04)^2}\left(\frac{152279893.3}{24}\right) = 6537.8$$

S($\bar{y}$)=80.85，所以估计该城镇每户年平均收入为

880.13±2(80.85)，即在 718.43—1041.83 元之间。

关于总值的估计在利用比率估计量时还需要知道总体的单位数 M_0，在已知 M_0 的基础上可以利用以下公式

$$\bar{\bar{Y}} = M_0 \frac{\Sigma y_i}{\Sigma M_i} \tag{9.21}$$

$$S^2(\bar{\bar{Y}}) = M_0^2 \frac{1-f}{n\overline{M}^2} \cdot \frac{\Sigma(y_i - M_i\bar{y})^2}{n-1}$$

$$= \frac{N^2(1-f)}{n} \cdot \frac{\Sigma(y_i - M_i\bar{y})^2}{n-1} \tag{9.22}$$

因此在上例中若已知该城镇共有 2500 户，欲估计全镇居民户的总收入即可利用此公式，这一结果留给读者自己计算。

3.无偏估计的估计量。前面的整群抽样估计公式，在估计总体总值时，

需要知道总体的单位数 M_0，但是使用整群抽样的原因之一往往是没有总体单位的抽样框，有时总体单位数也是未知的，但总体的群数 N 是已知的，这时对总体总值的估计可以采用另一个公式，令 $\bar{y}_t=\frac{1}{n}\sum_{i=1}^{n}y_i$，为样本群的均值，则总值的估计量及其方差公式为

$$\hat{Y}=N\bar{y}_i=\frac{N}{n}\sum_{i=1}^{n}y_i \tag{9.23}$$

$$S^2(\hat{Y})=N^2\frac{(N-n)}{Nn}\cdot\frac{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}{n-1} \tag{9.24}$$

显然，这种估计方法是以群为单位的简单随机抽样，它不考虑群的大小，但却是一个无偏估计量，另外这种方法在估计单位均值时应该知道总体单位数 M_0 或 $\overline{M}_N$。这两种方法采用哪一种方法较好，一方面取决于是否知道 M_0，另一方面要看群的大小和标志值之间的相关程度。当相关程度比较高时，比率估计的效率就比较高，如果相关程度比较低，群的大小就不能提供什么信息，在这种情况下就采用以群为单位的简单随机抽样方法为好。如果总体各群的容量相等，即 $M_1=M_2=\overline{M}_N$，那么这两种估计方法是等价的。

例 9.13　邮局欲估计辖区内每个家庭平均订报份数，该辖区共有 4000 户，划分为 400 个群，每群 10 户，现随机抽 4 个群，取得资料如表 9.7。

表 9.7　　4 个样本群订报数

群	订	报	数								y_i	$\bar{y}$
1	1	2	1	3	3	2	1	4	1	1	19	1.9
2	1	3	2	2	3	1	4	1	1	2	20	2
3	2	1	1	1	1	3	2	1	3	1	16	1.6
4	1	1	3	2	1	5	1	2	3	1	20	2

试对平均每户家庭的订报份数及总的订报份数作出区间估计，置信度为 95%。

解：(1)根据比率估计量[公式(9.18)或(9.19)]

$$\bar{y}=\frac{\Sigma y_i}{\Sigma M_i}=\frac{19+20+16+20}{4(10)}=1.875$$

根据公式(9.20)

$$S^2(\bar{y})=\frac{N-n}{Nn\overline{M}_N^2}\frac{\Sigma(y_i-M_i\bar{y})^2}{n-1}$$

$$=\frac{400-4}{400(4)(10)^2}\times\frac{10.75}{3}=0.0089$$

即平均每户订报的区间估计为 $1.875\pm2\sqrt{0.0089}$ 份或 1.69—2.06 份。其总的订报份数为

$$\hat{Y}=M_0\bar{y}=4000(1.875)=7500\text{份}$$

$$S^2(\hat{Y})=M_0^2V(\bar{y})=141900\text{区间估计为}6760-8254\text{份}$$

(2)由于本例中群的大小相等，因此也可以用无偏估计的 方法计算，根

据公式(9.23)，(9.24)。

$$\bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_i = 18.75\text{，}\ \bar{y} = \frac{\bar{y}_t}{\bar{M}} = 1.875$$

$$\hat{\bar{Y}} = N\bar{y}_t = 400 \times 18.75 = 7500$$

$$S^2(\hat{\bar{Y}}) = \frac{N^2(N-n)}{N_n} \cdot \frac{\Sigma(y_i - \bar{y}_t)^2}{n-1} = 141900$$

其估计值及方差都与比率估计的结果相同。

4.按规模大小戍比例的抽样。当群的规模大小不同时，如果群的标志值总量 Y_i 与群的规模之间存在高度相关时，按群的大小成比例抽样可以取得比较好的效果，这是一种不等概率的抽样。当第 i 群的大小为 M_1，总体的总单位数为 M_0，令该群被抽中的概率为 $\pi_i = \frac{M_i}{M_0}$。这种抽样方法，根据样本来估计总体平均数及其方差比较简单，其形式和简单随机抽样一样，称作自动加权，即

$$\hat{\bar{Y}}_{pps} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_i\text{，} \tag{9.25}$$

$$S^2(\hat{\bar{Y}}_{pps}) = \frac{1}{n(n-1)}\sum_{i=1}^{n} (\bar{y}_i - \bar{y}_{pps}) \tag{9.26}$$

例 9.14　一个大工厂欲估计上一季度每个工人的平均病假无数，该厂共有 8 个车间，人数分别为 1200，450，2100，860，2840，1910，290 及 3200 人，现用整群抽样抽取 3 个车间，每个车间被抽中的概率使之与该车间工人人数成比例。假设抽中第三车间，工人有 2100 人，病假共 4320 天；第六车间 1910 人，病假共 4160 天，第八车间 3200 人，病假共 5790 天。要求估计全厂工人的平均每人病假天数。

解：先求样本群的平均每人病假天数

$$\bar{y}_1 = \frac{4320}{2100} = 2.06\text{，}\ \bar{y}_2 = \frac{4160}{1910} = 2.18\text{，}\ \bar{y}_3 = \frac{5790}{3200} = 1.81$$

根据公式 (9.25)　$\hat{\bar{Y}}_{pps} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_i = 2.02$

根据公式 (9.26)

$$S^2(\hat{\bar{Y}}_{pps}) = \frac{1}{n(n-1)}\sum_{i=1}^{n} (\bar{y}_i - \hat{\bar{Y}}_{pps})^2 = 0.0119$$

全厂平均每人病假天数的置信区间为 2.02±0.22 天。

按照与群中单位多少成比例的抽样方法主要有两种：

一种是汉森与赫维茨于 1943 年提出的,其方法是将总体各群按一定顺序排列，然后根据其规模 M_i 进行累计，最后累计数为 M_0。于是利用随机数字表在 1—M_0 之间取一个数，设为 r，若 $\sum_{i=1}^{j=1} M_i < r < \sum_{i=1}^{n} M_i$ 则第 j 个群被抽中，并继续抽取直到抽满所需的样本容量为止。现结合一个例子来说明。假设 N＝7，共包括 30 个单位即 M_0＝30，现将各群按一定顺序排列并累计如表 9.8。

表 9.8　　　　　　　　按单位数多少成比例抽样计算表

群（i）	单位数（M_i）	累计 ΣM_i	累计范围
1	3	3	1——3
2	1	4	4
3	11	15	5——15
4	6	21	16——21
5	4	25	22——25
6	2	27	26——27
7	3	30	28——30

在抽选样本时从 1—30 之间取一个随机数，假设为 19，落入第 4 个群的累计范围之内，因此第 4 个群即为抽中的单位，显然采用这种抽选方法时，每个单位抽中的概率都和该单位的大小成比例，而且是重复抽样。

另一种方法是由拉希里提出的。因为汉森与赫维茨提出的方法在 N 很大时，累计的工作量就很大。而拉希里的方法可以免去累计，但需要抽选两次来决定是否抽中，首先是在 1—N 之间抽取一个随机数，假定抽中的数为 i，今总体中最大的群的大小为 M_{max}，因此第二次是从 1—M_{max} 之间抽一个随机数，假设为 r，当 r＜M_i 时，第 i 个单位就算被抽中；若 r ＞M_i，则第 i 个单位未被抽中需要重新在 1—N 之间抽取。这样继续下去，凡两次都符合条件才算抽中单位，直到抽满所需样本为止。

§9.4.3　总体比例的估计

假如要估计的是总体的比例，即总体单位中具有某种特征的单位数 A 与总体单位之比，在整群抽样中今 a_i 表示第 i 个群中具有某种特征的单位数，则估计总体比例的估计量及其方差的公式为

$$P = \frac{\Sigma a_i}{\Sigma M_i} \tag{9.27}$$

$$S^2(P) = \left(\frac{N-n}{Nn\overline{M}^2}\right)\left(\frac{\Sigma(a_i - M_i P)^2}{n-1}\right) \tag{9.28}$$

可以看出，整群抽样的比例估计也是一个比率型的估计量，一般是个有偏估计，因而要求样本的容量比较大。如果各群的容量相同，即 M_i 为一常数时，比例的估计量也是一个无偏估计。

例 9.15　在例 9.12 中，若需要研究该城镇中具有老人的户所占的比例时，在这 25 个居民小组中取得资料如表 9.9。试估计总体比例的 95%置信区间。

解：根据公式(9.27)

表 9.9　　　　　　　　某场面镇居民户中有老人户的资料

群（i）	居民户（M_i）	有老人户（a_i）	群（i）	居民户（M_i）	有老人(a_i)
1	8	4	14	10	5
2	12	7	15	9	4
3	4	1	16	3	1
4	5	3	17	6	4
5	6	3	18	5	2
6	6	4	19	3	3
7	7	4	20	4	1
8	5	2	21	6	3
9	8	3	22	8	3
10	3	2	23	7	4
11	2	1	24	3	0
12	6	3	25	8	3
13	5	2	**合计**	151	72

$$p=\frac{\Sigma a_i}{\Sigma M_i}=\frac{72}{151}=0.477$$

在计算方差时需要知道群的平均大小$\overline{M}_N$，在未知时需用样本$\overline{M}$来估计$\overline{M}_n=\frac{\Sigma M_i}{n}=6.04$，根据公式(9.28)

$$S^2(p)=\frac{N-n}{Nn\overline{M}^2}\cdot\frac{\Sigma(a_i-M_ip)^2}{n-1}=0.00055$$

所以当置信度为 95%时该城镇户中有老人的户约占 0.48%±0.05%，即在 0.43—0.53%之间。

§9.5 系统抽样

§9.5.1 系统抽样及其作用

1.关于系统抽样。系统抽样也称机械抽样或等距抽样。系统抽样是对研究的总体按一定的顺序排列，每隔一定的间隔抽取一个单位，并把这些抽取的单位组成样本进行观察用以推断总体。例如有一叠发票，我们欲估计共平均每张的金额以及总金额，由于发票太多，全面计量比较费时费力，可以从中每隔 10 张抽取 1 张，然后将抽中的发票组成样本进行观察，用这一样本的均值来估计推断总体均值。如果知道发票的总量也就可以推断其总值。

2.系统抽样的作用。系统抽样有以下作用：

(1)简便易行。从简单随机抽佯来说，在抽样之前需要对每一个单位加以编号，然后才能利用随机数字表等方法抽选样本。当总体单位数量很多时，编号与抽选的过程也比较麻烦，而等距抽样只要确定了抽样的起点和间隔，整个样本的所有单位也随之而自然确定。它可以充分利用现成的各种排列，例如某市的工矿企业可以按照原有的行政系统及分部门的习惯顺序排列，抽样时就可以直接利用这些顺序进行系统抽样。这种抽样很方便，也便于推广，为不熟悉抽样调查的人员所掌握，在多阶段抽样时，有利于检查抽样过程是否合乎要求。这种方法也适合于某些基层现场的抽样，例如在森林调查中往往很难在林地中划分抽样单位随机抽选，而系统抽样就比较方便。

(2)系统抽样的误差大小与总体单位的排列顺序有关。因此当对总体的结构有一定的了解时，可以利用已有的信息对总体单位进行排列后再进行系统抽样，就能提高抽样效率。在一般的情况下系统抽样使样本单位在总体单位中散布比较均匀，其估计量的方差要小于简单随机抽样。因此这是大规模抽样调查时一种比较常用的抽样方法。

3.系统抽样的分类。系统抽样根据总体单位的排列情况大致可以分为三类：

(1)按无关标志排队，即排列的顺序和所研究的标志是无关的。例如调查某工厂职工的平均年龄时，是按姓氏笔划排列的职工名单进行抽样的。显然，年龄和姓氏笔划之间没有必然的联系，因此这种排列的抽样也称为无序系统抽样，可以比作是随帆排列的。

(2)按有关标志排队。例如在进行农产量调查时，是将总体单位按当年估产或上年的平均自产的高低顺序排列的。这种按有关标志排列的系统抽样也称为有序的系统抽样。这种抽样的方法可以使标志值高低不同的单位，均有代表性的单位选人样本，因而抽样的误差比较小。

(3)根据各单位在总体中原有的自然位置顺序排列。这种排列通常处于上述两种情况之间。例如工厂中的工人名单按原有的工资名册顺序，工业产品质量检验时按生产的时间顺序每隔一定时间抽取一定数量的样本等。这种自然状态的排列，有时与调查的标志有一定的联系，但又不完全一致，这往往是为了抽样的方便。

§9.5.2 系统抽样的方法

1.直线等距抽样。设总体单位数为 N，欲抽取的样本容量为 n。当 N 为 n 的整倍数时，先算出系统抽样的间隔 k，$k=\frac{N}{n}$，也称为抽样距离。这种方法实际上是把 N 个单位排成一直线后分成 n 段，每段中有 k 个单位，从每段

中抽取一个样本单位。抽选时先在 1—k 中随机抽选一个数目，设为 i，则第 i 个单位为抽中单位，以后每隔 k 个单位为一抽中单位，即 i＋k，i＋2k…直到抽满 n 个数目为止。其相应的单位即为所要抽取的样本。可用图 9.3 来表示。

图 9.3　直线等距抽样示意图

例 9.16　某县共有 300 个村，欲抽 15 个村为样本，试用直线等距抽样方法抽取一个随机样本。

解：已知 N＝300，n＝15，可求出抽样间隔 $k=\frac{300}{15}=20$，于是在 1—20 之间随机抽取一个数字 i，设 i＝6，则第 6 个单位即为抽中单位，其余的样本单位为 26，46，…，286 共 15 个村组成样本。

上述系统抽样方法的第一个单位是随机确定的，因此称为随机定位直线系统抽样。当总体单位的排列是按有关标志排列时，其标志值往往是由低到高或由高到低形成一种趋势。因此第一个单位的位置偏高或偏低也会影响到以后的各单位，整个样本就会产生偏误。为了减少这种偏误，有人提出 i 取 1—k 的中间数值，当 k 为奇数时，$i=\frac{k+1}{2}$；当 k 为偶数时，$i=\frac{k+2}{2}$ 或 $\frac{k}{2}$。这种抽样方法称为中心定位系统抽样。

当总体单位的排列顺序呈现线性趋势时，采用中心定位系统抽样固然可以防止偏误，但是对该总体来说，排队确定后，只能抽取一个固定样本，失去了随机性。因此有许多统计学家提出了改进的方法，其中简单易行且效果较好的是塞蒂和辛提出的对称等距抽样，这种方法在我国国家统计局主持的有关抽样调查中得到广泛的应用。

塞蒂提出的方法是当 N＝nk，n 为偶数时，要从 N 个单位中抽取 n 个样本单位，就先将总体 N 个单位分成 $\frac{n}{2}$ 个组，使得每个组包含 2k 个单位，然后在 1—k 中随机地确定抽样起点，按下面的公式在每组中抽取距该组两端等距离的两个单位。

$$[i+2jk,2(j+1)k-i+1] \quad j=0,1,\cdots,\frac{n}{2}-1 \tag{9.29}$$

当 n 为奇数时，仍按上式进行，但 j 取值到 $\frac{n-1}{2}-1$ 为止，并要增加靠近终端的一个样本单位[i＋(n-1)k]。

例 9.17　根据例 9.15 提供的情况用塞蒂对称等距抽样方法抽取一个随机样本。

解：已知 N＝300，n＝15，k＝20，i＝6，根据公式计算出应抽取的单位号码为

j=0,i+2jk=6,2(j+1)k-i+1=35；

j=1,i+2jk=46,2(j+1)k-i+1=75；

M（由于n为奇数）

j=$\frac{n-1}{2}-1$,i+2jk=246,2(j+1)k-i+1=275；

j+(n-1)k=286

因此所抽的样本为(6，35，46，75，…，246，275，286)。前面的样本单位都是对称的，由于 n 是奇数，最后的样本单位是不对称的。

辛在塞蒂方法的基础上作了修正，所以也称作修正的对称等距方法。其不同点是抽选对总体两端等距离的成对单位，当随机起点为 i，n 是偶数时，其成对的对称单位是

$$[j + jk；(N = jk) - i + 1] \qquad j = 0,1,\cdots,\frac{n}{2} - 1。 \tag{9.30}$$

当 n 为奇数时则更加靠近中间的$[i+\frac{1}{2}(n-1)k]$。

例 9.18　仍用例 9.16 的情况用辛的修正对称等距抽样方法抽取样本。

解：根据公式(9.30)

j=0 时　j+jk=6，(N-jk)-i+1=295

j=1 时　i+jk=26，(N-jk)-i+1)=275

M

$$j = \frac{n-3}{2}\text{时} \quad j + jk = 126，(N - jk) - i + 1 = 175$$

因为 n 是奇数，还应增加一个单位

$$i+\frac{1}{2}(n-1)k=146$$

2.循环等距抽样。当 N 为有限总体而且 N 不能被 n 所整除，这时 k 不是一个整数。如果将 k 取一个比较接近的整数，样本的容量可能不一定是 n，样本的估计值可能产生偏误。为了使样本容量始终为 n，且估计量为无偏估计，可以采取循环等距抽样。它是将总体各单位排成首尾相接的循环圆形，用同样的方法确定抽样间隔 k，k 可以取最接近的整数。所不同的是在 1—N 中抽取一个随机起点，然后每隔 k 个单位抽取一个，直到抽满 n 个样本单位为止。

例 9.19　某总体单位 N＝23，要求抽取 n＝5 的等距样本，样本的估计量为总体的无偏估计。

解：可采用循环等距抽样方法，因为$k = \frac{23}{5} = 4.6$，所以可以取接近的整数 5，将总体单位排列成一个首尾相接的圆形如图 9.4，然后在 1—23 中抽取一个随机数字 i，假设 i＝16，则 i＋k 总体一共只有 23 个单位，按照循环的排列第 26 号单位实际上是 26-23＝3 号单位；31 号是总体的第 31-23＝8 号单位；36 号是第 36-23＝13 号单位。

图 9.4　循环等距抽样示意图

§9.5.3　系统抽样的估计和推断

1.总体平均数的估计。系统抽样时可以以样本平均数来估计总体平均数，即

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{9.31}$$

但系统抽样估计量的方差大小与总体中各单位的排列状况有关。若用 ρ 表示组内相关系数，反映系统样本内任意两个单位之间的相关系数，则系统抽样均值的方差为

$$V(\bar{y}_{sy}) = \frac{\sigma^2}{n}[1+(n-1)\rho] \qquad (9.32)$$

(1)如果总体单位是按随机顺序排列的，则可以认为ρ＝0，这样系统抽样的方差也就相当于简单随机抽样，可以按简单随机抽样的公式计算估计量的方差。

(2)当总体各单位的排列顺序具有线性趋势时,每个系统样本内既有标志值低的单位，也有标志值高的单位，这时往往ρ＜0，因此系统抽样的方差就会小于简单随机抽样。

(3)当总体各单位标志值存在周期变动趋势,而抽样的间隔又恰好与周期的变动相一致时，系统样本中各单位的标志值会比较一致，因此ρ＞0，这样系统抽样的方差将会大于简单随机抽样。

例 9.20　某果园欲估计每株果树的平均产量，采取在果园中每隔 7 株抽取 1 株的系统方法一共抽了 212 株果树作样本。对抽中的果树进行实际称重并得$\sum_{i=1}^{212} y_i$＝17066 公斤，$\sum_{i=1}^{212} y_i^2$＝1486800。求每株树的平均产量并确定 95%的置信区间。

解：在果园中果树的排列顺序可以认为与产量是无关的，则可以采用简单随机抽样的公式进行估计，因此

$$\bar{\bar{Y}} = \bar{y}_{sy} = \frac{1}{n}\sum_{i=1}^{212} y_i = \frac{1}{212}(17066) = 80.5(\text{公斤})$$

由于

$$s^2 = \Sigma(y_i - \bar{y})^2 / n-1 = \frac{1486800 - 212(80.5)^2}{211} = 535.48$$

$$s^2(\bar{y}_{sy}) = \frac{1-f}{n}s^2 = \frac{(1-1/7)}{212}(535.48) = 2.16$$

所以平均每株果树的产量的 95%置信区间为

$\bar{y}_{sy} \pm 2s(\bar{y}_{sy})$=80.5±2.9，即在 77.6 公斤到 83.4 公斤之间。

3.总体比例的估计。系统抽样也常常用于对总体比例的估计，其原理和平均数的估计一样。如果总体单位是随机排列的可以用简单随机抽样估计比例的公式

$$p_{sy} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{其中：} y_1 = \begin{cases} 1 & \text{具有某种特征} \\ 0 & \text{其他} \end{cases} \qquad (9.33)$$

$$s^2(p_{sy}) = \frac{p(1-p)}{n-1}(1-f) \qquad (9.34)$$

例 9.21　某粮店供应辖区内共 5775 户，现欲变更营业时间需征求这些户的意见，采取每隔 6 户抽取一户的办法，共抽选了 962 户，调查结果同意的为 652 户。试估计这一总体中同意的占多大比例。

解：已知 N=5775，n=962　Σy_i=652，

根据公式(9.33)和(9.34)

$$\bar{p}_{sy}=\frac{1}{n}\Sigma y_i=\frac{652}{962}=0.678$$

$$S^2(\bar{p}_{sy})=\frac{\bar{p}(1-\bar{p})}{n-1}(1-f)$$

$$=\frac{(0.678)(0.322)}{961}(1-\frac{1}{6})=0.00019$$

因此可以推断总体中同意的比例为$0.678\pm 2\sqrt{0.00019}$即在 65%到 70.6%之间，置信系数为 95%。

4.重复系统抽样。前面曾经提到，当总体中存在周期性变动，而且抽样的间隔为周期变动的整倍数时会降低系统抽样的效率。解决的办法可以在同一总体中重复抽取多个系统样本来对总体进行估计。设总体单位数为 N，原准备抽一个系统样本，其容量为 n，则抽样间隔为$k=\frac{N}{n}$。现改为抽于 n_s 个等容量的系统样本，则新的间隔应为 k'=n_sk，于是在 1——k'之间随机抽取 n_s 个随机数，得到 n_s 个系统样本。然后根据 n_s 个估计值作出对总体的估计值和抽样方差。这种方法称作重复系统抽样，其总体均值及其估计量方差的公式为

$$\bar{\bar{Y}}=\frac{1}{n_s}\sum_{i=1}^{n^s}\bar{y}_i \qquad (9.35)$$

$$s^2(\bar{\bar{Y}})=\frac{s_b^2}{n_s}(1-f),\text{其中}s_b^2=\frac{\sum_{i=1}^{n^s}(\bar{y}_i-\bar{\bar{Y}})^2}{n_s-1} \qquad (9.36)$$

例 9.22 有一个周期变动的总体，N＝240，原来准备抽取一个系统样本，n＝24，现为了消除周期变动的影响，改为抽 4 个容量为 6 的样本，所以 n_s=4，原来的抽样间隔$k=\frac{240}{24}=10$，现 k'=n_sk=4(10)=40，根据 4 个随机起点分别间隔 40 抽取一个单位，就得到 4 个容量为 6 的样本。若调查结果得到 $\bar{y}_1$＝19.17，$\bar{y}_2$=$\bar{y}_3$=$\bar{y}_4$=18.17。试估计总体的均值及其估计标准误。

解：根据公式(9.35)、(9.36)

$$\bar{\bar{Y}}=\frac{1}{n_s}\Sigma\bar{y}_i=\frac{1}{4}(19.17+18.17\times 3)=18.42$$

$$s_b^2=\Sigma(\bar{y}_i-\bar{Y})^2$$

$$=\frac{1}{4-1}[(19.17-18.42)^2+3(18.17-18.42^2)]=0.25$$

$$S^2(\bar{\bar{Y}})=\frac{{s_b}^2}{n_s}(1-f)=\frac{0.25}{4}(1-0.1)=0.05625$$

因此估计总体的均值为 18.42，估计标准误为$\sqrt{0.05625}$=0.237。

习　题

1.某工厂欲估计某一项作业所需的平均操作时间，从 98 人中随机抽取 8 人，记录他们的时间为 4.2，5.1，7.9，3.8，5.3，4.6，5.1，4.1。若操作时间在工人中呈正态分布，试以 95%的置信水平估计总体平均所需时间。

2.一社会学家研究在某一小城镇的家庭中二代以上居住在一起的户占全镇户数的比例，该镇共有 621 户，用简单随机抽样抽取 60 户，其中有 10 户系二代以上居住在一起，试估计全镇中有多少户二代以上居住在一起，并计算抽样标准误；若要求估计比例的误差不超过 5%,则应抽取多少户作样本？

3.某工厂希望估计某个月内由于事故引起的工时损失，由于工人、技术人员及行政管理人员的事故率是不同的，因而采用分层抽样，已知以下资料

工人	技术人员	管理人员
N1=132	N2=92	N3=27
σ_1^2=36	σ_2^2=25	σ_3^2=9

若用奈曼分配确定各层样本，样本容量 n＝30，调查结果 $\bar{y}_1$ ＝试估计总的工时损失数及其 95%的置信区间。

4.某工厂在全国有 452 家经销点，现想了解刊登广告以后销售量的变化，抽取了 20 个点作样本，取得广告前后两个季度的销售量资料如表 1。

表 1

样本	广告前	广告后	样本	广告前	广告后
1	208	239	11	599	626
2	400	428	12	510	538
3	440	472	13	828	888
4	259	276	14	473	510
5	351	363	15	924	998
6	880	942	16	110	171
7	273	294	17	829	889
8	487	514	18	257	265
.9	183	195	19	388	419
10	863	897	20	244	257

若已知广告前一季度的总销售量为 216256,试估计广告后一季度销售量的 95%置信区间。

5.汽车公司抽样检查在使用的车辆中不安全轮胎的比例数，在 175 辆中抽了 25 辆,其不安全轮胎数分别为试估计不安全轮胎的比例及估计的允许误差。(a＝0.05)

6.一个班级共有 40 个学生，分成 4 个学习小组，按照 1，2，3，4 的顺序排列，在一次考试中各组按考试成绩由低到高排列，因此不及格的人列在前面，如表 2。

(1)用每隔 10 人抽取：人的系统抽样方法，列出所有可能的样本，并计算不及格人数的比例及抽样方差；(2)用每隔 5 人抽取、人的系统抽样方法，列出所有可能样本，计算不及格人数的比例及抽样方差。(3)若已知总体比例为 0.4，分别以样本容量为 n＝4 和 n＝8 计算简单随机抽样的方差。与前面的结果相比较，你能得出什么结论？

表 2

	第1组	第2组	第3组	第4组
人数编号	1——11	12——20	21——28	29——40
不及格人数	4	3	5	4

7.卫生部门对 32 个城市的饮食业抽样检查卫生情况，采用二阶段抽样，第一阶段抽取 4 个城市，第二阶段在每个城市抽取一半饮食店进行检查，检查不合格的店所占比例，检查结果如表 3。要求估计推断这 32 个城市中不合卫生要求的饮食店所占的比例及 95%的置信区间。

表 3

城　市	城市饮食店	样本店	卫生不合格店数
1	25	13	4
2	10	5	1
3	18	9	4
4	16	8	2

第 10 章　统计质量管理

§10.1　引　言

质量是企业的生命，也是一切工作的根本。质量管理也称为质量学(Qualimetry)，是概卒论与数理统计在质量管理方面的应用。

对产品质量进行管理，早在奴隶社会和封建社会就已经开始，主要表现为弓、箭等产品和零部件的互换。进入资本主义社会以后，随着大工业生产的发展，对产品规格要求越来越精密。在 1840 年前后，对零件精度做出了公差的规定。进入 20 世纪以后，首先在工业部门中，要求实行标准化的生产。不仅要求各国国内标准化，而且要求国际上实行标准化。这时，产品质量虽有一些改善，但还不能预防次品的出现。对于产品质量的检验采取的仍是全部检验，特别是对于武器、弹药等破坏性检验的产品还没有较好的检验办法。1924 年，美国贝尔电话实验室的工程师休哈特(W.A.Shewhart)利用统计正态分布和小概率的原理，提出应用统计质量控制图对生产过程中的产品质量进行控制，及时发现生产异常情况，达到了预防次品发生的目的。这一方法的应用，极大地提高了产品质量水平，也开创了质量管理学。

本世纪初，t 统计量的出现开创了小样本统计推断的新纪元。在本世纪 30 年代，内曼(J.Neyman)和皮尔逊(E.S.Pearson)在前人研究成果的基础上提出了置信区间估计和假设检验的方怯。将这些统计方法应用在产品质量检验中，出现了用样本检验代替产品全部检验的新方法。第二次世界大战后，沃尔德(A.Wald)又提出了减少破坏性检验样本容量的序贯抽样检验方法。产品质量抽样控制和检验的方法起源于美国，并在 50 年代介绍到日本。日本企业界在学习应用生产中的控制和生产后的检验的同时,从 60 年代初开始生产前试验设计方法的研究和推广，并逐步形成了系统的试验设计理论。由于日本抓住了生产前质量提高的关键环节，真正起到了预防的作用，使日本产品的质量得到了根本的改善，在国际上逐渐树立了高质量的信誉，我们将在这一章里，分别介绍统计方法在生产前、生产中和生产后质量管理的应用。

§10.2　产品质量的试验设计

试验设计(Design of Experiments)是 20 年代费雪在农业试验中提出的，并与方差分析方法相结合成为生物学、遗传学、农学研究的科学方法。50—60 年代时，日本的田口玄一博士将正交试验设计方法应用于科学管理，逐渐普及了这一科学并极大地提高了产品质量水平。

§10.2.1　试验设计的基本概念

1.指标、因素与水平。试验设计是对试验方案进行最优设计，以降低试验误差和生产费用，减少试验工作量，并对试验结果进行科学分析。

在试验中用来衡量试验结果的量称为试验指标。例如在小麦增产试验中，小麦产量即为试验指标。在科学试验中，对试验指标产生影响的称为影响因素或因素，例如对小麦产量产生影响的品种、播种时间、播种方式、耕作方式、施肥方式、施肥量等都是因素。以上这些因素都是可以人为控制的，这类因素称为可控因素。还有一类人们不能控制的因素，如小麦生长期的温度、雨量等，这类因素称为不可控因素。在试验中，我们总是尽量减少不可控因素，或者想办法固定不可控因素，以将不可控因素变为可控因素。试验中因素变化的各种状态称为水平。如小麦施肥方式这一因素，可供选择的氮肥、磷肥、钾肥等都是施肥方式的水平。我们看到，以上试验设计的概念与方差分析的基本概念是相同的，因为这两种方法本是同一问题的不同研究侧面。

例 10.1　某化工厂生产一种化工产品，采收率低并且不稳定，一般在 60—80%间波动。希望通过试验设计，摸清该种产品的生产规律，从而找出最好的生产方案。

解：在明确了试验目的之后，就要对因素和水平进行分析和挑选。在确定因素和水平时，要挑选那些对试验结果影响明显且不清楚如何影响的因素和水平。本例中经过分析，挑选出反应温度、加碱量和催化剂种类 3 个因素，每个因素各 3 个水平进行试验分析，可得到因素与水平表如表 10.1。

表 10.1　　　　　　　　因素与水平表

水　平	因	素	
	A 反应温度（℃）	B 加碱量（公斤）	C 催化剂种类
1	A_1=80	B_1=35	C_1=甲
2	A_2=85	B_2=48	C_2=乙
3	A_3=90	B_3=55	C_3=丙

2.正交表。对于例 10.1 中这 3 个因素 3 个水平的试验，如要从中找到最优搭配，就要从 3×3×3＝33＝27 种搭配中进行挑选。若要做全面试验的搭配需要做 27 次试验。如果是 4 个因素，每个因素为 3 个水平，其全面试验需做 3×3×3×3＝34　＝81 次。正交试验就是用尽量少的试验次数以获取尽可能多信息的科学方法，其工具是应用正交表来安排试验方案。对于 4 个因素，每个因素 3 个水平的试验可用正交表如表 10.2。

表头上的 L 表示正交表，L 的下标 9 表示进行 9 次试验，括

表 10.2　　　　　　　　　　$L_9(3^4)$

试验号	列号（因素）1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

号内 3 表示每个因素最多安排 3 个水平，3 的指数 4 表示最多安排 4 个因素。由表 10.2 我们可以归纳出正交表具有下面两个性质：

(1)每列中不同水平出现的次数相同；

(2)任意两列(2 因素)交叉试验的次数相同，即任意两列的交叉同样是 11，12，21，22，13，31，23，32，33。其他的正交表还有。等，所有正交表都有以上两条性质。

§10.2.2　正交试验设计

1.用正交表安排试验方案。例 10.1 是 3 个因素 3 个水平的试验，现成的正交表中没有刚好安排 3^3 的表式，可用 $L_9(3^4)$进行安排，使用时去掉 $L_9(3^4)$任意一列即可。现在我们试验时假定去掉第 4 列，利用 $L_9(3^4)$正交表安排例 10.1 的方案如表 10.3。

表中每一横行表示一次试验。1 号试验简记为 $A_1B_1C_1$，表示 A、B、C3 个因素都用第 1 个水平进行试验。5 号试验记为 $A_2B_2C_3$，其他各号试验也都是以因素的下标表示水平号数。

表10.3　　　　　　某化工产品采必率试验方案表

试验号	因素（列号）A 反应温度（℃）	B 加碱量（公斤）	C 催化剂种类
1	1（80）	1（35）	1（甲）
2	1（80）	2（48）	2（乙）
3	1（80）	3（55）	3（丙）
4	2（85）	1（35）	2（乙）
5	2（85）	2（48）	3（丙）
6	2（85）	3（55）	1（甲）
7	3（90）	1（35）	3（丙）
8	3（90）	2（48）	1（甲）
9	3（90）	3（55）	2（乙）

2.试验结果分析。在利用正交表安排好方案后，就可以按方案进行试验了。该例的 9 次试验结果见表 10.4 的最右列，计算与分析过程见表 10.4 方案下的计算表。

表10.4　　　　　　　　　某化工产品采收率试验分析表

试验号	因素 A 反应温度（℃）	因素 B 加碱量（公斤）	因素 C 催化剂种类	试验结果(指标) 采收率（%）
1	1（80）	1（35）	1（甲）	51
2	1	2（48）	2（乙）	71
3	1	3（55）	3（丙）	58
4	2（85）	1	2	82
5	2	2	3	69
6	2	3	1	59
7	3（90）	1	3	77
8	3	2	1	85
9	3	3	2	84

续前表

试验号	因素 A 反应温度（℃）	因素 B 加碱量（公斤）	因素 C 催化剂种类	试验结果（指标） 采收率（%）
K_1	180	210	195	
K_2	210	225	237	
K_3	246	201	204	
$\overline{K}_1$	60	70	65	
$\overline{K}_2$	70	75	79	
$\overline{K}_3$	82	67	68	
R	22	8	14	

从表 10.4 中的 9 次试验结果看，第 8 号试验 $A_3B_2C_1$ 的采收率最高，为 85%；第 9 号试验的采收率其次，为 84%等等。但第 8 号试验方案不一定是我们所要寻找的最优搭配，因为该方案只试验了 9 次，仅是我们要做的 33＝27 次的 1/3，即我们的正交试验只是全部试验的一个样本。然而，利用正交表安排的试验样本不是完全随机的样本，而是一个可通过样本找出最优搭配的最好样本。通过正交试验寻找最优搭配的方法是：

首先，计算平均采收率。表中 K_1、K_2 和 K_3 中各数据是该列因素某水平 3 个试验结果之和。如 K_1 中第 1 列(温度)的 180 是 A 因素 1 号水平 3 个试验结果之和，即 51＋71＋58＝180。又如 K_2 中第 3 列(催化剂种类)的数据 237 是 C 因素 2 号水平 3 个试验结果之和，即 71＋82＋84＝237。其他各个数字都是按此方法得出的。在计算出各因素、各水平采收率之和 K_1、K_2 和 K_3 后，对这些总和数据均除以 3(即求和中试验结果个数)，得到下面的平均采收率 $\overline{K}_1$、$\overline{K}_2$ 和 $\overline{K}_3$。在计算出平均采收率后，还要分别计算各列三个平均采收率的极差 R。例如 A 因素的极差 22＝82-60。

其次，选出最优搭配，并进一步试验验证。从表 10.4 平均采收率的数据中可以看出，A 因素 3 个水平中以第 3 个水平的乎均采收率为最高，达 82%，

即先确定 A_3。接下来，要按照备因素间极差的大小顺序来安排 C 因素的最好水平，因为各因素的极差反映了该因素各水平安排得好坏的最大差别。显然，在安排各因素的最好水平时，按照极差大小的顺序来安排是合理的。C 因素应选第 2 个水平，即 C_2，因为取 C_2 时平均采收率高达 79%。最后确定 B 因素的最好水平。一般来说，选 B_2 平均采收率最高，是合理的，但如果考虑生产成本的话，选 B_1 也是可以考虑的，因为 B_1 虽然平均采收率低 5%，但少投入 13 公斤碱。这就需要进一步进行经济核算，如果少投入 13 公斤碱节约的费用比减少 5%的平均采收率造成的损失还大，就应该选 B_1 而不是 B_2 了。通过以上分析，我们找到了最优搭配 $A_3B_2C_2$ 和可供选择的搭配 $A_3B_1C_2$。由于这两个搭配还没有试验过，还需要对 $A_3B_2C_2$ 和 $A_3B_1C_2$ 补充进行试验，观察其采收率指标，并结合生产成本，最终选择最优搭配进行正式生产。

正交试验的指标数值还可以应用方差分析方法对各因素的影响是否显著进行检验，分析可以更加深入、广泛。

3.正交试验的直观解释。我们以 3 个因素，每个因素 2 个水平的问题为例，若要进行全面试验，需要 2^3＝8 次。但如果用正交表，则只需 $L_4(2^3)$，即 4 次即可。$L_4(2^3)$正交表如表 10.5。

表 10.5　　　$L_4(2^3)$正交表

试验号	因素		
	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

该例的 3 个因素有如一个正方体的三向座标，每一因素的 2 个水平就是每个方向上直线线段的两个端点，该立方体共 8 个角，代表 8 次全面试验。正交试验用其中的 4 个角(带黑点的)代替全面试验的 8 个角。我们发现，正方体的六个面每个面上都有 2 个角，12 条边上每条边上都有 1 个点。虽然只是在 8 个角中选出了 4 个角，但对 AB、AC 和 BC 任意两个因素来说都是全面试验，因而这 4 个点具有很强的代表性。如果所要寻找的最优搭配不在正交试验的 4 个点中(例如最优点是 $A_1B_1C_2$)，也会通过与该点相邻的 3 个较优搭配点($A_1B_1C_1$，$A_2B_1C_2$，$A_1B_2C_2$)表现出来，而这 3 个较优搭配点都是正交试验中的点，那么通过这 3 个点很容易就摸索到最优搭配点了。正交表就是通过非全面试验的但具有代表性的部分试验点找出最优搭配的理想工具。

图 10.1　正交表的直观解释图

4.试验方案的灵活安排。要搞好试验设计，安排好试验方案是重要的基础工作，主要是解决好挑因素、选水平、选正交表这三个问题。

正交表是安排多因素试验的有力工具，在试验中有时多增加一两个因素并不增加试验次数。因此，除了我们已经确知某些因素对试验结果影响很小

可以不安排外，要尽可能地在试验中多考察一些影响较大或对试验结果的影响尚不清楚的因素。同时，也要考虑节约试验费用，反对盲目拼凑因素。

在选水平时，既要考虑水平个数的确定，也要考虑各水平用量的确定。在考虑水平的个数时又与选取正交表相联系，因为正交表确定之后，每个因素的水平数也就确定了。对于“品种”、“催化剂种类”等离散水平的因素来说，水平的个数是自然形成的，而对于“温度”、“时间”、“加料量”等连续水平的因素，水平个数是人为确定的。在选取水平和用量时，要先估计因素用量的取值范围，不要漏掉合理用量。接下来用等分法来选取有代表性的水平，如某“温度”因素是在 50—100℃之间变化，若考虑 3 个水平，选取 60℃、75℃、90℃较合适，若考虑 5 个水平，选取 55℃、65℃、75℃、85℃、95℃较为合适。

又由于并非任意因素和水平个数的正交表都存在，所以在确定水平个数时还必须考虑是否已有现成的正交表。当应用某一正交表安排某一试验方案后，某个因素仍空有水平的位置，可在该空闲位置上再重复安排该因素的某个水平，这种方法称为拟水平法。当找不到能容纳下全部因素和水平的较小正交表，但减少某因素的一个水平后就有合适的较小正交表来安排该试验时，可以将减少的水平作为后备水平，视首批试验结果的情况再决定第二批试验中是否用后备水平去代替首批试验的某一水平。当找不到能容纳下全部因素和水平的较小正交表时，也可以考虑用几张较小的正交表代替安排下全部因素和水平的较大正交表，这种方法称为分离水平法。

总之，灵活地应用已有的正交表安排好试验方案对于试验设计是十分重要的，读者可参阅参考文献并按照正交设计的原理在实际工作中灵活应用并加以创造。

§10.2.3　多指标综合评分

前面所讨论的试验设计只有一个反映试验结果的指标。然而在实际工作中反映试验结果的指标常常不只一个，这就需要将多个指标综合化为一个指标或者同时对多指标进行分析以进行综合评价，其中常用的方法即多指标综合评分的方法。

1.定性指标定量化。实际工作中的试验指标有时是以产量、浓度、采收率等定量形式出现，有时则以颜色、手感、口味等定性形式反映出来，对于定性指标要应用正交试验方法进行分析，首先需要将定性指标定量化，然后对定量化的指标进行分析。对于定性指标，往往难于直接给予准确的定量，简单易行的方法是先对定性指标结果排序，或直接用排序的序号作为定量指标，或根据排序的结果进行综合评分。

例 10.2　某服装生产企业要研究国际服装流行色，重点是搞清楚该种流行色中红、黄、蓝、白、黑 5 种基本颜色的比例。

解：已知 5 因素试验的最小正交表是 $L_8(4^1\times2^4)$，该正交表有 1 个因素是 4 个水平，其他 4 个因素都是 2 个水平。经分析确定，该国际流行色的主色是红色，将红颜色作为重点因素，考虑它的 4 个水平。按 $L_8(4^1\times2^4)$正交表安排 8 次试验及其结果见表 10.6。

在对 8 次试验结果进行分析时，根据定性描述，按满分为 100 分，最不满意为 0 分，给出了定量化综合评分，进一步的分析可通过对定量化的分数按前面所讨论的方法进行。如从 8 次试验结果简单地看，$A_2B_2C_2D_2E_2$ 是好的搭

配。经过分析，推出 $A_2B_2C_1D_1E_1$ 这一最好的搭配。

2.综合评分法。例 10.2 所讨论的是单个定性指标定量化的问题，该方法对多个定性指标定量化的问题也有同样的意义。对于多个定量指标的试验结果，既可以按前述的方法进行综合评分，也可对定量指标进行加权，综合成一个定量指标进行分析，这些方法简称为综合评分法。

例 10.3　在彩色电视机荧光粉的生产过程中会产生大量含有锌、镉等有毒物质的废水。试找出使用沉淀法进行一级处理的最好条件。

解：根据生产过程选用 $L_8(4^1\times2^4)$ 安排试验，(见表 10.7)，并得到含镉量和含锌量两个指标。这两个指标都是越小越好。

表10.6　　某服装国际流行色试验分析表

试验号	因素 红 A	素 黄 B	蓝 C	白 D	黑 E	试验结果（指标） 定性描述 定量化(分)
1	1（A_1）	1（B_1）	2（C_2）	2（D_2）	1（E_1）	不满意30
2	3（A_3）	2（B_2）	2	2（D_1）	1	不太满意50
3	2（A_2）	2	22	2（E_2）	90	非常满意
4	4（A_4）	1	2	1	2	还凑合60
5	1	2	1（C1）	1	2	不太满意55
6	3	1	1	2	2	很不满意10
7	2	1	1	1	1	还满意70
8	4	2	1	2	1	比较满意85
K_1	85	170	230	235	235	总和=450
K_2	160	280	220	215	215	
K_3	60					
K_4	145					
R	100	110	10	20	20	

表 10.7 中对于含镉量和含锌量这两个指标分别应用综合排序方法和加权平均方法将两个指标综合成为一个定量指标。如果两个指标的重要程度不同，也可以给更重要的一个指标以较大的权数。我们看到，这两个综合评价方法的结果是相同的。直观地看，搭配 $A_4B_2C_1D_2E_1$ 是好的，进一步分析可得出 $A_4B_2C_1D_1E_1$ 是所要寻找的最优搭配。

§10.2.4　三次设计概述

产品的三次试验设计方法是日本的田口玄一博士在 70 年代首先提出的。该方法在日本企业中应用后，极大地提高了日本工业产品的质量，使得日本产品在国际市场上具有很强的竞争力。进入 80 年代后，美国企业界开始庄意这种方法，很多大学和研究机构都在研究这种方法。

表10.7　　　　　　　　污水去镉去锌试验分析表

试验号	因素 A PH值	B 凝聚剂	C 沉淀剂	D $CaCl_2$	素 E 废水浓度
1	1（7—8）	1（加）	2（Na_2CO_3）	2（加）	1（稀）
2	3（9—10）	2（不加）	2	1（不加）	1
3	2（8—9）	2	2	2	2（浓）
4	4（10—11）	1	2	1	2
5	1	2	1（NaOH）	1	2
6	3	1	1	2	2
7	2	1	1	1	1
8	4	2	1	2	1

试验号	试验结果(指标) X_1 含镉(毫克/升)	X_2 含锌(毫克/升)	综合排序 名次　分数	加权平均 $\overline{X}=\frac{X1+X_2}{2}$	名次
1	0.72	1.36	80	1.04	8
2	0.52	0.90	540	0.71	5
3	0.80	0.96	515	0.88	7
4	0.60	1.00	630	0.80	6
5	0.53	0.42	465	0.48	4
6	0.21	0.42	290	0.32	2
7	0.30	0.52	375	0.40	3
8	0.13	0.40	1100	0.27	1

所谓三次设计，是指产品在设计过程中要经过的三个阶段，即系统设计(第一阶段或第一次设计)、参数设计(第二阶段或第二次设计)、容差设计(第三阶段或第三次设计)。在产品设计的第一阶段(即系统设计阶段)主要是专业技术人员应用专业知识进行系统结构方面的初步设计，使得该系统的结构具有所期待的特定功能。例如，在电视机等电器设备中通常都需要将交流电转化为直流电的整流器。如何将交流电转化为直流电，采用哪些元器件等都是这个设计阶段需要考虑的问题。系统设计以后，产品的结构和元器件就初步确定了。这时，要考虑的是如何确定各元器件的参数值(即元器件的数值)。选择元器件之间最好的参数组合,使得该产品具有最佳性能是第二阶段(即参数设计阶段)的主要任务。在产品设计的第三阶段(即容差设计阶段)，主要是考虑在产品性能变化不大的情况下，如何选择最低成本的元器件进行组装，即考虑成本核算和经济效益的阶段。在实际设计工作中，这三个设计阶段是紧密相连的。当专业技术人员拿出了系统设计方案后，统计人员要与技术人员一同参与后两个阶段的设计，因为参数设计和容差设计阶段主要是应用统计方法进行设计和分析。如果参数设计和容差设计的结果不很理想，通常要将设计方案返回到第一阶段，修改或重新考虑系统设计方案，直至三次设计全部完成，才可以进入生产阶段进行生产。

下面我们主要介绍田口玄一的参数设计和容差设计的思想，并配合案例加以说明。

1.参数设计。当一个产品进行了系统设计以后，该产品的系统结构就确

定下来了。假定该产品共有 n 个元器件，其对应的参数为 x_1，x_2，x_3，…，x_n。同时，我们假定 y 为该产品的某个特征值(即指标值)，那么，特征值 y 主要取决于元器件 x_1，x_2，…，x_n。参数设计的主要任务就是要选择这 n 个元器件的某个参数组合，使得产品的特征值达到设计的最佳要求。

如果特征值 y 可以通过这 n 个元器件的参数 x_1，x_2，…，x_n 用函数形式直接表达出来，或者可以通过它们直接计算，这时我们称之为可计算性项目的设计；反之，如果 y 不能由 x_1，x_2，…，x_n 用函数形式表达，而只能通过试验或测试等手段得到其观察值，我们则称这类问题为试验项目的设计。关于试验项目的设计，我们在前面已加以介绍，下面我们主要介绍可计算性项目的设计。

例如某彩色电视机的电源整流器要求输入 220 伏交流电而输出 115 伏直流电。在整流器中某整流晶体管 A 的功效曲线如图 10.2。

图 10.2　某晶体管的功效曲线图

由图 10.2 可以看出，若要选择输出电压刚好为 115 伏的晶体管，A_2＝20 恰能满足输出电压的要求。但若设计该整流器寿命为 10 年，则晶体管会逐渐老化。在老化过程中会产生士 30%的波动范围，即晶体管 A_2=20 的横向波动为 20×(1 士 30%)，结果是在 14—26 的范围内波动。由功效曲线可看出输出电压在 98—121 伏的范围间波动，输出电压的波动高达 23 伏，对整部彩色电视机接收质量影响很大。如果晶体管选择 A_4＝40，横向波动 40 ×(1±30%)，结果是在 28—52 的范围内波动，输出电压波动范围是 122—127 伏，仅有 5 伏的波动范围。显然，如果使用 A_4＝40 的晶体管并改变其他元器件(电阻)的参数，使得该晶体管与电阻配合能达到 115 伏电压输出，会使整流器稳定得多。当然，读者会提出这样的建议，即提高晶体管的质量，使其输出电压更稳定、波动更小。这当然是一条改进的措施，但生产高精度的晶体管会伴随着晶体管成本的增加。所以一般不如一个普通晶体管配合一个电阻更经济，同时也保证了输出电压的稳定。

田口玄一的参数设计中包含着这样的统计思想，即考虑任一元器件时，首先是从总体(整个产品或某一部分整体)的最佳性能(总体的最小方差)出发，然后考虑部分(元器件)的最优。若用统计学的术语来说，即首先考虑总体的最小方差(最稳定)，然后再考虑每一部分的无偏。同时还要考虑元器件的价格问题，要尽量选择耐老化、价格低的元器件，使得产品既质量可靠，又价格低廉，具有竞争力。

2.容差设计。经过参数设计阶段，找到一组元器件的最佳参数组合之后，接下来要考虑元器件成本最低和经济损失最小的问题。

在田口的容差设计中，考虑了多种损失函数。我们只介绍一种平均损失函数，但可由此了解容差设计的方法和思想。定义如下形式为平均损失函数

$$L = A \times \frac{S^2}{\Delta^2} \tag{10.1}$$

式中：L 表示平均损失函数；A 表示当某件产品不合规格时的损失值；S^2 表示产品质量特征的实际方差；Δ表示产品质量特征波动的容许误差。一般来说，容许误差Δ和不合规格时的损失值 A 都是固定的。也就是说，产品质量特征

值越稳定，实际方差就越小，产品平均损失也就越小。下面我们用一个平板玻璃生产的例子来看看如何应用统计方法分析产品质量问题并提高产品质量。

例 10.4　某玻璃厂生产一种平板玻璃，已知不合规格时每块的损失是 150 元。玻璃厚度的容许误差Δ=3㎜。现随机抽取 20 块玻璃检验玻璃厚度的质量，发现这 20 块玻璃的实际值与规格标准相差的误差(单位：㎜)分别为

0.3　0.6　-0.5　-0.2　0.0　1.0　1.2　0.8　-0.6　0.9

0.0　0.2　0.8　1.1　-0.5　-0.2　0.0　0.3　0.8　1.3

经计算，这 20 块玻璃与目标值的方差为

$$S^2=\frac{\sum_{i=1}^{20}(x_i-\mu)^2}{20}$$

$$=\frac{(0.3)^2+(0.6)^2+\cdots+(1.3)^2}{20}$$

$$=0.4795(mm)^2$$

将计算的方差和已知数据代入(10.1)式，有

$$L=150\times\frac{0.4795}{3^2}=7.99(元)$$

即平均每块玻璃损失 7.99 元。在计算出平均损失函数值之后，容差设计阶段要求进一步分析损失函数值中有多少是随机因素产生的，又有多少是系统因素产生的。要分解这一实际方差，需要对实际离差平方和进行方差分析。

我们已计算出总的离差平方和

$$SS_T=\sum_{i=1}^{20}(x_i-\mu)^2$$

$$=(0.3)^2+(0.6)^2+\cdots+(1.3)^2=9.59$$

$$SS_A=\sum_{i=1}^{20}(\bar{x}-\mu)^2=20(\bar{x}-\mu)^2=20(\frac{\sum_{i=1}^{20}x_i}{20}-\frac{20\cdot\mu}{20})^2$$

组间离差平方和

$$=20((\frac{\sum_{i=1}^{20}x_i-20\cdot\mu}{20})^2$$

$$=20(\frac{(x_1-\mu)+(x_2-\mu)+\cdots+(x_{20}-\mu)^{20}}{20})$$

$$=20(\frac{0.3+0.6+\cdots+1.3}{20})^2=2.66$$

则组内离差平方和

$$SS_E=SS_T-SS_A=9.59-2.66=6.93$$

将这些结果进行分析，则有方差分析表 10.8。

表 10.8　　　　　　　　方差分析表

方差来源	离差平方和	自由度	方差	F 值
系统误差 SS_A	2.66	1	2.66	7.29
随机误差 SS_E	6.93	19	0.365	
总　和 SS_T	9.59	20	0.4795	

若取显著性水平 a＝0.05，查表 $F_{0.05}(1,19)=4.38$。由 $F=7.29>F_{0.05}(1,19)=4.38$ 可知，系统误差是显著的。这里的系统误差，即平板玻璃生产时实际玻璃厚度的均值 $\bar{x}$ 与目标值 μ 之间的误差。这个误差是显著的，表明生产线还未调整到最佳状态，即使实际生产的均值 $\bar{x}$ 与目标值 μ 完全一致，也还有改进的余地。显然，随机误差所造成的平均损失为

$$L = 150 \times \frac{0.365}{3^2} = 6.08(\text{元})$$

那么，在总的平均损失中由系统因素造成的平均损失就是 7.99-6.08=1.91(元)

如将平板玻璃生产线调整成为无偏的，即实际均值与目标值一致后，平均每块玻璃可减少损失 1.91 元。

由此例可以看出田口方法的容差设计就是要对方差进行分析，计算各种损失函数，从而避免系统因素造成的损失，以获得最佳经济效益。田口玄一的质量管理方法不仅包括三次设计，还包括进入生产阶段后线内质量管理方法(三次设计统称为线外质量管理)，这部分内容可参阅参考献。

§10.3 产品质量的抽样控制

对产品质量进行控制，主要是通过统计控制图及时发现故障及其隐息，从而对生产过程中的产品质量进行监控。

§10.3.1 统计控制图的原理

现代工业产品的质量特征都有一定的规格和标准要求。例如灯泡、电池要有一定的使用寿命；钢丝绳、化学纤维要有一定的拉力强度；袋装食品如奶粉、糖果等要达到规定的重量；零部件的尺寸都有一定的技术要求，如此等等。但在生产过程中，产品之间的差异总是客观存在的。即使同一个工人，使用同一台机器，应用同样的原材料加工同样的产品，生产出的产品也不可能完全一样。因此，在制定产品质量标准时，一般都规定一个质量标准，同时给出允许的误差范围。例如奶粉自动装袋机要求每袋装足 500 克，但允许有±2 克的误差，通常将允许误差称为公差。

1.影响产品质量的因素。影响产品质量的因素很多，大致可归结为 5M1E 共 6 个方面，即人员(Man)、机器设备(Machine)、原材料(Material)、生产加工的方法(Method)、测量工具(Measurement)和环境(Enviroument)。从统计的角度，我们可将影响产品质量的众多因素分为系统性原因和随机性原因两大类。

系统性原因是某种特定的原因，往往造成产品与质量标准发生较大的整体偏差(如质量加工实际均值$\bar{x}$与标准值μ间的误差)。例如加工机器未按质量标准调试好、原材料质量不合格、加工方法不对等等。由系统性原因产生的误差，一经查明都是可以纠正的。如是机器的问题要重新调试；如是原材料问题可以更换；如是加工方法问题应予改进，等等。

随机性原因是由众多无法排除的偶然因素产生的，是客观存在并且不可避免的。例如机器的微小振动、原材料质量的细微差别、生产环境中温度的变化等等都会对产品质量产生影响，会给产品带来质量差异。这类随机性原因产生的质量差异虽然不可避免，但由于这类原因众多，差异有正有负，在互相抵消后会围绕着某一平均水平上下波动。

2.控制图的统计原理。在介绍了影响产品质量的两类原因之后，由其两类不同原因的性质可知，我们进行产品质量控制的主要目的是控制系统性原因的出现。这是因为系统性原因通常是影响产品质量的主要矛盾，同时系统性原因也是应该避免并且可以避免的。而随机性原因既难以避免，一般说来其波动对产品质量影响也不大。那么，如何区别产品质量特征的波动是系统性原因还是随机性原因呢？产品质量控制图即是通过区别系统性原因还是随机性原因来达到控制不合格品出现的目的的。

假定某一产品质量标准为μ，生产过程按这一标准进行，由于存在着并且只存在不可避免的随机性原因的影响，实际产品质量值$\bar{x}$会以μ为中心，围绕着μ形成一个两头小中间大的误差的正态分布。用数学符号表示，即

$$x_i=\mu+e_i,\ e_i\sim N(0,\sigma^2)$$

则

$$x_i\sim(\mu,\sigma^2)$$

由正态分布的原理，我们知道 $x_i\le\mu\pm3\sigma$的概率为 99.73%，即观察质量值 X_i 落在标准值μ正负 3 个标准差之内的概率为 99.73%。换句话说，如果产品的生产过程是按标准值进行的并且只存在随机性原因的影响，产品实际质量

值超过标准值 3 个标准差范围之外的概率还不到 1%。这个不到 1%的概率在统计学上称为小概率事件。对于这类不到 1%的小概率事件，在单独的一、二次试验中，由于其发生的概率太小，小到可以忽视其出现的地步，我们通常认为它不会发生。例如乘飞机、火车、或汽车旅行，发生意外事件都是可能的。但这种可能性太小，因而在我们某次旅行时，认为这个出现意外的小概率事件不会发生，一般不会因为害怕小概率的意外事件发生而不去出差或旅行。在产品质量控制中，我们就是应用正态分布的 3σ原理(或小概率原理)来进行控制的，即当实际产品质量特征值超出以μ为中心的正负 3 个标准差的控制范围，我们就认为产品质量出现了问题，即出现了系统性的质量原因。由前边假设检验的原理可知，在我们做出这种判断时，可能将本来是正常的生产过程误判为异常情况出现，然而犯这种错误(α错误)的概率是很小的，还不到 1%。同理，我们也是认为这种错误在个别的一、二次控制检验中是不会出现的。

3.工程能力分析。在对生产过程进行产品质量控制之前，需要了解工程能力是否满足产品质量的要求，就需要进行工程能力分析(Process Capability Analysis)。工程能力分析通常要计算工程能力指数 C_p

$$C_P = \frac{\text{公差下限} - \text{公差上限}}{6\sigma} = \frac{T}{6\sigma}$$

式中：C_p 表示工程能力指数；T 表示公差范围，是产品质量标准允许的最大误差范围，例如，奶粉自动装袋机的重量标准是 500 克±2 克，公差上限是 502 克，公差下限是 498 克，公差范围是 4 克，6б是实际产品质量特征的 6 个标准差，即正态分布控制图的最大范围。我们不难看出，工程能力指数反映的是质量要求的范围与实际产品质量变动范围之间的比较数值。换句话说，C_p 是衡量实际生产过程能否满足公差要求的一个相对数。显然，如果 C_p，值越大，表明实际误差范围小于公差要求范围，产品加工过程的质量是好的；反之，如果 C_p 值越小，表明实际误差范围大于公差要求范围，产品加工过程的能力已经不能满足产品质量的要求，工程质量有问题。在实际应用时要注意，б的数值一定要在生产过程稳定的前提下取得，否则计算出的 C_p 不可靠。在计算出 C_p 后，不同的 C_p 反映不同的工程能力，详见表 10.9。

表10.9 工程能力指数分析表

C_p值	加工等级	能力分析
$C_p > 1.67$	特级加工	加工精度很高，要考虑加工成本是否浪费
$1.67 \geq C_p > 1.33$	A级加工	加工精度足够，对产品质量检验可大大简化
$1.33 \geq C_p > 1.00$	B级加工	加工精度恰好，对产品质量进行抽样检验
$1.00 \geq C_p > 0.67$	C级加工	加工精度不够，对产品质量进行全面检验
$0.67 \geq C_p$	D级加工	必须改进工程能力

需要注意的是，前面所介绍的工程能力指数是在公差中心值与实际产品质量特征均值相同的条件下应用的。如果公差中心值与产品均值不同，要设法调整加工过程，使其均值与公差中心(即规格标准要求)相同。因为工程能力指数是在产品质量均值无偏的情况下用以反映实际方差大小的一个比较指

标。同时，如果实际产品质量均值与公差中心值有差异，这个差一定是系统性的质量影响因素，应加以去除。我们在下面所介绍的产品质量控制方法，主要是针对实际加工过程中绝大多数的 B 级加工过程而应用的。其他的加工等级应考虑其他改进或调整措施来改进产品质量。

§10.3.2　控制图的作法

对于产品质量的控制，依其产品质量特征的不同特点，可分为计量控制、计件控制和计点控制。这三种产品质量特征分别服从正态分布、二项分布和普哇松分布，但当样本容量充分大及其他条件满足时，二项分布与普哇松分布都以正态分布为其极限分布，因而正态分布的 3σ 原理对这些控制图都是适用的。

1.计量控制图。计量控制图应用于质量特征可以直接用物理的度量单位来加以计量的产品质量控制过程，如机器零件的尺寸；食品装袋机装袋的重量；砖与水泥等建筑材料的抗压强度；电子元器件的电流、电压和耐用时间等。这些变量都是连续型的变量，在生产过程正常的条件下，产品的质量特征服从正态分布。

计量控制图的种类很多，主要有均值与极差($\overline{x}$—R)控制图、中位数与极差($\overline{x}$—R)控制图、均值与标准差($\overline{x}$—R)控制图，两极(G—H)控制图，单个值控制图等等，其中应用最广的是均值与极差控制图。下面，我们就通过这对控制图的一个实例介绍计量控制图的作法。

例 10.5　某厂电解工序每日分三班工作，现取了 30 天的数据，要应用这些数据做出 $\overline{x}$—R 控制图。数据见表 10.10。

表10.10　　　　$\overline{x}$—R控制图的数据与计算表

工序名称：电解工段　　　　采样时间：1988年6月1日——6月30日

指标：电流效率　　　　采样人：×××

单位：%　　　　负责人：×××

月、日	样组编号	早班	中班	晚班	$\overline{x}$	R	备注
6.1	1	92.2	91.7	96.8	93.57	5.1	
2	2	96.5	98.4	96.8	97.23	1.9	
3	3	99.5	92.4	99.2	97.00	7.2	
4	4	93.5	98.7	90.2	97.13	8.5	
5	5	92.2	95.7	95.1	94.33	3.5	
6	6	96.1	90.9	94.3	93.77	5.2	
7	7	99.5	90.6	95.8	95.30	8.9	

控制图的作法如下：

(1)收集数据。绘制控制图的数据要尽可能的多，最好能在 100 个以上。同时注意在生产过程稳定正常的条件下搜集数据，否则利用该数据绘制的控制图难以达到区别系统性原因和随机性原因，以及时发现系统性原因影响的目的。

(2)妥善分组。一般情况下可按同一天、同一个班次、同一台机器设备进行数据抽取和分组。每组样本容量 3—5 个为宜。本例中以天分组，每组中 n＝3。

(3)计算各组的均值 $\overline{X}$。如第一组 6 月 1 日

$$\bar{x} = \frac{92.2 + 91.7 + 96.8}{3} = 93.57$$

(4)计算总平均值 $\bar{\bar{X}}$。

$$\bar{\bar{x}} = \frac{93.57 + 97.23 + \cdots + 95.70 + 94.93}{30}$$

$$= \frac{2842}{30} = 94.70$$

(5)计算各组的极差 R。如第一组 6 月 1 日

R=96.8-91.7=5.1

(6)计算平均极差。

$$\bar{R} = \frac{5.1 + 1.9 + \cdots + 7.6 + 4.5}{30}$$

$$= \frac{155.6}{30} = 5.187$$

(7)计算控制图的控制限。

$\bar{x}$ **图的中心线**$(CL) = \bar{\bar{x}} = 94.70$

上控制限$(UCL) = \bar{\bar{x}} + 3\sigma_{\bar{x}}$

$= \bar{\bar{x}} + A_2\bar{R}$

$= 94.70 + 1.023 \times 5.187$

$= 100.046$

下控制限$(LCL) = \bar{\bar{x}} - 3\sigma_{\bar{x}}$

$= \bar{\bar{x}} - A_2\bar{R}$

$= 94.70 - 1.023 \times 5.187$

$= 89.434$

R**图的中心线**$(CL) = \bar{R} = 5.187$

上控制限$(UCL) = \bar{R} + 3\sigma_R$

$= D_4\bar{R}$

$= 2.575 \times 5.187$

$= 13.357$

下控制限$(LCL) = \bar{R} - 3\sigma_R$

$= D_3\bar{R}$

= **不予考虑**

其中 A_2、D_3 和 D_4 是样本容量 n 的函数，其系数值见表 10.11。

表10.11　　　　　　　　　　上下控制限系数表

n	A_2	D_3	D_4
2	1.880	—	3.267
3	1.023	—	2.275
4	0.729	—	2.282
5	0.577	—	2.115
6	0.483	—	2.004
7	0.419	0.076	1.924
8	0.373	0.136	1.864

其中 D_3 系数值当 $n\leq 6$ 时接近于 0，故下控制限可不予考虑。

(8)画出 $\bar{x}$ 和 R 控制图，并将 $\bar{x}$ 和 R 点描在图内。在画控制图时，中心线画实线，上下控制限画虚线，画出 $\bar{x}$ 和 R 图见图 10.3。

在描点绘图时，要注意：第一，$\bar{X}$ 图的点用点表示，R 图的点用叉表示；第二，越出上下控制限的点用红色的“◎”画出，引起警惕。因为此时要停止加工，检查生产过程是否异常及异常的原因；第三，用折线将图中点子联起来以观察点子的流动特点。

图 10.3　$\bar{X}$—R 控制图

(9)根据正态分布的特点判断制图期间的生产过程是否正常和稳定。如不正常，不稳定，则要采取措施，改进不正常状态，然后再重抽样绘图。如果已进入正常和稳定状态，则可延长 $\bar{X}$ 和 R 控制图的控制限，对生产过程进行控制和管理。对于比例，我们由图 10.3 观察，生产过程是正常和稳定的。这样，就可以延长控制图对 7 月份的生产进行控制了。

以上我们重点介绍了 $\bar{X}$—R 图的作法，其他计量控制图与 $\bar{X}$—R 大同小异，其制图过程与控制原理都是相同的，故不再一一介绍了。

2.计件控制图。对于某些产品的质量特征难以用数理来表示或者虽然可用数量表示，但由于数量特征太多，对每一个数量特征都进行控制费用太高，这种情况下均采用计件控制方法。计件控制时，人们常用合格与不合格表示产品质量，故也称为属性控制图。通常情况下，人们是从一批产品中抽取 n 件进行检验，观察其中的不合格品率并对其不合格品率进行控制。这种控制图称为不合格品率控制图，或简称 P 图。

例 10.6　某车间生产某种型号的零件，每天从共产品中抽出一个样本，容量在 740 伴左右。现连续抽取 25 天的样本，各天的次品件数和次品率如表 10.12。

不合格品率(P)控制图的作法如下：

(1)搜集数据。在稳定的生产过程中取 20—25 组数据。每组数据样本容量 n 的确定要视产品次品率 P 的大小而定。一般而言，当 P 很小时 n 要取得大一些；P 较大时 n 可取得稍小些。但 n 一般不得小于 50，nP 在 5 左右较适宜。本例中 nP 都在 50 左右，说明样本容量 n 取得较大。如从检验费用节约的角度，可以取得稍小些，平均取 70 个左右即可。

(2)计算各组的 P。如第一组

$$p=\frac{48}{724}=0.066$$

(3)计算平均次品率 P 和平均样本容量 n

$$\overline{p}=\frac{1318}{18533}=0.071$$

$$\overline{n}=\frac{18533}{25}=741$$

(4)计算控制限

中心线(CL)=$\overline{p}$=0.071

上控制限$(UCL)=\overline{p}+3\cdot\sqrt{\overline{p}(1-\overline{p})/\overline{n}}$

$$=0.071+3\cdot(0.0094)$$

$$=0.099$$

下控制限$(LCL)=\overline{p}-3\cdot\sqrt{\overline{p}(1-\overline{p})/\overline{n}}$

$$=0.071-3\cdot(0.0094)$$

$$=0.043$$

(5)画出控制图并描上点子。画法与图 10.3 相同，因而略去。

需要说明的是，在 P 图中由于每组样本容量 n_i 不完全相同，严格来说，在计算上下控制限时要用公式

$$UCL_i=\overline{p}+3\cdot\sqrt{\overline{p}(1-\overline{p})/n_i}$$

$$LCL_i=\overline{P}-3\cdot\sqrt{\overline{P}(1-\overline{P})/n_i}$$

这样，严格的控制限要随各组样本容量的变化而变化。通常，在各组样本容量相差不太大时，可用其均值$\overline{n}$代入作为近似方法。

3.计点控制图。有些产品的质量特征是用单位面积上有多少缺陷数来表示的，如一匹棉布上有多少疵点，一个铸件上有多少气孔、砂眼等等。计点控制图就是用来控制产品的缺陷数是否超过规定的界限的方法。在生产情况正常时，单位面积上的缺陷数服从普哇松分布，而普哇松分布的均值和方差都是参数λ，标准差为$\sqrt{\lambda}$。普哇松分布也可用正态分布来近似计算。

例 10.7　从某车间生产的铸件中随机地抽取 10 件进行检验。这 10 件的气孔个数分别为

3,2,0,5,4,6,0,7,7,6

试根据这些数据作出缺陷数控制图(C 图)。

(1)计算出中心线(CL)=$\overline{c}=\frac{\Sigma c}{k}=\frac{40}{10}=4$

(2)计算上下控制限

$$上控制限(UCL) = \bar{c} + 3\sqrt{\bar{c}}$$
$$= 4 + 3\sqrt{4}$$
$$= 10$$
$$下控制限(LCL) = \bar{c} - 3\sqrt{\bar{c}}$$
$$= 4 - 3\sqrt{4}$$
$$= 0$$

(3)画出控制图并描点。由于其画图方法与图 10.3 相同，故也略去。

在应用控制图时需注意如下问题：

(1)生产过程发生了变化，如换了新原料或新工艺流程后，需要重新作图。

(2)对于加工同一产品的不同机器应分别绘制控制图。

(3) $\bar{x}$—R 图的 $\bar{x}$ 图与 R 图要结合起来进行判断并积累经验，灵活应用。例如在机械行业中，某次检查发现 $\bar{x}$ 越出控制下限，但 R 图正常，经检查是车床机油太少，齿轮旋转不灵所致。以后再发现此种情况时，就不必将机器全部拆开检修，而要着重检查机油和齿轮的情况。这样可以节约人力与物力。

(4)控制图的作用是及时发现系统性原因，当问题刚刚出现时就加以纠正，可以减少不合格品，提高生产过程中的产品质量。

§10.3.3　控制图的诊断与分析

1.生产过程异常的诊断。如何应用控制图及时发现生产过程出现了异常呢？下面给出一些主要的诊断标准：

(1)点子越出了控制限；

(2)点子虽然没有越出控制限，但点子的排列不是随机的，如下情况出现也应诊断为异常，见表 10.13。

以上这些生产异常的种种情况，判断的标准都是出现的概率很小，小到不可能出现。因而根据小概率原理诊断为生产过程中产品质量特征的分布已不再是原来正常状态下的正态分布，故诊断为生产异常。例如点子在控制限附近出现中“连续 3 点中有 2 点出现在控制界限附近”的情况就是一个小概率事件。这里所说的在控制界限附近是指点子落在正态分布(控制限内)第二个标准差和第三个标准差(控制限上)之间的区域。根据正态分布的概率得知，点子落在这个区域的概率为 0.9973-0.9545=0.0428。则任意 3 点中 2 个点子落在该区域的概率为

P(任意 3 点中 2 点落在 2σ 至 3σ 之间)= $C_3^2(0.0428)^2(0.9545)^1$

$$= 0.0052 = 0.52\%$$

其他诊断为异常的各种情况也都是类似的小概率事件。

表 10.13　生产异常情况诊断表

2.生产过程异常的原因分析。当通过控制图诊断出生产过程出现了异常后，需要对质量问题的原因进行分析，其中行之有效的方法是鱼刺图和排列图。

(1)鱼刺图。鱼刺图又称为因果分析图或石川图，是日本石川馨教授发明的，因而称为石川图。这种图由于其形状很像鱼骨头，一般习惯称为鱼刺图。在分析质量原因时，将出现的质量问题放在鱼头的位置，然后将主要原因列

出，再依次分析影响主要原因的次要原因，影响次要原因的再次要原因等。下面我们对切削刀出现裂缝的质量问题应用鱼刺图进行分析。经分析发现，切削刀质量问题主要是机器设备和人员两方面的问题，则着重从这两方面查找原因。分析过程见图 10.4。

图 10.4　鱼刺分析图

在利用鱼刺图对众多原因进行分析后，应从其中挑出主要原因来重点加以解决，这样就解决了主要矛盾。在从众多原因中抓主要原因时，可以应用排列图。

(2)排列图。排列图是美国质量管理专家朱兰(Juran)应用了意大利经济学家巴雷特(Pareto)在研究社会财富的分配时发现的极少数人占有多数社会财富，而多数人仅占有少量社会财富的原理，提出质量问题中极其重要的少数与无关紧要的多数的研究方法。例如某厂切削刀质量问题表现为表 10.14 的不合格项目和不合格品件数。

若将这些不合格品项目的件数或所占百分比画成直方图，即不合格品项目的排列图，见图 10.5。

由表 10.14 与图 10.5 我们不难看到在众多的不合格品项目

表10.14　　某厂切削刀质量排列表

不合格项目	不合格品件数	占百分比（%）
1. 料　短	42	47
2. 裂　缝	28	32
3. 刃　口	7	8
4. 硬　度	4	4
5. 光洁度	2	2
6. 其　他	6	7
合　计	89	100

图 10.5　排列图

中，由不合格品件数看，料短是最主要的质量问题，裂缝问题其次，这两类问题导致的不合格品件数占全部不合格品的 79%。因而，若解决了这两类质量问题就解决了问题的大部分。以上仅是简单地从不合格品件数来进行分析的。若将不合格品导致的经济损失考虑进去，则裂缝又成为第一主要质量问题，见表 10.15。

现对裂缝这一主要原因作进一步研究，发现主要由 A、B 两台机器所致。在这两台机器所生产出的 28 件次品中，机器 A 占 24 件，机器 B 占 4 件(见图 10.6A)，则我们可以肯定机器 A 是导致产品质量问题的主要部分。我们进而对机器 A 所生产的 24 件次品进行分析(见图 10.6B(a))，这 24 件次品是由操纵这台机器的三班的甲、乙、丙三人平均造成的，则可以肯定是机器 A 的原因；若如图 10.6B(b)所示，在操纵这台机器的三班工人中，甲工人生产出 21 件次品，乙工人生产出 2 件次品，丙工人生产出 1 件次品，则可以断定问题源于甲工人的操纵。

图10.15　　　　　　　　　某厂切削刀质量损失分析表

不合格项目	损失金额（元）	累计损失金额（元）	累计百分比（%）
1.裂缝	2×28=56	56	48
2.料短	1×42=42	98	84
3.硬度	2×4=8	106	91
4.刃口	1×7=7	113	97
5.光洁度	0.8×2=1.6	114.6	98
6.其他	0.4×6=2.4	117	100
合　计	117	—	—

图 10.6　某厂切削刀裂缝原因分析排列图

按照这种层层抓主要问题的方法就可以很快解决主要矛盾。如果应用各种方法均找不出主要问题，可考虑应用试验设计来进行分析。

§10.4　产品质量的抽样检验

产品质量的抽样检验是在某道生产过程之后或者全部生产过程结束之后所进行的最后一道产品质量管理工作。这种检验工作通常是生产者与接收者(接收者既可以是下道工序也可以是商业部门)共同进行，从成批产品中抽取样本以决定是否接收整批产品。我们将在这一节中介绍一次抽样检验、二次抽样检验和序贯抽样检验方法。

§10.4.1　一次抽样检验

1.一次抽样检验方案。一次抽样检验方案是指通过抽取一个样本就做出接收或拒收决定的抽样检验方案。

对于一次抽样方案，要在检验前确定样本容量 n(也称检验量)和样本中容许的不合格品件数 c(也称判定数)这两个数值。如果样本中实际不合格品件数 d≤c，则认为整批产品合格，加以接收；如 d＞c，则认为整批产品不合格，加以拒收。一次抽样方案的检验过程如图 10.7。

图 10.7　一次抽样方案流程图

2.接收概率的计算。从直观看来，对于同样质量的一批产品，在批量大小不同时，似乎我们只要根据批量大小按比例地抽取样本容量，就可以得到同样的抽样检验效果。例如，方案(100；5，0)和方案(500；25，0)表面上看起来似乎是一样的，实际是两种完全不同的方案。区别主要在于两个方案有完全不同的接收概率。

设批量为 N 的一批产品，其不合格品率为 P，则该批量中不合格品的总件数 D=NP。现从 N 件中抽取 n 件检验，出现的不合格品件数 d≤c 的概率属于超几何分布，即

$$P(d \le c) = \sum_{d=0}^{c} \frac{C_D^d \cdot C_{N-D}^{n-d}}{C_N^n} \tag{10.3}$$

在前面例举的两个方案中，若 P=0.05，当批量为 100 时，D=100×0.05=5，d=0，则

$$P(d \le 0) = \frac{C_5^0 \cdot C_{100-5}^{5-0}}{C_{500}^{25}} = 0.77$$

当批量为 500 时，D＝500×0.05=25，d=0，则

$$P(d \le 0) = \frac{C_{25}^0 \cdot C_{500-25}^{25-0}}{C_{500}^{25}} = 0.29$$

由于超几何分布计算的工作量很大，所以当 N 很大，而 n 相对较小，即 $\frac{n}{N} \le 0.1$ 时，可以用二项分布作为超几何分布的近似。当 P 很小，nP 为一不大的常数时，可进一步用普哇松分布来近似。

例 10.8　设产品批量 N＝1000，不合格品率 P=0.04。试按方案(1000；30，1)计算这批产品的接收概率。

解：首先用超几何分布

$$P(d \le 1) = \sum_{d=0}^{1} \frac{C_{40}^{d} \cdot C_{1000-40}^{30-d}}{C_{1000}^{30}}$$

$$= \frac{C_{40}^{0} \cdot C_{960}^{30}}{C_{1000}^{30}} + \frac{C_{40}^{1} \cdot C_{960}^{29}}{C_{1000}^{30}}$$

$$= 0.6603$$

由于 $\frac{n}{N}$=0.03＜0.10，可用二项分布近似计算

$$P(d \le 1) = \sum_{d=0}^{1} C_{n}^{d} P^{d} (1-P)^{n-d}$$

$$= C_{30}^{0}(0.04)^{0}(0.96)^{30} + C_{30}^{1}(0.04)^{1}(0.96)^{29}$$

$$= 0.6612$$

又由于 nP=1.2，为一不大常数，也可用普哇松分布近似

$$P(d \le 1) = \sum_{d=0}^{1} \frac{(nP)^{d}}{d\,!} e^{-np}$$

$$= \frac{(1.2)^{0}}{0\,!} e^{-1.2} + \frac{(1.2)^{1}}{11} e^{-1.2}$$

$$= 0.6626$$

3.抽样特征曲线(OC 曲线)。当方案(N； n，c)确定以后，随着不合格品率 P 的变化，接收概率 L(P)也会随之发生变化。现仍以方案(1000；30，1)为例，我们计算得到表 10.16。

表 10.16　　　　　　　抽样特征曲线数值算表

P	nP	接收概率L（P）
0.0	0.6	0.879
0.0	1.2	0.661
0.06	1.8	0.455
0.08	2.4	0.296
0.10	3.0	0.184
0.12	3.6	0.110
0.14	4.2	0.064
0.16	4.8	0.036
0.18	5.4	0.020
0.20	6.0	0.011
M	M	M

由表 10.16 可描出图 10.8。

图 10.8　方案(1000,30,1)的抽样特征曲线

任一抽样检验方案，其目的在于正确判别合格批与不合格批。一个好的抽样方案，当产品质量较好时，应以较高概率接收；当产品质量变坏时，接收概率应迅速减小；而当产品质量坏到某个规定限度时，该方案应以很高概率拒收。因而抽样特征曲线是评定和比较不同抽样方案的有效工具。

4.检验中可能犯的两种错误α和β。由于方案抽取的是随机样本，并应用样本信息对整批产品进行推断，所以就可能犯错误。在这里所犯的两类错

误与假设检验中所犯的两类错误是一样的。将本来可以接收的合格批误判为不合格批，犯的是第一类错误，用 a 表示，一般常取 5%或 1%。在某检验方案下，达到或低于 α 的较小的不合格品率 P0 称为可接受质量水平(AcceptableQualityLeve1，简称 AQL)，即有

$$
\begin{aligned}
\alpha &= 1 - L(AQL) \\
&= 1 - \sum_{d=0}^{c} \frac{(nP_0)^d}{d\,!} e^{-nP_0} \qquad (10.4)
\end{aligned}
$$

这就是说，当产品的实际不合格品率 P 等于一个很小的 P_0 时，我们本应予以整批接收，然而却将其拒收了，显然犯的是“认真为假”的α错误。当我们将不合格批误判为合格批时，犯的是第二类错误，用β表示，一般也常取 5%和 1%。在某检验方案下，当产品的不合格品率 P 大到我们只愿意以很小的概率β来接收时，此时的 P=P1 被称为批容许的最大不合格品率(LotT01erancePercentDefective，简称 LTPD)，即

$$
\beta = L(LTPD) = \sum_{d=0}^{c} \frac{(nP_1)^d}{d\,!} e^{-np_1} \qquad (10.5)
$$

在实际检验中，由于存在着随机抽样误差，α和β的出现是难免的。由两类错误的性质可知，生产者在检验时更担心犯第一类错误，接收者在检验时较担心犯第二类错误。在实际检验时，双方共同商定，当 $P \le P_0$ 时犯第一类错误应小于 a 而当 P≥P1 时犯第二类错误应小于β。因此，抽样方案的 n 和 c 实际上是根据双方商定的 P_0、P_1、α和β这 4 个已知数值计算得到的，即解如下的联立方程组

$$
\begin{cases}
1 - \alpha = L(AQL) = \sum_{d=0}^{c} \frac{(nP_0)^d}{d\,!} e^{-np_1} \\
\beta = L(LTPD) = \sum_{d=0}^{c} \frac{(nP_1)^d}{d\,!} e^{-nP_1}
\end{cases}
$$

为了方便使用者，不必每次都通过已知的 P_0、P_1，α和β去解 n 和 c,统计学家已为我们编好现成的一次方案检查表，可直接通过查表求出方案的 n 和 c 值，有兴趣的读者可阅读参考文献上的有关参考书。

5.退回筛选验收制度。退回筛选是世界各国许多公司企业在对产品进行抽样检验时所采取的一种验收制度。当产品经抽样检验为合格品时，要将样本中的不合格品换成合格品后，才准许该批产品出厂：当产品检验拒收时，则要进行百分之百的全部检验，要将所有的不合格品全部换成合格品才能出厂。

对于退回筛选的验收制度，作为生产者关心的是如何能减少检验的数量，而作为接收者则更关心出厂之后的质量水平。下面我们先介绍生产者关心的平均检验件数 I 的计算。当某检验批的总体不合格品率 $\overline{P}$ 未知时，通常我们用样本平均不合格品率 P 来代替，于是有

$$
\begin{aligned}
I &= n \cdot \sum_{d=0}^{c} \frac{(n\overline{P})^d}{d!} e^{-n\overline{P}} + N(1 - \sum_{d=0}^{c} \frac{(n\overline{P})^d}{d!} e^{-n\overline{P}}) \\
&= N - (N-n)\sum_{d=0}^{c} \frac{(n\overline{P})^d}{d!} e^{-n\overline{P}} \\
&= n + (N-n) - (N-n)\sum_{d=0}^{c} \frac{(n\overline{P})^d}{d!} e^{-n\overline{P}} \\
&= n + (N-n)(1 - \sum_{d=0}^{c} \frac{(n\overline{P})^d}{d!} e^{-n\overline{P}})
\end{aligned} \tag{10.6}
$$

上式中的平均检验件数 I 实际上是一个加权算术平均数，即当整批产品接收时要检验 n 件，而当整批产品拒收时则要检验 N 件。所以平均检验件数 I 等于 n 乘以接收概串加上 N 乘以拒收概率，即

$$I=nL(\overline{P})+N(1-L(\overline{P}) \tag{10.7}$$

由上式可以看出，平均检验件数 I 是由两个因素决定的。一是若 P 较大，则 $L(\overline{P})$较小，$1-L(\overline{P})$则较大，I 也就较大；二是 N 大 I 也随之增大。因而生产者若要降低 I 的数值，其最有效的办法就是降低平均不合格品率 $\overline{P}$

生产者关心的是平均检验件数，因为这宜接关系到生产成本和经济效益。接收者则关心产品出厂后的平均质量水平(Average Outgoing Quality，简称 AOQ)，其计算公式为

$$
\begin{aligned}
AOQ &= \frac{N\overline{P} - I\overline{P}}{N} \\
&= \overline{P}\frac{N-I}{N}
\end{aligned}
$$

$$
\begin{aligned}
&= \overline{P}\frac{(\overline{N}-n)-(N-n)(1-\sum_{d=0}^{c} \frac{(n\overline{P})^d}{d!} e^{-n\overline{P}})}{N} \\
&= \overline{P}\frac{N-n}{N}\sum_{d=0}^{c} \frac{(n\overline{P})^d}{d!} e^{-n\overline{P}}
\end{aligned} \tag{10.7}
$$

平均质量水平 AOQ 实际计算的是平均出厂不合格品率，即平均未检验的不合格品件数占批量总数的百分比。当批量 N 很大，样本 n 很小时，式中的 $\frac{N-n}{N}$ 趋向于 1，则 $AOQ = \overline{P} \cdot L(\overline{P})$，即平均不合格品率与接收概率的乘积。显然，$\overline{P}$ 大则 $L(\overline{P})$小；$\overline{P}$ 小则 $L(\overline{P})$大。AOQ 有一个极大值称为平均出厂不合格品界限，接收者对这个界限很关心，因为它是接收者可能蒙受的最大损失。在实际检验工作中，生产者总希望平均检验件数越少越好，接收者也希望平均出厂不合格品率及界限越小越好，双方利益矛盾斗争的结果要求提高产品质量水平并根据双方的利益决定抽样检验方案。

§10.4.2　二次抽样检验

1.二次抽样检验方案。二次抽样检验方案是分两个阶段进行，必要时抽取两个样本并做出接收或拒收的抽样检验方案。对于二次抽样方案，需要确定 4 个数值，即第一个样本容量 n_1，第二个样本容量 n_2，第一合格判定数 c_1 和第二合格判定数 c_2。二次抽样方案流程图如图 10.9。

2.接收概率的计算。二次抽样方案的接收概率是由两部分组成的，第一部分 $P(d_1 \leq c_1)$是第一次抽样中就予以接收的概率。第二部分 $P(d_1 + d_2 \leq c_2)$是第二次抽样后才决定接收的概率，其中当 $d_1=c_1+1$ 时，尚不能决定整批产品是否接收，再抽检第二个样本。这时，若 d_2 取值在 0，1 ，⋯，c_2-c_1-1 时，两次抽检的不合格品件数之和仍小于或等于 c_2，仍可做出接收整批的判决。同理，当 $d_1=c_1+2$ 时，d_2 的取值在 0，1，⋯，c_2-c_1-2 时，整批仍可接收。若 N 很大，n_1、n_2 和 P 较小，同时 n_1P 和 n_2P 为不大的常数时，可用普哇松分布作为超几何分布的近似。我们可用普哇松分布写出二次抽样方案的接收概率的接收概率。

图 10.9　二次抽样方案流程图

$$
\begin{aligned}
L(P) &= P(d_1 \leq c_1) + P(d_1 + d_2 \leq c_2, c_1 < d_1 \leq c_2) \\
&= \sum_{d_1=0}^{c_1} \frac{(n_1P)d_1}{d_1!} e^{-n_1P} + \sum_{d_1=c_1+1}^{s_2} \frac{(n_1P)d_1}{d_1} e^{-n_1P} \\
&\quad \cdot \left[\sum_{d_2=0}^{3-d_1} \frac{(n_2P)d_2}{d_2!} e^{-n_2P}\right]
\end{aligned}
\tag{10.8}
$$

例 10.9　试就方案(1000；36，59，0，3)计算 P=0.01 的接收概率。

解：已知 N＝1000 较大，$n_1=36$，$n_2=59$ 较小，P=0.01 也较小。$n_1P=0.36$、$n_2P=0.59$ 均为不大的常数，可用普哇松分布近似。

$$
\begin{aligned}
L(P) &= \sum_{d_1=0}^{0} \frac{(0.36)^0}{0!} e^{-0.36} + \sum_{d_1=1}^{3} \frac{(0.36)^{d_1}}{d_1!} e^{-0.36} \\
&\quad \cdot \left[\sum_{d_2=0}^{3-d_1} \frac{(0.59)d_2}{d_2!} e^{-0.59}\right]
\end{aligned}
$$

$=0.6980+0.2893=0.9873$

3.平均检验件数与平均出厂不合格品率。在二次抽样方案中平均检验件数 I 可用下式计算

$$
\begin{aligned}
I &= n_1 \cdot P(d_1 \leq c_1) + (n_1 + n_2) \cdot P(c_1 < d_1 \leq c_2, d_1 + d_2 \leq c_2) \\
&\quad + N(1 - L(P))
\end{aligned}
\tag{10.9}
$$

我们看到这实际上也是一个加权算术平均数，用第一次就接收的 n_1 乘以其概率 $P(d_1 \leq c_1)$，加上二次抽样后接收的(n_1+n_2)乘以其概率 $P(c_1 < d_1 \leq c_2, d_1+d_2 \leq c_2)$，然后再加上 N 乘以拒收概率。这三个概率之和为 1。

现仍以例 10.9 中计算在方案(1000；36，59，0，3)时 P=0.01 条件的 I。我们可惜用刚刚计算出的接收概率值，并应用(10.9)式计算

$I=36(0.698)+(36+59)(0.2983)+1000(0.0127)$

$=65.31$

在计算出平均检验伴数 I 之后，平均出厂不合格品率仍用一次抽样方案中的公式，即

$$
AOQ = P\frac{N-1}{N} \tag{10.10}
$$

若将例 10.9 的数据代入，则有

$$AOQ = 0.01\frac{1000-65.31}{1000}$$
$$= 0.0093$$

二次抽样方案同样可以画出抽样特征曲线，同样应考虑犯两类错误的问题，这些问题与一次抽样方案原理相同，这里不再介绍。二次抽样方案的确定也是在考虑所犯两类错误及相应的不合格品率之后通过较复杂的计算得到的。为了方便使用，读者可查阅专门手册，并可通过一系统表格得到方案中的各个数值。

§10.4.3　序贯抽样检验

在抽样检验中，我们总希望在不影响判断准确性的前提下，尽可能地减少样本容量，特别是在进行破坏性检验的时候。为此，沃尔德在第二次世界大战后提出了序贯抽样理论。这种方法对样本的容量事先不做规定，而是按第一、第二的顺序逐个抽出检验。然后将每个观察值的信息累加起来，直到累加的信息能够做出接收或拒收的决定时为止。理论可以证明，应用这种抽样检验方法，可以充分利用样本数据的信息，所抽样本容量是各种抽样方案中最小的。

序贯抽样的直观原理是：设有一批产品，当不合格品率 $P=P_0$ 时，我们认为产品质量较好，愿意以 $1-\alpha$ 的较大概率接收。此时犯将 $P=P_0$ 的较好质量产品当作较差产品拒收的错误仅为 α。当不合格品率 $P=P_1$ 时，说明该批产品的质量较差，我们将以 $1-\beta$ 的极大概率拒收，而犯将 $P=P_1$ 较差产品质量当作较好产品接收的错误仅为 β。这时，我们先计算 A 和 B 两个检验临界值

$$A=\frac{\beta}{1-\alpha} \qquad B=\frac{1-\beta}{\alpha} \tag{10.11}$$

式中：A 是一个接收概率比，是一个很小的数值；β 是一个拒收概率比，是一个较大的数值。

接下来，我们就可应用序贯方法进行抽样检验了。第一步先抽 1 个产品，如这个产品是合格品，记为 $d_1=0$；如这个产品是不合格品，记为 $d_1=1$。以后也按同样的方法抽样和判断，即

$$d_i=\begin{cases}0, & \text{合格品}\\ 1, & \text{不合格品}\end{cases}$$

然后计算似然比值

$$L_1=\frac{P_1^{d_1}(1-P_1)1-d_1}{P_0^{d_1}(1-P_0)1-d_1} \tag{10.12}$$

显然，分子 $P_1^{d_1}(1-P_1)^{1-d_1}$ 表示不合格品率为 P_1 时，出现不合格品为 d_1 个的概率。分母 $P_0^{d_1}(1-P_0)^{1-d_1}$ 表示不合格品率为 P_0 时出现不合格品为 d_1 个的概率。如果第一个产品是合格品，则 $d_1=0$，似然比值 $L_1=\frac{1-P_1}{1-P_0}$。是个小于 1 的较小数值。若第 1 个产品是不合格品，则 $d_1=1$，似然比值 $L_1=\frac{P_1}{P_0}$ 是个大于 1 的较大数值。若 $L_1\le A$，说明不合格品率 $P=P_0$ 的概率较大，应作为合格

批加以接收；若 $L_1 \geq B$，说明不合格品率 $P=P_1$ 的概率较大，应作为不合格批加以拒收；若 $A < L_1 < B$，尚不能作出判断，需要继续抽样检验。在抽取了第 2 个产品进行检验后，则用下面的似然比值

$$L_2 = \frac{P_1^{d_1+d_2}(1-P_1)2-d_1-d_2}{P_0^{d_1+d_2}(1-P_0)2-d_1-d_2}$$

$$= \frac{P_1\sum_{i=1}^{2} d_i(1-P_1)2-\sum_{i=1}^{2} d_i}{P_0\sum_{i=1}^{2} d_i(1-P_0)2-\sum_{i=1}^{2} d_i} \qquad (10.13)$$

与 A 和 B 值比较并检验。若 $A < L_2 < B$，则再抽第 3 个产品。这样，抽到第 n 个产品时其检验统计量(似然比值)

$$L_n = \frac{P_1\sum_{i=1}^{2} b_i(1-P_1)n-\sum_{i=1}^{n} d_i}{P_0\sum_{i=1}^{n}\ (1-P_0)n-\sum_{i=1}^{n} d_i} \qquad (10.14)$$

从直观上看，若 n 个产品全部是合格品，似然比值 $L_n = \frac{(1-P_1)^n}{(1-P_0)^n}$ 由于 $P_1 > P_0$，则 L_n 趋向于 0。反之，L_n 趋向于∞。沃尔德证明了应用序贯抽样方法在有限步的检验后终止的概率为 1，即在有限步检验后就可做出或者接收或者拒收的判决。

在实际工作中，若每次检验都计算似然比值是比较麻烦的。统计学家将这种方法用控制图的形式来进行检验，其上下控制限的计算如下：

由于第 n 次抽检后判断为合格批的条件是

$$\frac{P_1\Sigma d_i(1-P_1)^{n-}\Sigma d_i}{P_0\Sigma d_i(1-P_0)^{n-}\Sigma d_i} \leq A \qquad (10.15)$$

两边取对数

$$\Sigma d_i \log\frac{P_1}{P_0} + (n-\Sigma d_i)\log\frac{1-P_1}{1-P_0} \leq \log A$$

$$\Sigma d_i(\log\frac{P_1}{P_0} - \log\frac{1-P_1}{1-P_0}) \leq \log A - n\log\frac{1-P_1}{1-P_0}$$

于是

$$\Sigma d_i \leq \frac{\log A}{\log\frac{P_1}{P_0} - \log\frac{1-P_1}{1-P_0}} - \frac{n\log\frac{1-P_1}{1-P_0}}{\log\frac{P_1}{P_0} - \log\frac{1-P_1}{1-P_0}}$$

记

$$h_1 = \frac{\log\frac{1-\alpha}{\beta}}{\log\frac{P_1}{P_0} - \log\frac{1-P_1}{1-P_0}} \qquad s = \frac{\log\frac{1-P_0}{1-P_1}}{\log\frac{P_1}{P_0} - \log\frac{1-P_1}{1-P_0}} \qquad (10.16)$$

上式简化为$\Sigma d_i \le -h_1+sn$ 时，为接受域。应用同样方法推导，并令

$$h_2=\frac{\log\frac{1-\beta}{\alpha}}{\log\frac{P_1}{P_0}-\log\frac{1-P_1}{1-P_0}} \tag{10.17}$$

可以求得$\Sigma d_i \ge h_2+sn$ 时，为拒绝域。序贯抽样检验的图解法 如图 10.10。

图 10.10　贯抽样的图解法图示

例 10.10　设有一批产品采用序贯抽样方法检验，规定 P_0=0.02，P_1＝0.20，α=0.05，β＝0.10。试计算出控制限。

解：

$$A=\frac{\beta}{1-\alpha}=\frac{0.1}{0.95}=0.1053 \qquad B=\frac{1-\beta}{\alpha}=\frac{0.9}{0.05}=18$$

$$\log A=-0.9776 \qquad \log B=1.2553$$

$$\log\frac{P_1}{P_0}=1 \qquad \log\frac{1-P_1}{1-P_0}=\log\frac{0.80}{0.98}=-0.0881$$

$$\log\frac{1}{A}=\log\frac{1-\alpha}{\beta}=0.9777 \qquad \log\frac{1-P_0}{1-P_1}=\log\frac{0.98}{0.80}=0.0881$$

$$h_1=\frac{0.9777}{1-(-0.0881)}=0.8985 \qquad h_2=\frac{1.2553}{1.0881}=1.1536$$

$$s=\frac{0.0881}{1-(-0.0881)}=0.0810$$

所以上下控制限分别为

(上控制限)d_n=1.1536+0.081·n

(下控制限)d_n=-0.8985+0.081·n

序贯抽样方法同样可以描出抽样特征曲线井计算平均样本容量等。尽管序贯抽样方法在理论上有很多优越之处，但实际操作起来较麻烦。因而主要应用于破坏性检验产品和数量较少价值较高产品的检验。

习　　题

1.某厂为了考察铁损情况，选用如表 1 的因素水平表进行试验。

表 1

水平	因素			
	A 退火温度(℃)	B 退火时间(小时)	C 原料产地	D 轧程分配(mm)
1	1000	10	甲地	0.3
2	1200	13	乙地	0.35

除要求考察 4 个因素的作用外，还要考察 A×B、A×C、B×C 的交互作用。

(1)如果选用 $L_8(2^7)$正交表，试排出试验方案。

(2)如果将 A、B、C、D 放在 $L_8(2^7)$表的第 1，2，4，7 列上，将 A×B、A×C 和 B×C 放在第 3，5，6 列上，所得试验结果依次为：0.82，0.85，0.70，0.75，0.74，0.79，0.80，0.87。试分析试验结果。

2.简述三次设计的方法和统计思想。

3.简述统计控制图的原理。

4.某铸件的重量抽样测量值如表 2，试画出 X—R 图并进行诊断分析。

5.试就方案(30，50，1，3)计算 P=0.05 的平均抽检量 ASN。

6.试就方案(1000；30，50，1，3)计算 P＝0.05 的平均检验件数 I。

7.设有一批产品采用序贯抽样方法，已知 P_0=0.02，P_1=0.10，α=0.01，β=0.05。试绘制抽样检验图。

第 11 章　统计决策

§11.1　引言

§11.1.1　统计决策的三要素

决策是人们为达到某一目标，从若干可能的方案(或措施、途径、行动等)中经过分析，选出最优方案的行为。在我们的日常生活和工作中，处处离不开决策。例如遇到阴天出门是带雨伞还是不带雨伞；在商店中遇到某种商品是买还是不买；从北京站到天安门广场是乘地铁还是乘公共汽车；一种新的产品是投产还是不投产等等。

要了解和研究决策问题，首先要考虑构成决策问题的三个基本要素，即策略集、自然状态集和损失函数：

1.策略集。策略集是决策者全部可供选择策略的集合。也称为决策者的行动集或行动空间。

例 11.1 某厂生产出 10000 件产品，按规定次品率必须不高于 5%才能出厂。现从该批产品中随机地抽取了 50 件产品进行检验,发现其中有 4 件次品。试问该批产品能否出厂。

解：很显然，这是一个假设检验的问题。我们先建立假设

$$H_0:P(\text{次品})\le 0.05 \qquad H_1:P(\text{次品})>0.05$$

并选择适当的检验统计量进行检验。最后会在接受或拒绝两种策略(或行动)中选择其一。因而该例的策略集可取为 $A=\{a=(a_1,a_2):a_1$ 为接受该批产品，a_2 为拒绝该批产品$\}$ 。其中 A 表示策略集，a_1 和 a_2 表示可选的策略。由这个例子可以看出假设检验问题以致整个统计推断问题可以视为广义的统计决策。

例 11.2　某木器加工厂要对是否生产一种新型组合家具做出决策。影响决策的不确定性因素是本市居民购买这种家具的户数占全市居民总户数的比例θ根据过去与这种家具类似的销售资料可得到θ的取值和先验概率，见表 11.1。

表 11.1　　某木器加工厂某种家具的销售资料

购买率θ	$\theta_1=0.01$	$\theta_2=0.02$	$\theta_3=0.03$	$\theta_4=0.04$
θ的概率 P(θ)	0.2	0.4	0.3	0.1

解：假定本市有 2 万户居民，这种新型组合家具销售后每套可净盈利 20 元。投产时的生产准备费用为 1 万元，试问该厂是否应投产。

在这个例子中，工厂的负责人也只有两种策略，即 a_1 表示投产，a_2 表示不投产。策略集为 $A=\{a_1,a_2\}$。以上两个例子的策略集都是 2 个点。对于其他决策问题，也可以是 3 个点、n 个点，甚至可以取某一实数区间或整个实数轴。

2.自然状态集。自然状态集是决策问题所有自然状态(即决策所面对的客观条件)的集合，表示为■＝{θ}。其中■表示自然状态集，θ表示自然状态。在进行统计推断时，θ一般表示未知的总体参数，则■相应地被称为参数集或参数空间。在例 11.1 中，参数就是这批产品的次品率，即θ＝P(次品)。则■={θ：0≤θ≤1}，即次品率可在[0，1]内取值。例 11.2 的自然状态集为■

$=\{\theta_1,\theta_2,\theta_3,\theta_4\}$即共有四种自然状态(四种购买率)。其他的统计推断和决策问题我们也都可按此方法找到自然状态集。

3.损失函数。任何一个决策问题，当面对某种自然状态θ而采取某种行动a时，总会产生一定的行动结果。损失函数是对决策行动结果的计量表达式，用L(θ，a)表示，其含义是当自然状态为θ的某一个值时采取某一行动所遭受到的损失。这种决策行动结果的计量是以面对某一自然状态θ而采取的最理想策略作为目标来比较的。显然，最理想的策略(最优行动)的损失为0，其他各种策略多多少少都会有一定的损失。在对决策结果进行计量和分析时，还有一种与损失函数作用相同的表示方法，即报偿值或报偿表。这是一种以实际收益(盈利或亏损)来衡量决策结果的方法，我们在本章中主要应用报偿表来进行决策分析。

在例 11.1　中，若用损失函数表示决策结果，最简单的是用 0—1 损失函数，即不论检验结论是接受 H_0。还是接受 H_1，都可能判断正确或判断错误。当我们判断正确时，损失为 0，而当我们判断错误时，损失为 1。如果要用报偿表来表示例 11.1 的判断结果，则需要知道每件次品的损失费用及其他数据，并应用前一章的方法计算平均检验件数和平均出厂不合格品率来算出报偿值。

对于例 11.2，由策略集和自然状态集所构成的报偿表见表 11.2

表 11.2　　某木器厂的决策报偿表　　(单位：元)

策　略	自然状态(购买率θ)			
	θ_1=0.01	θ_2=0.02	θ_3=0.03	θ_4=0.04
a_1 （投产）	-6000	-2000	2000	6000
a_2 （不投产）	0	0	0	0

表 11.2 中当不投产时，不论购买率如何变化，报偿表中数值均为 0，既没有收入也没有损失。而当选择投产的策略时，则随着购买率的变化报偿值也有显著变化，其报偿值是这样计算的

$$投产时的报偿值=20000\times\theta_i\times20-10000$$

当θ_1=0.01 时，报偿值是－6000 元：当θ_2=0.02 时，报偿值是－2000 元；当θ_3=0.03 时，报偿值是 2000 元；当θ_4＝0.04 时，报偿值是 6000 元。显然，购买率θ_i＝0.025 时是该厂的盈亏点，低于该购买率企业会亏损，高于该购买率企业则盈利。对于这个例子，若用损失函教表示，只要改变表 11.1 中各报偿值的符号即可。

§11.1.2　决策的分类

我们知道，决策的问题有些很简单，也有些很复杂。简单到出门时要决定是否带伞，复杂到关系到国计民生的政策是否出台、是否执行。例如我国计划生育政策的制定、改革开放政策和措施的制定、三峡工程的决定等都是很复杂的决策。从决策应用的范围和领域讲，几乎所有的科学研究和生产实际问题都离不开决策，特别是在社会、经济和政治生活中，科学决策的问题就更普遍，也更重要了。对于多种多样的决策问题，可以从不同的角度、用不同的方法进行多种分类。我们在这里仅按决策三要素的不同及其与决策方

法的联系将决策分为确定型决策、不确定型决策和对抗型决策三种。

1.确定型决策。确定型决策是指自然状态及其发生的状态都是已知的，需要决策者根据自然状态的变化及其约束条件从策略集中挑选出最优策略或行动。例如某人想买一台彩色电视机，商店中有不同价格不同型号的若干种，是买质量好价格贵的,还是买质量稍差但价格便宜的呢？是买 25 英寸大屏幕的、还是买 18 英寸一般屏幕的呢？假定购买者已知各种型号彩电的质量，这时，他需要根据他可支付的货币量、根据各种彩电的质量和价格、根据自己对彩电的其他要求从中做出决策和选择。这类决策问题通常用运筹学中的数学规划方法求一定约束条件下某一方程组的最大值或最小值。

2.不确定型决策。不确定型决策是指自然状态及其发生都是不确定的，因其不确定的程度不同又可以分为两种。一种是各自然状态发生的概串是已知的，我们在决策时可以应用这些概率来进行分析判断。如例 11.2 中，虽然我们并不清楚新型组合家具投产后的实际购买率，但我们可根据历史资料得到几种可能的购买率及其发生的概率，这些信息对我们的决策无疑是很有帮助的。当然，由于各种自然状态或多或少都有可能发生，而我们仅能根据概率做出某一种决策选择，所以任一决策都是带有一定风险的，即可能会犯决策错误，因而这种不确定型决策又称风险型决策。另一种不确定型决策是当各自然状态的发生完全未知时，即无任何概率或先验信息可供参考时的决策，这后一种不确定型决策也称为完全不确定情况下的决策。显然，在不确定型决策中，由于自然状态及其发生的可能性是未知的，势必要应用已知概率或已知信息或者应用其他方法对其发生的可能性给予推断，这无疑要应用统计方法，因而统计决策主要是讨论解决不确定型决策问题，这也是本章的主要内容。

3.对抗型决策。对抗型决策是指人与人之间、不同利益或不同军事集团之间的竞争。在决策的三要素中，前两种决策(确定型决策和不确定型决策)都存在着自然状态，可以理解为决策者(策略集)与自然界(自然状态集)之间的一种抗争。例如出门带不带伞的问题可以理解为人与自然在做斗争或在做游戏。你带了伞，它没有下雨则它胜了。而它下雨时你带了伞则你胜了。因而，美国统计学家沃尔德形象地称统计决策为“人与自然的博奕”。那么，对抗型决策就是“人与人的博弈”了，即将原决策三要素中的自然状态集也换成人或利益集团。这就是人与人之间的对抗了。例如企业之间的竞争、敌对双方的军事对抗等都属于对抗型决策。这种决策也同样可列出损失函数或报偿表，并根据这些数值应用运筹学中的对策论等方法做出决策，选择最优。

§11.2 风险型决策

风险型决策是指当自然状态发生的可能性不确定，但根据经验或理论又有若干估息可以利用时的决策。这种信息通常表现为自然状态发生的概率。例如百货公司要进一批大衣在冬季销售，在上半年就要向工厂订货。大衣销路主要取决于冬天的气温。如果天气寒冷，大衣的销量就大，需要多进一些货；如果天气暖和，销售量减少，应该少进货。显然，如果进货多，卖不出去，会造成资金积压，公司将会受到损失；如果进货少供不应求，则既给顾客带来损失也使自己应得的收入减少。这就要求百货公司根据已往年份的气温和气象预测来决定进货多少。即使我们有气象预测，也要冒测不准的风险。对于风险型决策，最常用的决策方法是应用期望值来进行决策。

§11.2.1 期望报偿

期望报偿的计算公式为

$$E(a_i)=\sum_{j=1}^{m}X_{ij}\cdot P(\theta_j) \tag{11.1}$$

式中：$E(a_i)$表示第 i 个策略(行动或方案)的期望报偿；X_{ij}表示采取第 i 个策略出现第 j 种自然状态时的报偿；$P(\theta_j)$表示第 j 种自然状态出现的概率，共有 m 种自然状态。

对于例 11.2，我们将表 11.1 中的自然状态的概率放到表 11.2 的报偿表中去，有表 11.3。

按照期望报偿的计算公式，两个策略的期望报偿分别为

$E(a_1)=(-6000)(0.2)+(-2000)(0.4)$

$+(2000)(0.3)+(6000)(0.1)=-800$(元)

$E(a_2)=(0)(0.2)+(0)(0.4)+(0)(0.3)+(0)(0.1)=0$(元)

表 11.3　　某木器厂的先验概率报偿表　　(单位：元)

策略	自然状态（购买率θ）			
	$\theta_1=0.01$	$\theta_2=0.02$	$\theta_3=0.03$	$\theta_4=0.04$
	$p(\theta_1)=0.2$	$p(\theta_2)=0.4$	$p(\theta_3)=0.3$	$p(\theta_4)=0.1$
a_1(投产)	−6000	−2000	2000	6000
a_2（不投产）	0	0	0	0

就期望报偿来看，若投产，平均损失 800 元；若不投产，平均损失为 0 元。显然决策者要选择不投产。

§11.2.2 完全情报价值

在例 11.2 的决策问题中,如果某市场调查机构愿有偿向你提供新式组合家具的市场购买率信息，你最多愿意支付多少钱来购买，这就是风险型决策中的完全情报价值。完全情报价值的计算公式是

$$\begin{matrix}\text{完全情}\\\text{报价值}\end{matrix}=\begin{matrix}\text{确定情况下}\\\text{的期望报偿}\end{matrix}-\begin{matrix}\text{不确定情况下的}\\\text{最大期望报偿}\end{matrix} \tag{11.2}$$

所谓确定情况下的期望报偿是指各自然状态下的最大报偿值的数学期望。在表 11.3 的例子中，

$$\text{确定情况下的期望报偿}=(0)\times(0.2)+(0)(0.4)+(2000)(0.3)+(6000)(0.1)=1200(\text{元})$$

这样，就可以计算出

$$\text{完全情报价值}=1200-0=1200(\text{元})$$

式中，确定情况下的期望报偿是当市场调查机构能够给我们提供购买率的准确信息时的期望报偿，即若市场调查机构告知θ＝0.01，我们决策者显然要选择不投产的策略，此时报偿值最大，为 0；若告诉我们θ＝0.02，我们也会选择-2000(投产)和 0(不投产)这两个报偿值的最大者，即不投产：若θ＝0.03 或θ＝0.04，则我们应选择投产策略，因为这时可以盈利了。在 4 种不同自然状态下的 4 个最大报偿值就是半市场调查机构提供给我们信息时的最大报偿。再将这 4 个最大报偿乘以 4 种自然状态下的概率就是确定情况下的期望报偿(平均报偿)。用这个最大收益值减去不确定情况下的最大期望报偿应该是决策者购买市场调查机构信息的最大支付额，即完全情报的价值。

§11.2.3　其他风险型决策标准和方法

由于期望报偿考虑的是在大量的重复试验中产生的平均报偿值，因此该决策标准一般只适应于下面的情况：自然状态的概率是客观分初规律的反映，比较稳定；决策不是解决一次性的问题，而是解决多次重复的问题；决策失误的后果不会对决策者带来严重的后果。如果实际决策问题不符合这些情况，原则上期望报偿标准就不适用，需要考虑其他标准和方法。

1.最大可能标准。最大可能标准是以最大可能出现的自然状态下报偿值的大小作为决策的标准。例如某服装厂看到市场上出现西装热，考虑是否在原有设备基础上增加西装生产量。现有两种策略可供选择，一是增加一套设备进行更大规模的生产，二是在原有设备基础上小批量生产。未来的自然状态也是两种情况，一是西装热继续保持下去，二是西装热降温，根据预测两种自然状态的概率分别为 0.3 和 0.7。由此可做出报偿表如表 11.4。

表 11.4　　某服装厂西服生产决策报偿表　　(单位：万元)

自然状态	根　率 P(θ)	策略 增加设备 a1	不增加设备 a2
西装热继续θ1	0.3	200	50
西装热降温θ2	0.7	—50	10
期望报偿		25	22

我们在表 11.4 下面一栏计算出了期望报偿值。从期望报偿的标准看，应选择增加设备这一策略。但从自然状态的概率看，西装热降温的可能性较大。在西装热降温时，显然不增加设备的策略较好。于是，按照最大可能标准，应选择不增加设备的策略。应该说明的是，这种决策标准适用于各种自然状态的概率相差较大，其中某一自然状态的概率显著地高于共他概率，同时期望报偿值又差别不大的情况。如果自然状态较多，而且概率差别又不大，不同方案的期望报偿值又相差较大，这时采用最大可能标准进行决策就不一定

合适。

2.效用标准。效用是某人对某事物价值的一种主观测度，在经济学与管理学中效用都有较广泛的应用。在决策问题中，效用值表示不同的决策者对同一策略结果的不同偏好程度。在某些情况下，即使是同样的货币价值量，不同的决策者对此评价也不一样，他们个人的效用值不尽相同。因此通常用比较效用值或效用的期望值来进行决策称为效用标准决策。

例如在某一地区打一口油井需投资 100 万元。投资后有两种可能，一是可能有油，可获利润 200 万元，出油的概率为 0.75；二是可能没油，100 万元投资就损失了，没油的概率为 0.25。我们根据以上条件可计算出投资的期望利润为

$$E(a)=(200)(0.75)+(-100)(0.25)=125(\text{万元})$$

对于投资者来说还有另一种无风险的投资，投入 100 万元可获利润 20 万元。如果纯用期望报偿值进行决策，显然投资者应选择投资打油井的策略。但这同样的投资和利润额，对于不同的投资者来说，其效用值是不同的。假设投资者甲有 100 万元资本，投资者乙有 2000 万元资本。投资者甲会较谨慎地进行策略的选择，即他不敢轻易地将全部资本都投在油井上。因为如果不出油，他就会破产。一般他讲，投资者甲会更倾向于不冒风险的投资。而对投资者乙来说，可能更倾向于油井的投资，因为出油的概率较大，可获利 200 万元。即使没有油，损失 100 万元对投资者乙关系也不太大。由这个例子可以看出，同样的报偿值对不同的投资者来说效用值是不同的，即投资者甲认为无风险投资的效用值较大，而投资者乙则认为带有风险但可能获取较大利润的油井投资效用值更大。对于较复杂的决策问题，可由决策者对不同策略的不同结果按其效用值排序或给分，然后按效用值的顺序进行决策。

§11.2.4　决策树

前面介绍的期望报偿决策方法是通过计算并比较期望报偿值进行的。除此之外，期望报偿决策还可以用图形直观地表示和比较，即决策树方法。决策树又可以分为两类，一类是单阶段决策，另一类是多阶段决策。

1.单阶段决策。现在我们仍用例 11.2 来介绍决策树的方法。

第一步，建树。首先用一个小方块表示决策点，由此点向右按策略数分成几条树干，并在每条树干上注明策略的内容。本例共 2 个策略，即投产和不投产，故画出两个树干。然后，在每一树干上用圆圈表示节点。由节点向右按自然状态数分出几条细枝。每条细枝上应注明自然状态的内容和概率。该例中每个策略都有 4 种自然状态，即每个节点连接 4 条细枝。最后，在细枝的末梢注明该策略在该自然状态下的报偿值。本例按此画法得到图 11.1。

图 11.1　某木器厂新型组合家具的决策树

第二步，决策。先计算各策略的期望报偿值，并注在节点上。投产的期望报偿是-800 元，不投产的期望报偿为 0。然后比较各策略的期望报偿值，将较小的期望报偿值对应的策略去掉，即剪掉投产的树干。这样，剩下的不投产策略应成为选择的决策。

2.多阶段决策。多阶段决策树实际上是单阶段决策树的重复或复合，即第一阶段决策树的末梢作为第二阶段决策树的树根，从而形成枝叶繁茂的大树。下面我们举例说明决策树在多阶段动态决策中的应用。

例 11.3 某服装厂根据市场需求情况，下半年计划增加领带生产项目。根据初步核算，如产品能顺利推销，每条领带可获利 0.4 元；如卖不出去而削价处理每条要亏损 0.3 元。现有两个方案(即策略)可供选择：一是增添新设备，第三季度可生产领带 100 万条，但成本需增加设备析旧费 6 万元；二是不增设备，采取加班办法，第三季度可生产 85 万条，但需增加加班费 3 万元。共自然状态可分为需求量大和需求量小两种情况。需求量大时可销售 140 万条，需求量小时只能销售 80 万条。方案的实施准备分两个季度进行，第四季度的方案要视第三季度的情况而定。如果第三季度添置设备后市场需求量大，则第四季度将从如下两个方案中挑选。一是继续用新设备生产，但新设备正常运转后，第四季度产量可增至 120 万条；另一方案是在新设备基础上再加班，这时产量可达 140 万条，但此时既要支付机器折旧费又要支付加班费共 9 万元。如果第三季度需求量小，则第四季度只考虑用新设备继续生产。在第三季度未添新设备的情况下，如果需求量大，则第四季度仍可考虑添置设备或加班两种方案。但若第三季度需求量小，就不再添置设备，而仅用加班方式生产。不过连续两季加班，工人疲劳，第四季度只能生产 80 万条。第四季度的自然状态，假设与第三季度无关，需求大和需求小的概率仍分别为 0.6 和 0.4。现在要求从下半年两个季度的总利润出发决定第三和第四季度备应采取的方案。

解，在多阶段决策的情况下，用列表的方法比较困难，而用决策树方法方便宜观。

第一步仍是建树。先从左到右把可能的方案和可能的自然状态及其概率用树干、树叶和树节(节点)画出。然后再算出各方案在各种可能状态下的下半年两个季度的利润总和，写在树末梢，其图形如图 11.2。

图 11.2　某服装厂领带生产决策树的建树图

图 11.2 决策树最右边树梢的报偿值都是第三季度和第四季度两个季度的利润总额，各树梢的利润计算为

第三季度利润＋第四季度利润=该方案的总利润

(1)[(100)(0.4)-6]+[120(0.4)-6]=76(万元)

(2)[100(0.4)-6]+[80(0.4)-6-40(0.3)]=48(万元)

(3)[100(0.4)-6]+[140(0.4)-9]=81(万元)

(4)[100(0.4)-6]+[80(0.4)-9-60(0.3)]=39(万元)

(5)[80(0.4)-6-20(0.3)]+[120(0.4)-6]=62(万元)

(6)[80(0.4)-6-20(0.3)]+[80(0.4)-6-40(0.3)]=34(万元)

(7)[85(0.4)-3]+[100(0.4)-6]=65(万元)

(8)31+[80(0.4)-6-20(0.3)]=51(万元)

(9)31+[80(0.4)-3]=60(万元)

(10)[80(0.4)-3-5(0.3)]+[80(0.4)-3]=56.5(万元)

第二步是利用决策树进行决策。过程是从右向左逐步计算各方案的利润总额的期望值。先计算右边第一列节点的期望利润。然后在不同策略之间淘汰掉不好的方案，即剪枝(打上双截号)，将选中的方案写在方框上。然后以方框作为树梢再计算期望利润，最后再比较、剪枝，就得到了所选定的最优方案，即第三季度采取添置新设备生产的方案，第四季度用新添设备继续生产，其过程期望利润为 59.2 万元。第二步决策过程见图 11.3。

图 11.3　某服装厂领带生产决策树的决策过程图

§11.3　完全不确定型决策

在决策时不知道各种自然状态出现的概率，显然无法采用期望报偿或最大可能标准，这类决策问题即完全不确定型决策。在这类决策问题中由于只考虑报偿值，而不同的决策者对同一报偿矩阵(报偿表)又有不同的效用值，因而对于完全不确定型决策的标准较多。现通过一例来加以介绍。

例 11.4　设某厂要对一种新产品的生产进行决策。决策者有 3 种策略可供选择，即(1)建立新车间大量生产。(2)改造原有车间达到中等产量。(3)利用原有设备小批试产。市场对该产品的需求情况有 4 种可能。(1)需求量很大，产品畅销。(2)需求量较大，产品销路尚好。(3)需求量不大，产品销路较差。(4)需求量很小，产品滞销。这 3 种策略和 4 种自然状态及各报偿值如表 11.5。

表 11.5　　某厂新产品投产的决策报偿表　　(单位：万元)

策　略	自　然　状　态			
	畅销	尚好	较差	滞销
大量生产	80	40	-30	-70
中等产量	55	37	-15	-40
小批试产	31	31	9	-1

§11.3.1　小中求大的标准(悲观准则)

小中求大的标准是从各策略最小的报偿值中选择最大报偿值所对应的策略作为决策者的决策。在本例中，3 种策略的最小报偿值分别是－70，－40，－1。其中最大值为－1，则－1 所对应的策略“小批试产”就是该标准下所选择的策略。我们看到，这种决策准则在从最小报偿值中选最大的过程实际上是从最坏处考虑，在最坏情况中选一个相对好的。因而是一种保守的、稳妥的决策过程，故也称为悲观准则，其决策结果必然选取无所作为的策略。

§11.3.2　大中求大的标准(乐观准则)

这种标准与小中求大标准恰好相反，它是从各策略最大的报偿值中选择最大的所对应的策略作为决策者的决策。在本例中，3 种策略的最大报偿值分别为 80，55， 31。其中最大值是 80，则 80 所对应的策略“大量生产”应作为本标准下的策略选择。显然，这种标准是将决策建立在最乐观的估计之上。在实际决策中，这种标准决策的风险较大。

§11.3.3　现实主义标准(折衷准则)

这是悲观与乐观准则的折衷。人们一般认为，完全悲观、完全乐观都是不现实的。现实的态度应是既不悲观，也不乐观。于是，人们提出用一个乐观系数 α $(0\le\alpha\le1)$ 将乐观与悲观标准折衷，即将乐观结果与悲观结果加权

$$\begin{matrix}\text{现实}\\\text{估计值}\end{matrix}=\begin{matrix}\text{乐观}\\\text{结果}\end{matrix}\times\alpha+\begin{matrix}\text{悲观}\\\text{结果}\end{matrix}\times(1-\alpha)\tag{11.3}$$

其中 α 值随乐观态度而定。若 α 取 1，则是完全乐观；若 α 取 0，则是完全悲观；若 α 取 0.5，则态度既不乐观，也不悲观。假设本例中乐观系数为 0.7(比较乐观)，则 3 种策略的现实估计值分别为 3 个策略中策略 1(大量生产)的估计值最大，应选择大量生产的策略。我们看到，这个标准的关键是乐

观系数α的选择。显然，α的选择是因人而异的，主观性也较强。

§11.3.4　最大后悔中求最小的标准

这种标准是从决策者决策后常产生后悔这一角度考虑问题，选择决策后最小的后悔值作为策略选择的标准。后悔值的计算，通常用各自然状态下所做决策与最优决策的差额表示。本例中，当产品畅销时，选择大量生产策略可得利润 80 万元，当然不会后悔，因为此策略是该自然状态下的最优策略。此时，如选小批试产的策略就不是最理想的策略，与最优策略相差 80－31＝49(万元)，即有 49 万元的后悔值。这样，经对表 11.5 的整理，得到与表 11.5 对应的后悔值表 11.6。

表 11.6　　某厂新产品投产的决策后悔值表　　(单位：万元)

策　略	自　然　状　态				各策略的
	畅销	尚好	较差	滞销	最大后悔值
大量生产	0	0	39	69	69
中等产量	25	3	24	39	39
小批试产	49	9	0	0	49

在表 11.6 最右列的最大后悔值列中，最小的是 39 所对应的策略“中等产量”，故按此标准应选“中等产量”的策略。显然。这种标准既不是十分悲观地选“小批生产”，也不是十分乐观地选“大量生产”，而是比较稳当地选取“中等产量”的适中标准。

以上介绍的 4 种标准都没有考虑自然状态的概率，是从不同的角度和标准考虑决策问题的。结果 3 种策略都可以选到，都是从某种主观态度出发的，在实际问题中应结合实际情况灵活选择和应用。

§11.4　贝叶斯决策

在本书的§3.3 中我们介绍过贝叶斯公式，这里讨论的贝叶斯决策是贝叶斯公式在决策方面的应用。贝叶斯决策已成为统计决策的最主要方法之一。

§11.4.1　先验概率及其确定

先验概率也称为验前概率(Prior Probability)，是在试验之前通过历史经验、书本知识和主观判断对各自然状态的发生可能性给予的概率赋值。先验概率并不是也不应该是拍脑门随意给出的，而应该是对过去客观概率的总结，应该体现自然状态发生的主观判断和历史信息。之所以称为先验信息或先验概率，是为了与抽样调查或试验之后所获信息或概率加以区别。

应用贝叶斯公式进行决策是无法回避先验概率的，如果有历史资料、理论数据等作为先验概串，当然是最理想的。如例 11.2 中某木器加工厂要生产一种新式组合家具。对于各种购买率出现的概串可以用过去与这种新式组合家具类似的市场购买率之概率代替。显然，在有历史数据资料可以利用时，应用历史资料作为先验概率是一种常用的方法。有时，当存在理论数据或总体数据时也可用来作为先验概率。如果没有较客观准确的历史资料或理论资料，我们介绍几种确定先验概率的方法。

1.等可能法。等可能法即在没有任何先验信息可以利用的情况下，可以认为各种自然状态出现的概率是相同的。假如有 k 种自然状态，那么每种自然状态的概率即为$\frac{1}{k}$。用概率分布形式表示就是

θ_i	θ_1,	θ_2,	…,	θ_k,
$P(\theta_i)$	$\frac{1}{k}$	$\frac{1}{k}$	…	$\frac{1}{k}$

这种方法也称为同等无知原则，是贝叶斯 200 多年前提出贝叶斯公式时就给出的一种先验概率赋值方法。这种方法的思想是，在我们无法对自然状态出现的概率大小给予正确的赋值时，认为各概率相同是最合理的估计或假定。

2.主观判断法。在无较客观的历史资料可利用时，主观判断法是一种常用的先验概率确定的方法。在主观判断法中，又有一个人的判断和多人共同判断之别。个人判断是决策者经过对自然状态过去发生的情况和对未来发生的估计所做出的主观判断。在判断过程中，决策者可以对自然状态中两两状态发生可能性的大小分别比较和分析，从而给出个人的主观概率。多人共同判断是为了避免个人判断主观差异过大，而请对此自然状态了解较多的专家、学者或经验丰富的人共同做出判断。由于每个人的学识经历不同，观察问题的角度不同，最终每个人所给予的先验概率数值也会不同。此时可以通过讨论得到一个比较一致的意见，也可以将各个不同的概率加权，取其平均作为这些人的代表值。

除了以上介绍的两种方法外，还可以应用共轭分布族法、广义贝叶斯估计、经验贝叶斯方法等来确定先验概率。总之，解决先验概率的问题是应用贝叶斯公式进行统计推断时最重要的问题。由于贝叶斯公式的后验概率是由先验概率和样本信息综合而成的，数理统计证明

$$\mu = W_1\bar{x} + W_2\mu_0 \tag{11.4}$$

式中：μ 是后验密度均值；$\bar{x}$ 是样本均值；μ_0 是先验概率密度的均值：W_1 和 W_2 分别是 $\bar{x}$ 和 μ_0 的权数，即 $W_1+W_2=1$。当样本容量 $n\to\infty$ 时，$W_1\to1$ 而 $W_2\to0$。这表明当样本容量充分大时，抽样信息对后验密度的影响占有主要地位，而先验概率的影响会随着 n 的增大而减小。这等于告诉我们，如果我们的先验概率是有把握的，可以抽取较小样本；如果我们的先验概率是假定的，是没有把握的，则应抽取较大样本，提高客观数据资料的影响和作用。

§11.4.2　应用后验概率的决策

我们在前面期望报偿的决策方法中，介绍的是直接利用先验概率计算期望报偿值，然后选择最大期望报偿值的策略。贝叶斯决策的特点是应用贝叶斯公式利用样本数据信息对先验概率进行修正，用修正后的后验概率计算期望报偿值，并进行决策。

例 11.5　A 城市的报纸发行部门考虑在 B 城市发行他们的报纸。经分析，发行利润主要受两个因素的影响：一是由 A 城到 B 城的运输方式，甲法是用火车运输，乙法是用汽车运输；二是 B 城 50000 户居民对该报纸的订阅率。根据历史经验并进行分析，得到如表 11.7 的先验概率和报偿表。

表 11.7　　A 城报纸发行的报偿表

B城 订户数	订阅率 （自然状态） θ_i	先验概率 P （θi）	每天利润（元） （策略） 甲法	 乙法
20000	0.4	0.3	900	1200
25000	0.5	0.5	1500	1500
30000	0.6	0.2	2200	1800

若用先验概率直接计算期望报偿值，则有

甲法：900×0.3+1500×0.5+2200×0.2=1460(元)

乙法：1200×0.3+1500×0.5+1800×0.2=1470(元)

显然，按先验概率计算的期望报偿应选择乙法运输，即选用汽车运输方式。然而，历史经验毕竟只代表过去，仅可作为当前决策的参考，还需要样本数据进行补充。贝叶斯公式恰恰提供了将先验信息和样本信息结合起来的方法。对于本例，要了解 B 城实际订阅情况，有效的办法就是进行抽样调查。现从 B 城 50000 户居民中随机地抽取了 25 户，向他们提供了 A 城报纸样品，其中 14 户表示愿意订阅。在有了样本信息的情况下，可用贝叶斯公式对先验概卒进行修正或调整。我们令 X 表示样本信息，则有贝叶斯公式

$$P(\theta_i|X) = \frac{P(\theta_i)P(X|\theta_i)}{\sum_{j=1}^{3} P(\theta_i)P(X\theta_i)} \tag{11.5}$$

式中：$P(\theta_i|X)$ 是后验概率，即以样本信息 X 为条件的订阅率；$P(\theta_i)$ 是先验概率；$P(X|\theta_i)$ 是样本信息，即当 θ 分别为 0.4，0.5 和 0.6 时抽取 25 户其中有 14 户订阅的概率，亦即

$$P(X|\theta_1 = 0.4) = C_{25}^{14}(0.4)^{14}(0.6)^{11} = 0.0434$$
$$P(X|\theta_2 = 0.5) = C_{25}^{14}(0.5)^{14}(0.5)^{11} = 0.1328$$
$$P(X|\theta_3 = 0.6) = C_{25}^{14}(0.6)^{14}(0.4)^{11} = 0.1465$$

接下来就可以应用贝叶斯公式计算 3 个后验概率了

$$P(\theta_i|X) = \frac{P(\theta_i)P(X|\theta_i)}{\sum_{j=1}^{3} P(\theta_i)P(X\theta_i)}$$
$$= \frac{(0.3)\cdot(0.0434)}{(0.3)(0.0434)+(0.5)(0.1328)+(0.2)(0.1465)}$$
$$= 0.1197$$
$$P(\theta_2|X) = 0.6108$$
$$P(\theta_3|X) = 0.2659$$

在有了后验概率后，就应该用后验概率代替先验概率计算期望报偿值

甲法：900×0.1197+1500×0.6108+2200×0.2695=1617(元)

乙法：1200×0.1197+1500×0.6108+1800×0.2695=1545(元)

在应用后验概率计算期望报偿后，甲法的期望报偿较明显地大于乙法，即用火车运输要优于汽车运输。

为什么应川后验概率计算期望报偿后所做决策与应用先验概率计算期望报偿的决策相反呢？原因是样本信息起了作用。在随机抽取的 25 户家庭中，有 14 户表示订阅，订阅率为 $\frac{14}{25} = 0.56$。但我们不能简单地直接用订阅率 0.56 进行计算和决策，因为 0.56 是随机样本的比率，存在着抽样误差。较理想的办法是应用贝叶斯公式使样本信息和先验信息结合起来，即利用样本信息对先验概率进行修正和调整。从先验概串 $P(\theta_i)$和后验概率 $P(\theta_i|X)$的比较中可以看出，θ_1=0.4 的后验概率由 0.3 降为 0.1197，θ_2=0.5 的后验概率由 0.5 提高到 0.6108，θ_3=0.6 的后验概率由 0.2 提高到 0.2695。这是由于样本中的订阅率高达 0.56，使θ_2=0.5 和θ_3＝0.6 的可能性加大了，而θ_1=0.4 的可能性减小了。我们通过此例看到应用贝叶斯公式既考虑先验信息又利用样本信息，实际上是一种认识——实践——再认识的过程，它比单纯依靠样本信息进行决策的实践——认识过程更完善，决策的效果通常更好。

习　　题

1.某轮胎厂拟生产一种新轮胎，现有 3 种工艺设计方案可供选择，每种方案的生产费用如表 1。

表 1

方　　　　案	固定成本（万元）	每只轮胎可变成本（元）
A	60	300
B	90	200
C	120	150

据分析可能有三种销售量，分别为 4000 只、7000 只和 10000 只。这三种销售量的概率分别为 0.3、0.5 和 0.2。轮胎每只售价为 750 元。要求：

(1)按总利润编制一个报偿表；

(2)按期望报偿进行决策，哪个方案最优，

(3)计算完全情报价值；

(4)分别用“小中求大”、“大中求大”和“最大后悔中求最小”标准进行决策。

2.某儿童玩具厂拟对是否生产一种新的活动玩具进行决策。第一阶段先要考虑是否需要试销阶段，第二阶段要决定是否向全国推销。根据分析知道：如果试销，成功与失败的概卒各占 0.50；如果不试销，而直接向全国推销，成功的概率为 0.40，失败的概率为 0.60。在试销成功时，第二阶段在全国销售成功的概率为 0.80，失败的概率为 0.20；而在试销失败时，第二阶段在全国销售成功的概率仅为 0.10，失败的概率为 0.90。又知，如果在全国销售成功，可获 500 万元利润；如果失败，则亏损 100 万元。若放弃这个计划，不影响企业利润。若第一阶段试销，需投入 10 万元试销费用。要求用决策树方法进行决策。

3.在例 11.2 中木器加工厂新型组合家具的决策问题中,该厂对全市居民家庭进行随机抽样，在抽取的 20 户中，有 2 户对该新型家具感兴趣，表示要购买。试用样本信息修改先验概卒并用后验概率进行计算和决策。

第 12 章　国民经济核算

§2.1　引　言

应用经济统计方法，其应用对象是经济领域中各种可以用数量表示出来的经济现象。因此，在许多应用统计学教科书中，除了对统计方法本身进行研究外，还对作为应用对象的统计数字(也常称为统计指标数字)进行讨论。在实际生活中，各种类型的统计数字何止成千上万，但是对大多数的统计数字来说，人们一般凭直观就能理解。比如说，一个国家的人口数、钢产量。这些数字的含义从它们的名称上就能很好地理解。但也有一些统计数字，如国民生产总值、国民收入等，有自己特定的经济含义，对于这些数字，如果不加以说明，在应用时就容易犯错误。还有一种情况是，统计数字本身并不难理解，但表现这些数字的方法却有一定的难度。这时，对于应用者来说，就应该掌握这些表现数字的方法。统计数字的类型虽然很多，但本章我们只研究国民经济核算中遇到的各种统计数字的含义、关系及表列方法。

§12.1.1　五种重要的国民经济核算方法

一个国家的国民经济活动是非常复杂的，它是一个由生产、流通、消费、积累活动交织而成并且不断循环往复的过程。它既表现为各种物质产品和服务的运动，又表现为和这些实物运动相向对流的货币运动。如果将整个社会再生产过程作为一个整体，以货币为计量单位进行价值量的综合统计计算，并以帐户形式把它们反映出来，就是我们在这里讨论的国民经济核算。它是我们进行许多重要的宏观经济分析和政策研究的基础。

一个比较完整的国民经济核算中，都包括这样五种核算，即国民收入核算、投入产出核算、国际收支平衡核算、资金流量核算和国民资产负债核算。国民收入核算是国民经济核算的核心，现代的国民经济核算就是在国民收入核算的基础上发展起来的。目前，几乎世界上的每一个国家都定期地公布有关国民收入的各项数字，用于反映国情、进行经济分析和国际比较。投入产出核算，或称为投入产出分析，是美国经济学家瓦西里·列昂惕夫于 1936 年提出的一种表列数据的经济分析的方法，它在联系中研究经济发展将可能对各个部门提出的要求，为国民经济的宏观控制和计划管理提供了一种非常好的方法。它提出的时间虽然不长，发展却非常迅速。现在，世界上大多数国家都编制过投入产出表。在我国，现在也几乎每一个省都进行编制。国际收支平衡对任何国家的经济发展都至关重要。国际货币基金组织提出的国际收支平衡表，利用复式记帐的原理，科学地反映了一个国家经常收支、资本收支和黄金外汇储备变动的情况，已成为各国了解和控制国际收支的重要工具。资金流量核算反映的是各种储蓄和提留，是如何通过复杂的金融活动最后转化为投资的，社会上的各种金融资产和金融负债又是如何发生的。这对于我们了解和掌握社会上的资金梳向及其规模，搞好宏观控制，显然也非常重要。国民资产负债表则把企业资产负债表的原理应用于整个国民经济，对整个国家期初期末的资产和负债进行核算，这种核算虽然应用得还不够广泛，但在进行这种核算的国家，已经显示了巨大的作用。这 5 种最为重要的国民经济核算，我们将在本章中逐一进行讨论。

五种国民经济核算虽然都能在特定的领域中发挥重大的作用，但它们所反映的毕竟只是国民经济的某一个侧面，如果把这些核算方法结合起来，形

成一个完整的体系，来全面系统地反映国民经济的运行情况，就能使国民经济核算发挥更大的作用。1953 年，联合国统计处公布的《国民经济核算体系和辅助表》，就是经过努力的初步结果。由于条件的限制，当时公布的这个体系并不成熟，它表现的还只是国民收入和国际收支方面的一些流量。但从这个报告中，已经可以看出国民经济核算的发展方向了。该报告指出：“如果从本报告所描述的这样一个国民经济核算体系着手，就可以把生产帐户进一步划分，以便显示出各产业部门之间的商品流量，这些流量正是投入产出研究的中心特征。同样地，通过对各帐户作出适当的详细说明，就有可能把全部有关的金融流量引入本体系，再加上这个经济的各不同部门的资产负债表，那么这个核算体系的结构就可以完整了。最后，还能够把这个体系中的主要产品的流量和存量按不变价格来表示。”在这之后的 15 年间，各国的国民经济核算又有了很大的发展，不但在理论上进行了各种探索，还进行了大量的统计实践；计量经济模型作为经济分析研究的手段，得到了越来越广泛的应用；电子计算机的数据处理能力又有了突飞猛进的发展。在这种形势下，联合国组织了一个由英国剑桥大学应用经济系主任理查德·斯通教授任主席的专家小组，经过两年多的努力，于 1968 年推出了新版的《国民经济核算体系》(A System of National Accounts，简称为新 SNA)。新 SNA 完整系统地记录了国民经济的流量和存量的全面结构，既包括了旧 SNA 中有关国民收入和国际收支平衡的全部帐户，也包括了详细的投入产出表、资金流量表和国民资产负债表，它把综合程度不同的数据汇集在一起，组成了一个清晰而相互联系的体系。同时，它还对各种流量如何用不变价格表示出来进行了细致的讨论，以便把各种数据组织成为可以进行动态分析的时间序列。《国民经济核算体系》中指出：新体系的提出，像旧 SNA 那样，是为那些愿意改进、详尽描述和扩大他们的国民经济核算和他们的基本统计体系的各国机构提供国际准则。由于它把全部流量和存量的定义和分类都结合起来并连接到一个统一的结构之中，新 8NA 就成为一个极好的工具，用来从事计划经济和社会分析所需大批有关的基本数据的搜集和整理工作。这一体系的发表，极大地促进了各国政府统计工作和经济科学研究的发展，斯通教授因为这一贡献，被授于 1986 年诺贝尔经济学奖。联合国的新 SNA 发表后，各主要发达资本主义国家都开始了由旧的国民收入统计向新 SNA 的过渡。现在，已经过渡到新 SNA 的国家，全世界已达 40 多个。

§2.1.2　我国国民经济核算的发展

我国的国民经济核算一直是比较薄弱的。过去，我们是按照原苏联的办法，着重核算物质生产部门的劳动产品，并按此范围观察经济结构和增长速度。这套方法的核算范围比较狭窄，手段也过于简单，满足不了我国国民经济管理和国际间比较的需要。根据国务院统一核算领导小组的决定，近些年来，我国也开始进行对国民经济核算体系的研究，并且考虑如何在我国建立这样一个体系。就实际情况看，我国过去并不存在一个按旧 SNA 的要求进行的国民收入统计，国民生产总值的统计也刚刚开始，正在摸索经验。因此，核算体系必须从头建立。国家统计局花了近 3 年的时间，组织各方面力量调查研究，反复讨论，综合了各方面的意见和观点，制订了《我国国民经济核算体系方案》，作为今后的国民经济核算工作的框架。

这个核算体系共包括 6 大部分 19 张表，它吸收了联合国的《国民经济核算体系》中的某些建议和方法，也保留了原有的原苏联式的平衡表，同时在

分类时采取了介于 SNA 和原苏联东欧各国的《物质产品平衡表体系》(英文简称 MPS)之间的分类方法。这样做的原因，当然是考虑到了我国的特点，以使现在的方案便于人们接受和在统计工作中切实可行。该方案指出：这个方案的指导思想，是“以马克思主义理论为指导，从我国社会主义初级阶段的实践出发，吸收国际上两大核算体系的长处，建立具有我国特色的国民经济核算体系。”

应该说明的是，我国的国民经济核算体系和联合国的国民经挤核算体系虽然使用的是同样的名称，但这两个“核算”的概念是不同的。联合国的国民经济核算体系的英文原名是 A System of National Accounts，这个 Accounts 的本来意思是会计帐户，而 National Accounts 可以翻译成国民帐户、社会帐户等，它指的是用会计记帐的方法，以货币来计量和反映整个国民经济的活动情况。“国民经济核算体系”是一套由各种宏观会计帐户构成的一个在数量上相互衔接的体系。而我国的国民经济核算体系中的“核算”二字实际上是指统计计算，它包括和国民经济有关的各种以实物和货币单位表示的流量和存量的计算。它们可能用帐户表示出来并在帐户间建立直接的联系，也可能不用帐户表示，相互间并不在数字上衔接。为了避免混淆，最好的方法是在名词上把这两个体系区别开来，或者对联合国的体系采用新的译名(如国民帐户体系)，或者对我国的体系换一个名字。但在目前情况下，只好在使用中加以必要的说明。

另一方面，我们也将看到，我国的国民经济核算体系中，对于价值量的核算，使用的方法和联合国体系的方法在原理上是一致的，这正是我国的国民经济核算体系中比较复杂的部分。而对于实物量的核算，在统计工作中虽然可能存在各种各样的困难，但指标本身却是比较容易理解的。因此，本章我们在讨论了五种该算方法之后，将着重讨论如何把这些方法结合进一个完整的体系，对国民经济活动中的价值量进行核算。这种核算的含义，和联合国的国民经济核算体系的含义是一致的。在此之后，我们再把核算的范围作进一步的扩展，把价值量以外的内容也包括进来。这正是我国的国民经济核算体系的特征。

§12.2　国民收入核算

目前在世界上，由于对“生产”这个概念的不同理解，在进行国民收入统计时，存在着两种不同的方法。反映在联合国的《国民经济核算年鉴》中，就是“中央计划经济”国家实行的“物质产品平衡体系”(即 MPS)和“市场经济”国家实行的“国民经济核算体系”之间计算国民收入上的差别，这也是两个体系之间最基本的差别。在 1981 年的《国民经济核算年鉴》中提供的“国民收入”，属于“国民经济核算体系”的有欧、美、亚非拉的 144 个国家，属于“物质产品平衡体系”的有原苏联、东欧备国及古巴等 10 个国家。我国在新的核算体系建立起来之前，也是按照“物质产品平衡体系”的表式向联合国提供有关数字的。这两个体系在统计口径上的差别主要在于，“物质产品平衡体系”认为生产就是物质产品的生产，不包括服务活动，而“国民经济核算体系”则认为生产活动不仅应该包括物质产品的生产，还应包括服务。在统计史上，前一种看法被称为“限制性生产”的观点，后一种看法则被称为“综合性生产”的观点。在本章中，我们先讨论“中央计划经济”国家的国民收入统计，然后再讨论“　市场经济”国家的国民收入统计。

1.“中央计划经济”国家的国民收入统计。联合国年鉴中的“中央计划经济”国家，主要是指原苏联、东欧等社会主义国家。在这些国家中，人们一直认为，马克思的再生产理论，指的是物质生产领域的再生产，而以马克思主义的再生产理论为理论基础的国民收入统计，当然只应该在物质生产领域中计算国民收入。

按照这些国家经济学家们的一般理解，国民收入这个概念的涵义可以表述如下：国民收入，从价值方面看，是全社会物质生产领域的劳动者在一定时期年(例如一年)新创造的价值总量；从使用价值方面看，是新创造的物质财富的总和。

从国民收入和社会总产品的关系看，国民收入是全社会物质生产领域一定时期内所生产的社会总产值的一个组成部分。一年的社会总产值是当年消耗的物化劳动和活劳动所创造价值的总和，即 $c+v+m$，其中物化劳动用于补偿垫付的生产资料资金，即在生产中消耗了的生产资料的价值等于 c；新投入的活劳动所创造的价值，就是国民收入，即 $v+m$。这也就是说，在价值上，从社会总产值中拍除了物质消耗价值后余下来的价值，就是国民收人；在使用价值上，从社会总产品(即社会产品总量)中扣除了消耗的生产资料后所余下来的产品，就是实物形态的国民收入，即新创造的物质财富的总和。

国民收入统计研究国民收入的生产、分配、再分配和最终使用的各个阶段。与之相适应，国民收入的计算方法也有生产法、分配法和最终使用法之分。采用生产法和分配法计算的对象，都是国民收入生产额。因此，用这两种方法计算的结果应当是一致的。但用最终使用法计算的对象是国民收入最终使用额而不是国民收入生产额。因此，用最终使用法计算的结果与用前两种方法计算的结果往往不同。

(1)计算国民收人生产额的“生产法”。国民收入是由工业、农业、建筑业、运输业、邮电业、商业、公共饮食业及物资供销业(包括仓储业)等几大物质生产部门的劳动者的活劳动所创造的。

用生产法来计算国民收入生产额，就是按照国民收入的形成过程，用生产成果扣除物质消耗来计算国民收入的生产量。

由于国民收入是社会总产值的一个组成部分，社会总产值又是由国民经济的各个物质生产部门的总产值构成的，所以要计算整个国民经济的国民收人生产额，可以先从各部门的总产值中，分别扣除共生产中的物质消耗，得到各部门的净产值，然后再把各个部门的净产值加总在一起，就得到了国民收入生产额。这种从社会总产值中扣除物质消耗，以求得国民收入生产额的计算方法，通常就称为“生产法”。其计算公式为

$$\begin{aligned}\begin{matrix}\text{国民收入}\\\text{生产额}\end{matrix} &= \begin{matrix}\text{各物质生产部门}\\\text{净产值的总和}\end{matrix}\\ &= \sum\left(\begin{matrix}\text{各物质生产}\\\text{部门的总产值}\end{matrix} - \begin{matrix}\text{各物质生产部门}\\\text{物质消耗的价值}\end{matrix}\right)\\ &= \begin{matrix}\text{各物质生产部门}\\\text{总产值的总和}\end{matrix} - \begin{matrix}\text{各物质生产部门的}\\\text{物质消耗价值的总和}\end{matrix}\\ &= \text{社会总产值} - \text{物质消耗}\end{aligned}$$

(2)计算国民收人生产额的“分配法”。根据国民收入初次分配指标倒算国民收人生产额的方法，通常称为“分配法”。用公式表示为：

$$\text{国民收入}=v+m$$

式中：v是劳动者的劳动报酬；m是劳动者的劳动为国家、集体等创造的财富。具体他说，在社会主义国家，国民收入包括物质生产领域内劳动者的原始收入和各种经济类型企业纯收入及农村经济纯收入两大部分。

劳动者的原始收入，指的是物质生产部门劳动者的劳动报酬，它包括国营企业、集体企业和其他经济类型企业职工的工资和提取的职工福利基金，农村居民净收入和城乡个体劳动者的净收入。

各种经济类型企业纯收入指的是国营企业、集体企业和其他经济类型企业的税金、利润、利.良支出及其他纯收入。农村经济纯收入主要指的是农村经济的集体提留和交纳的税金。

将以上两部分相加，得到的就是国民收入生产额，用公式表示为

$$\begin{matrix}\text{国民}\\\text{收入}\end{matrix} = \begin{matrix}\text{劳动者}\\\text{原始收入}\end{matrix} + \begin{matrix}\text{各种经济类型}\\\text{企业纯收入}\end{matrix} + \begin{matrix}\text{农村经济}\\\text{纯收入}\end{matrix}$$

从理论上说，用生产法计算的国民收入和用分配法计算的国民收入，结果应该是相等的。从公式上看

$$(c+v+m)-c=v+m$$

式中等号左边是用生产法计算的国民收入生产额，等号右边是用分配法计算的国民收入生产额，二者当然是应该相等的。但在实际计算中，常常由于各种具体处理上的原因出现不一致的现象，这种误差称为统计误差。从原理上看，生产法较好地反映了国民收入生产额的来源，应成为计算国民收入生产额的基本方法。但在具体计算上，它需要把管理工作做得更细，占有非常丰富的资料。所以在实际工作中，它使用得远不如分配法广泛。

国民收入经过物质生产领域内部的初次分配以后，还要在全社会范围内进行再分配。国民收入经过分配和再分配的结果，分别形成物质生产领域各企业、非物质生产领域各单位以及居民三方面的最终收入。一般地说，国民收入经过错综复杂的再分配后，物质生产领域的再分配收入小于其再分配支出；非物质生产领域的再分配收入大于其再分配支出；至于居民，则总是通过再分配而增加收入的。

在社会主义国家，国民收入再分配的主要渠道有：国家财政预算、银行信贷和劳务购买。国家财政以上缴税收、利润等形式，从各类企业、单位及个人那里集中一部分纯收入，有时也以公债或国库券的形式集中一部分收入。然后再通过国家财政预算支出进行分配，即通过基本建设投资、文教科学卫生事业费、国防费、行政管理费拨款等形式，分配给物质生产领域各个行业和非物质生产领域的各个部门；银行信贷则通过存款放款、支付和收取利息等进行国民收入的再分配；劳务购买则是指居民间收入的再分配，如雇请保姆等，也属于国民收入的再分配。

具体地说，国民收入通过分配和再分配形成的最终收入，可以从以下三个方面计算

物质生产领域

$$\text{最终收入} = \text{初次分配收入} + \text{再分配收入} - \text{再分配支出}$$

非物质生产领域

$$\text{最终收入} = \text{再分配收入} - \text{再分配支出} + \text{再分配余额}$$

居民

$$\text{最终收入} = \text{初次分配收入} + \text{再分配收入} - \text{再分配支出}$$

从全国来说，再分配收入之和等于再分配支出之和；物质生产领域、非物质生产领域和居民三方面最终收入的总和，等于初次分配收入的总和，即等于国民收人生产额。其计算公式为

$$\text{物质生产领域的最终收入} + \text{非物质生产领域的最终收入} + \text{居民的最终收入} = \text{最终收入的总和}$$

$$= \text{国民收入初次分配收入的总和} = \text{国民收入生产额}$$

(3)计算国民收入最终使用额的“最终使用法”。国民收人生产额经过分配与再分配，形成最终收入总额。最终收入总额加上进口额减掉出口额，就是国民收入可供使用额。国民收入可供使用额减去本期损失，就等于国民收入最终使用额。国民收入最终使用额按其最终使用的用途，则可分为消费基金和积累基金。

消费基金是当年生产的国民收入的一部分，是以货币表现的用于居民个人和社会集团实际消费的那部分物质产品，它的实物形态是消费品。积累基金是当年生产的国民收入中的另一部分，是以货币表现的用于扩大再生产、增加非物质生产领域的非生产基金和建立物资储备的那部分物质产品，它的实物形态是扩大再生产用的生产资料和当年未消费的消费品。用公式表示，国民收入最终使用额为

$$\text{国民收入最终使用额} = \text{消费总额} + \text{积累总额} = \text{消费基金} + \text{积累基金}$$

2.“市场经济”国家的国民收入统计。在西方经济统计中，“国民收入”这个概念有两种含义：一种是广义的国民收入，它诣的是一套反映国民经济生产、支出、分配情况的指标体系，包活国民收人生产额(即国民生产总值)、国民收入分配总额、国民支出总额等；一种是狭义的国民收入，即广义国民收入指标体系中的国民收入分配总额。

按照西方国家在计算国民收入时所持的观点，国民生产、国民收入和国民支出是国民经济活动中的三个方面。或者说，有关货物和服务的经济活动可以从生产、分配和使用三个方面来考察，而国民生产总额(或称为国民生产总值)、国民收入分配总额和国民支出总额则是对这三方面活动计量的结果。

(1)国民生产总额的计算。国民生产总额指的是以市场价格(生产者价格)表示的一个国家在一定时期(通常是一年)内整个经济所生产的物质产品(或称为货物)和服务的价值的总和。计算国民生产总额常使用的方法是生产法(即部门法)，即按各个部门的生产成果汇总综合计算国民生产总额。具体地说，先计算出各部门的企业增加值，然后再汇总求得整个经济的国民生产总额。

下面分别说明不同部门的增加值的计算方法。

a.农业、矿业、制造业、建筑业等部门，直接按其生产的增加值计算。对于一个企业而言,它的增加值为它的总产出的价值减去其中间投入(不包括折旧)的价值。从价值上看，它包括工资、利润、税金和折旧这四大部分。把一个部门的所有企业的增加值加在一起，就可以得到部门的增加值。

b.商业、运输业、生活服务业等部门，则从部门、企业的营业总收入中，减去为营业用而购自其他部门的产品和服务之后，所获得的“ 增加值 ”(即纯收入)。

c.对于卫生、教育等部门及非营利组织的服务部门(如慈善机构)、政府部门(包括中央和地方政府的行政管理部门及军事部门等)，均以部门、机关的工作人员的薪金工资来度量工作服务的产值，作为部门的增加值。

d.对于银行及类似的金融机构的增加值的计算，是一个值得特别说明的问题。这种增加值的特别处理，是各国在统计实践中创造的，现在已被国民经济核算体系所采用。从性质上说，银行和其他类似的金融机构进行存款和贷款活动，可以取得两种收入，一种是发放和收回贷款以及进行其他银行业务的服务费用；一种是利息收入。服务费用只占银行和类似的全融机构(以下简称银行)收入的一小部分。银行的主要收入，来源于它们所得到的财产收入超过它们所付出的财产收入的部分。这一部分基本上由利息构成。由于银行办理存款和取款业务不收费或收费很少。同时还需支付各种为存款或取款而发生的各种费用，如果把银行的交易也像其他产业部门那样处理，它们的营业盈余就可能会是负数。

在各国统计实践中，这种异常现象常常通过在实际支付的费用之外估算一种服务费用加以避免。它等于银行通过贷款所取得的收入以及用掌握的存款进行投资所取得的收入，超过在这些存款上所支付的利息那一部分。在原则上，这笔服务费用不应包括由银行自有资金进行投资所得到的财产收入，但在实际处理中，考虑到各种具体困难，这笔收入往往也同时包括在内。

估算的服务费用一方面是银行的产出，另一方面也应看成是各个产业部门的中间消耗。因此，在计算国民生产总额时，各产业部门对银行支付的利息减去由银行获得的利息之差，就应该从部门增加值中扣除。但实际上往往不容易做到。所以，在计算国民生产总额时，还应该重新把这一部分从总量上扣除。

e.此外，对一些非企业、机关、单位组织的服务活动，如房屋租用(包括自用)、家庭服务等，也分别按服务的产值计算增 加值。具体计算方法是，房屋按房租减维修费计算(自有房屋按 类似的房屋估算房租) 家庭服务则按

服务工人的工资计算。

将上述所有各部门的增加值综合汇总，就可以得到一个国家 在其国土上所有的部门所生产的货物和服务的总额数字，这个数 字称为国内生产总额(或称为国内生产总值)。

国内生产总额不完全归该国居民所有，如果外国的投资者进 行了投资，他将要汇回利润，如果外国的劳动者取得了报酬，他 将要取走工资，这种由外国国民投入生产要素而获得的收入，称 为付给国外的要素收入；反之，本国的生产者如果在国外投入了 生产要素(资本、劳动等)，他将从国外获得要素收入，这一笔 收入并不包括在国内生产总额中。因此，如果从国民的收入这个 角度上来考察一国的生产总额，就应该在国内生产总额中扣除付 给国外的要素收入，这样得到的结果，才是国民生产总额。

这一关系可以用以下公式表示

$$\text{(a)}\quad \text{国内生产总额} = \sum_{i=1}^{n}(\text{增加值})_i$$

式中：n 表示国民经济的部门数，(增加值)表示各部门的 增加值。

$$\text{(b)}\quad \begin{matrix}\text{国民生}\\\text{产总额}\end{matrix} = \begin{matrix}\text{国内生}\\\text{产总额}\end{matrix} - \begin{matrix}\text{付给国外}\\\text{的要素收入}\end{matrix} + \begin{matrix}\text{得自国外}\\\text{的要素收入}\end{matrix}$$

$$= \text{国内生产总额} - \text{国外要素收入净额}$$

国民生产总额可以用现行价格和固定价格两种价格进行计 算。在用固定价格计算的国民生产总额的时间序列中，已剔除了价格变动因素的影响，只包含物最变化因素，这样就能比较好地进行各种动态分析。以固定价格计算的国民生产总额是以某一年(如 1975 年)的现行价格为准，对其他年度的国民生产总额及其部门构成按该年的市场价格进行估算的结果。它通常是以某一年(如 1975 年)为基年编制专用的综合价格指数，通常称为国民生产总额换算系数，并通过这些系数消除价格变动的影响。例如，以 1975 年为基年编制换算系数后，再用各年的以现价计算的各部门增加值除以这些换算系数，并把得到的结果加以汇总，就可以得到各年的以 1975 年的固定价格计算的国民生产总额。用公式表示，可以写成

$$\begin{matrix}\text{按某年固定价格计}\\\text{算的国民生产总额}\end{matrix} = \sum \frac{\begin{matrix}\text{按现价计算的}\\\text{各部门增加值}\end{matrix}}{\begin{matrix}\text{相应各部门的国民}\\\text{生产总额折算系数}\end{matrix}}$$

(2)国民支出总额的计算。国民支出总额是一个国家的居民、企业、政府在一年中为了最终消费和积累所支出的全部金额。由于生产和使用是同一个事物的两个方面，所以生产总量应等于使用总量。但因为生产的产品中有一部分将用于出口，本国将无法使用，而使用的产品中，除了包括本国生产的以外，还有从国外进口的(它未被包括在国民生产总额中)。所以，国民生产总额和国民支出总额在数量上往往是不等的(相差了一个净出口)。

国民支出总额是按照“ 支出法 ”来计算的，即按照居民最终消费支出、政府最终消费支出、固定资本形成总额和储备的增加，加总计算国民支出总额。由于各种最终消费支出都是使用者的最后支出，它们都应该按购买者价格进行计算。

(3)国民收入分配总额的计算。国民收入分配总额指的是整个国民经济

中，一切国民通过提供生产要素(土地、资本、管理、劳动)而获得的收入。从国民生产总值的价值构成来看，它主要包括 4 项内容，即雇员报酬、营业盈余、间接税和固定资本损耗。这些内容并不完全包括在国民收入分配总额中。国民收入分配总额的统计，有两项基本原则：第一，它必须提供生产要素获得的收入；第二，它必须是个人的收入。

下面我们将分别考察国民生产总额中的 4 项内容。

先看雇员报酬和营业盈余。雇员报酬指的是雇员为雇主提供劳动而获得的收入。营业盈余是指个人通过提供财产而获得的利润、利息、地租等收入，这两项均应计人国民收入分配总额。

间接税主要包括销售税、货物税和不动产税。这些税虽然是由工商企业支付的，但是企业将把这些税额计入产品的售价中转嫁给消费者。这部分收入虽然是由企业生产的，但却由国家收走了，不再是个人的要素收入，所以不计入国民收入分配总额。

固定资本损耗或折旧是企业在生产中提取的、用于更新损耗掉的固定资本的价值，它将由企业保存，不用于价值分配，虽然它是当年生产的，也不计人国民收入分配总额。

这样，在国民收入分配总额中，实际上只包括两大部分，即雇员报酬和营业盈余。这样得到的国民收入分配总额，常称为按要素成本计算的国民收入。

应该说明的是，这里所说的要素成本价格，指的并不是哪一种产品具有一种被称为要素成本的价格，而是指上面所说的核算原则，它和计算国民生产总额(或国民支出总额)时采用的市价，恰恰是相互对应的。

在国民收入分配总额中，从大的项目上说，主要包括雇员报酬和营业盈余，但在具体计算时，还有一些小的调整，一是津贴减当前政府企业盈余，二是工商业转让支付。津贴指的是政府给予私营企业或个人的补助金。这些费用并不是由那些受补助的企业或个人生产出来的，所以不包括在它们的增加值中，但这些企业之所以得到了补助，是因为他们进行了生产(如农民生产了谷物)，而且得到的补助往往和他们的产量联系起来，所以这是一项要素收入，应该把它们计入国民收入分配总额。当前政府企业盈余，主要指的是国营企业的营业盈余，这是一项生产成果，当然会包含在增加值中并计人国民生产总额。但由于它是政府的收入，不属于个人，所以要在国民收入分配总额中扣除。那么，为什么要把津贴和当前政府企业盈余放在一起呢？这主要是因为政府企业盈余一般都用于对某些需要扶持的企业或个人进行补助，二者的关系比较密切。工商业转让支付，指的是工商企业在未获得货物或服务的情况下，无偿地支付给个人或团体的金额。企业之所以有能力进行这种支付，是因为它进行了生产，这部分支付的金额当然也包括在增加值中，但由于这部分支付和生产要素无关，尽管它是支付给个人的，仍然不包括在国民收入分配总额中。

按要素成本价格计算的国民收入或国民收入分配总额主要采用收入法(或分配法)来计算，按照各要素参加生产分配的份额，分别计算各部门工资、薪金、利润、利息、祖金的收入金额，进而汇总求得按要素成本计算的国民收入。

(4)国民生产、国民支出和国民收入指标之间的数量关系。从前面的分析可以看出，国民生产总额、国民支出总额和国民收入分配总额三者之间，有

着下列的数量关系

a. $$\begin{array}{c}\text{国民生}\\\text{产总额}\end{array} - \text{出口} + \text{进口} = \begin{array}{c}\text{国民支}\\\text{出总额}\end{array}$$

或 $$\begin{array}{c}\text{国民支}\\\text{出总额}\end{array} - \text{进口} + \text{出口} = \begin{array}{c}\text{国民生}\\\text{产总额}\end{array}$$

b. $$\begin{array}{c}\text{国民生}\\\text{产总额}\end{array} - \begin{array}{c}\text{资本消耗}\\\text{的补偿}\end{array} - \begin{array}{c}\text{间接税及}\\\text{非税性负担}\end{array}$$

$$- \begin{array}{c}\text{工商业转}\\\text{让支付}\end{array} + \begin{array}{c}\text{津贴减去当前}\\\text{政府企业盈余}\end{array} = \begin{array}{c}\text{国民收入}\\\text{分配总额}\end{array}$$

(5)国民收入核算的进一步分析。欧美等国家的国民收入统计除了计算国民收入分配总额之外，为了考察全国居民个人的购买力及具收入使用情况，还以国民收入分配总额为基础，进一步综合计算全国居民的个人收入指标及其支配使用的情况。

国民收入虽然按生产要素进行分配，但一方面各项要素收入并不完全分配给居民个人，如公司未分配的利润就留作公司企业的收入；另一方面，居民个人除去由生产活动取得收入，即初次分配以外，还取得各种“再分类收入”(如救济金、退役金、政府支付给私人的公债利息等)。因此，各国除了计算国民收入分配总额之外，还计算全国的个人收入总额。个人收入指：全部居民个人在一年内，从各种来源，通过劳动、财产等生产要素参加生产及通过“无偿转让”所得到的一切收入之和。个人收入中既包括薪金和工资的收入、业主收入、个人的租金、红利、利息等收入，也包括个人取得的那些不属于要素分配，而来自政府、企业等的转让收入。

个人收入与国民收入之间既有联系，又有区别，它们之间的数量关系，可以用公式表示为

$$\begin{array}{c}\text{国民收入}\\\text{分配总额}\end{array} - \left(\begin{array}{c}\text{公司未分}\\\text{配利润}\end{array} + \begin{array}{c}\text{公司}\\\text{利润税}\end{array} + \begin{array}{c}\text{社会保}\\\text{险费等}\end{array}\right)$$

$$+ \left(\begin{array}{c}\text{政府支付}\\\text{的利息额}\end{array} + \begin{array}{c}\text{政府转}\\\text{让支付}\end{array} + \begin{array}{c}\text{商业转}\\\text{让支付}\end{array}\right) = \begin{array}{c}\text{个人收}\\\text{入总额}\end{array}$$

为了考察居民个人购买力及其收入使用情况，欧美各国还计算居民个人经过缴纳种种个人税以后，自己可以支配使用的收入，即个人可支配收入额。它和个人收入总额的关系为

$$\begin{array}{c}\text{个人收}\\\text{入总额}\end{array} - \text{个人税} = \begin{array}{c}\text{个人可支配}\\\text{收入额}\end{array}$$

个人可支配收入额还可以进一步分解，即

$$\begin{array}{c}\text{个人可支}\\\text{配收入额}\end{array} = \begin{array}{c}\text{个人}\\\text{支出额}\end{array} + \begin{array}{c}\text{个人}\\\text{储蓄}\end{array}$$

$$\begin{array}{c}\text{个人}\\\text{支出额}\end{array} = \begin{array}{c}\text{个人消费}\\\text{支出额}\end{array} + \begin{array}{c}\text{消费者}\\\text{所付利息}\end{array}$$

§12.3　投入产出表

§12.3.1　投入产出来的结构

表 12.1 列出的是一个具有三个国民经济部门的投入产出表,我们将用这个表来说明投入产出分析的基本原理。为分析的便利，假定整个国民经济中只有这三个部门，每个部门只生产一种产品。下面我们就来考察这个表。

表 12.1 中的各行表示某一种商品的产出的总量和使用的去向。例如，从第 1 行我们可以看出，A 部门一年共生产了 103 个单位的 A 产品，即总产出为 103。这 103 的产品都使用到什么地方去了呢？从表 12.1 中可以看到，用于 B 部门的中间消耗 20，C 部门的中间消耗 45，用于各种最终需求 33(包括居民最终消费支出 10，政府最终消费支出 10，固定资本形成总额 10，储备的增加 3)，用于出口 5，合计起来正好是 103。

表 12.1 中的各列表示产品的投入或来源，从第 1 列可以看出，A 部门一共提供了 103 的 A 产品，即总投入 103。这 103 是怎么来的呢？有 100 是由 A 部门自己生产的，还有 3 是进口的。在生产过程中，投入了 30 个单位的 E 产品作为中间投入，各种最初投入即增加值为 70(包括雇员报酬 40，营业盈余 15，间接税 10，固定资本损耗 5)，生产过程的总投人为 100。

投入产出表的这种结构是很科学的，它客观地反映了国民经济各部门间的相互关系，同时，也是对国民收入核算的一种发展。利用它所提供的数字，可以计算各种国民收入指标。应该说明的是，由于投入产出表反映的是一个经济(国家或地区)的投入产出联系，它的立足点在国内，所以利用它计算出来的都是以国土原则核算的指标，即国内生产总额和国内支出总额，而要计算国民收入分配总额，还必须在表中的雇员报酬和营业盈余中，

表 12.1　　一个具有三个部门的投入产出表

		中间消耗			最终需求				出口	总产出
		A	B	C	居民最终消费支出	政府最终消费支出	固定资本形成总额	储备的增加		
中间投入	A	0	20	45	10	10	10	3	5	103
	B	30	0	30	95	20	10	10	10	205
	C	0	80	0	20	10	28	10	6	154
增加值	雇员报酬	40	50	42						
	营业盈余	15	25	16						
	间接税	10	15	11						
	固定资本损耗	5	10	6						
进口		3	5	4						
总投入		103	205	154						

表 12.2　　按经济活动分类的国内生产总值

部门	增加值
A	70
B	100
C	75
国内生产总值	245

加上本国居民从国外获得的雇员报酬和营业盈余，再减去外国居民从本国获

得的雇员报酬和营业盈余。表 12.2 和表 12.3 列出的，就是该经济生产、分配和支出方面的情况。

这两个表中的数字，都是根据表 10.1 中的数字加工而来的。表 12.2 中 A 部门的增加值，等于表 12.1 中第 1 列的国内总投入 100 减去中间投入 30，依次类推。在表 12.3 的左边，各项数字是投入产出表中有关项目按行汇总的结果，如雇员报酬 132.是 A、B、C 三个部门的雇员报酬相加的结果(132＝40＋50＋42)，左边一栏的数字，则是投入产出表中有关内容按列汇总的结果。如居民最终消费支出，就是居民消费的三种产品的总和(125=10＋95＋20)。

表 12.3　　国内生产总值的成本和使用构成

雇员报酬	132	居民最终消费支出	125
营业盈余	56	政府最终消费支出	40
间接税	36	固定资本形成总额	48
固定资本损耗	21	储备的增加	23
		出口	21
		减：进口	12
对国内生产总值支付的费用	245	国内生产总值	245

§12.3.2　直接消耗系数矩阵和列昂惕夫逆阵

表 12.1 中各部门的关系虽然表达得比较清楚，但结构却略显复杂。为了分析上的便利，可以将这个表进行适当的压缩和调整。首先，将进口数字和出口数字对抵，得到三个部门的净出口(分别为 2，5，2)。然后，再把净出口也作为最终需求，和最终需求的其他项目归并，就能得出各种产品的总的最终需求的数字，如对于 A 产品，最终需求为 10＋10＋10＋3＋2＝35，依次类推。在投入方面，也可以把最初投入(增加值)的内容汇总，得出三个总的数字，这样，就可得到一个更为简单的投入产出表(见表 12.4)。

在表 12.4 中，对于任何一个部门或一种产品，都有

总投入=总产出

与此同时，由于各个部门的最初投入都可以表示为

最初投入=总投入-中间投入

表 12.4　　简化的投入产出表

		中间消耗			最终需求	总产出
		A	B	C		
中间投入	A	0	20	45	35	100
	B	30	0	30	140	200
	C	0	80	0	70	150
最初投入		70	100	75	245	
总投入		100	200	150		450

各个部门的最终需求又可以表示为

最终需求=总产出-中间消耗

由于总投入等于总产出，各部门的中间投入之和等于各部门的中间消耗之和。因此，对整个国民经济来说，又有

∑最初投入=∑最终需求

即最初投入的总量等于最终需求的总量。当然，对某一个部门来说，最初投入和最终需求却可以不等。

表 12.4 中的内容也可以用符号表示出来。表 12.5 列出的就是这样一个投入产出表，它是一个 n 部门乘以 n 部门的流量矩阵。

表 12.5 用符号表示的投入产出表

	中间消耗				最终需求	总产出
中	w_{11}	w_{12}	………	w_{1n}	e_1	q_1
间	w_{21}	w_{22}	………	w_{2n}	e_2	q_2
投			………		M	M
入	w_{n1}	w_{n2}	………	w_{nn}	e_n	q_n
最初投入	z_1	z_2	………	z_n		
总投入	q_1	q_2	………	q_n		

从表 12.5 中，可以看出中间消耗、最终需求和总产出之间的关系，即

$$\begin{aligned} w_{11}+w_{12}+\cdots+W_{1n}+e_1&=q_1\\ w_{21}+w_{22}+\cdots+W_{2n}+e_2&=q_2\\ &\cdots\\ w_{n1}+w_{n2}+\cdots+W_{nn}+e_n&=q_n \end{aligned} \qquad (12.1)$$

方程组(12.1)说明，一个部门的总产出，将等于各部门用于中间消耗的该部门产出和用于最终需求的该部门产出之和。

再来看中间投入和总投入的关系。表 12.5 的每一列说明，产出 q_j 的产品，需要各部门的产品作中间投入以及作最初投入的情况。显然，每个部门生产出 1 个单位的产品，需要各部门投入的产出为 $a_{ij}=w_{ij}/q_j$。例如，A 部门每生产出 1 个单位的产品，就需要 B 产品 $a_{21}=\frac{w_{21}}{q_1}=\frac{30}{100}=0.3$。通过这个方法，可以计算出所有的 a_{ij} 的数值，它们为

$$\begin{pmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33} \end{pmatrix}=\begin{pmatrix} 0 & 0.1 & 0.3\\ 0.3 & 0 & 0.2\\ 0 & 0.4 & 0 \end{pmatrix}$$

以上这个矩阵的每一个元素称为直接消耗系数，也称为投入系数，它反映了中间投入和产出之间的技术关系。而这个矩阵，就称为直接消耗系数矩阵。由于 $a_{ij}=\frac{w_{ij}}{q_j}$，所以有 $w_{ij}=a_{ij}q_j$，将它代人方程组(12.1)，得

$$\begin{aligned} a_{11}q_1+a_{12}q_2+\cdots+a_{1n}q_n+e_1&=q_1\\ a_{21}q_1+a_{22}q_2+\cdots+a_{2n}q_n+e_2&=q_2\\ &\cdots\\ a_{n1}q_1+a_{n2}q_2+\cdots+a_{nn}q_n+e_n&=q_n \end{aligned} \qquad (12.2)$$

如果用矩阵形式把上式表示出来，则有

$$A_q+e=q \qquad (12.3)$$

其中

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \qquad e=\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \qquad q=\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix}$$

将(12.3)移项整理后可得

$$q-Aq=e$$

$$(I-A)q=e$$

$$(I-A)^{-1}e=q \tag{12.4}$$

(12.4)式中 I 是单位阵，即

$$I=\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

矩阵$(I-A)^{-1}$反映着最终需求和总产出之间的变换关系，称为列昂惕夫逆阵。如果投入和产出之间的关系是稳定的，那么最终需求增加了，总产出也应增加。当然，在现实生活中，投入和产出之间的关系不可能完全是线性的，但在最终需求增量变化不大的情况下，利用这种关系，可以得出相当好的近似结果，还可以大大降低分析的复杂性，这正是投入产出法得到广泛应用的原因，在假定投入产出关系不变的情况下，或者换句话说，假定直接消耗系数矩阵或列昂惕夫逆阵保持不变，那么已知最终需求就可以求出总产出；反之，已知总产出也可以求出最终需求。

根据前面所计算的直接消耗系数矩阵，可以具体算出它的列昂惕大逆阵

$$(I-A)^{-1}=\begin{pmatrix} 1.077 & 0.257 & 0.375 \\ 0.351 & 1.171 & 0.340 \\ 0.141 & 0.468 & 1.136 \end{pmatrix}$$

如果假设每个部门的最终需求都增加 10%，即

$$e_1=\begin{pmatrix} 35 \\ 140 \\ 70 \end{pmatrix}\times 110\%=\begin{pmatrix} 38.5 \\ 154 \\ 77 \end{pmatrix}$$

那么，就可以得出所需的新的总产出的情况

$$q_1=(I-A)^{-1}e_1=\begin{pmatrix} 1.077 & 0.257 & 0.375 \\ 0.351 & 1.171 & 0.340 \\ 0.141 & 0.468 & 1.136 \end{pmatrix}\times\begin{pmatrix} 38.5 \\ 154 \\ 77 \end{pmatrix}=\begin{pmatrix} 109.9 \\ 220 \\ 165 \end{pmatrix}$$

显然，这对我们进行经济预测和产业结构的分析是非常有用的。

§12.4 国际收支平衡表

国际收支平衡表计量的是一个国家与外国之间的全部经济往来活动。编制国际收支平衡表可以了解本国与外国之间的财产、服务、资本的流动，并研究这些流动对本国的对外支付能力和对国内有效需求的影响。

具体地说，国际收支平衡表反映的是常住者同非常住者之间进行的交易。那么，什么是常住者呢？在国际收支平衡表中，常住产业部门和其他生产者是指在一国境内从事生产的单位，包括外国企业在这个国家的分支机构。常住政府团体则是指本国境内的中央和地方政府机构及驻外使领馆和军事单位。常住居民户是指在该国境内的所有个人，除去居住不足一年的外国旅客以及外国驻在这个国家的使领馆人员和军事人员。

国际收支平衡表采用复式记帐的表列方法，用贷方(收方)表示把商品、资产等转移给其他国家，取得各种货币收入(如商品出口)，用借方(支方)表示从其他国家获得商品、资产，形成各种货币支出。因此，国际收支平衡表上的贷方和借方(或收入和支出)实际上表示的是货币的贷入和借出。对利用外汇进行

国际收支结算的国家，贷方项目表示外国向本国支付外汇，借方项目则表示本国向外国支付外汇。

在国际收支平衡表中，借贷两方的全部货币往来项目概括起来可以分为：1.贸易收支；2.非贸易收支；3.转让收支；4.长期资本收支；5.短期资本收支；6.金融。这些项目反映了国际间货币收支往来的基本活动内容。下面分别加以说明。

1.贸易收支。贸易收支是由商品出口和进口所引起的货币收支。商品出口所引起的货币收入构成收支平衡表贷方的内容；商品进口所引起的货币支出则构成收支平衡表借方的内容。

2.非贸易收支。非贸易收支是由提供服务和使用服务所引起的货币收支。非贸易收支一般地分为三种：第一种为本国在国际间用运输工具和通讯工具为外国人提供服务(等于商品出口)而引起的货币收入；或本国人在国际间使用了其他国家的运输工具和通讯工具(等于商品进口)而付出的货币。第二种为旅游方面的货币收支。第三种为国际间投资造成的货币收支。向国外投资，将以利润、股息等形式取得货币收入；外国在本国投资，本国必须付出利润、股息等。这种利润、股息方面的收支，都将被归于非贸易收支。至于资本本身的流入流出，则不归在这一部分，它们将是资本收支所要表现的内容。

3.转让收支。转让收支包括不等价的物品、服务、金钱等的交换或转让。这方面的货币收支往来包括：政府和私人的汇款、各国政府间的援助、赔款、私人捐赠等等。

以上三项收支的合计，称为经常收支(或经常帐户)。

4.长期资本收支。长期资本收支也称为长期信贷往来，一般指期限 1 年以上，10—20 年不等的国际信贷往来引起的货币收支。它的收入包括：由于在国外出售有价证券、企业、房屋、土地等而获得的收入；获得新的外国借款；因外国在本国进行新的直接投资而收入的货币。而长期资本支出则与上列收入相反，指本国购买外国的有价证券；借款给外国；或在外国进行直接投资等。

5.短期资本收支。短期资本收支属于在一年以下的短期信用。短期资本收入包括：自外国获得短期信贷的货币收入；外国短期借贷资本流入的货币收入(如外国人在本国银行存活期存款、外国人购买本国短期票据等)；收回本国过去存放在国外的借贷资本(如从外国银行的活期存款中提取资金等)。短期资本的货币支出与以上列入贷方的短期资本货币收入相反，属于应列入借方的本国借贷资本流出的货币支出。

经常收支与长期资本收支之和称为基础收支，以表示一个国家国际收支的基础状态。基础收支再加上短期资本收支，则称为综合收支，表示一个国家收支的综合状态。在综合收支的下面，另列有称为，“ 金融帐户 ”的黄金外汇储备的资产和负债(6)，专门列出外汇银行部门和政府部问对外的资产和负债。金融帐户的动向和综合收支的动向经常是一致的。

这里，我们以日本统计月报中 1980 年的国际收支平衡数字为例，说明这种表的表列方法(见表 12.6)。

该国际收支平衡表(即国际货币基金组织方式国际收支原表)在横行表示出：贷方的货币收入(或外汇收入)，借方的货币(外汇)支出和收支差额；纵的方面则作出了几个阶段的收支平衡。经常收支平衡为关于商品、服务等经常往来的货币收支平衡；基础收支平衡为具有基础性质的财物、服务、资本的货币收支平衡：综合收支为全部国际收支往来的平衡，也称为“ 净清偿结算平衡 ”；黄金外汇储备的收支平衡，属于在综合收支基础

上，进一步清理其他内容(如商业银行的存款)的收支往来，因而需要由政府组织外汇收支及储运黄金，这在国际上也称为官方结算平衡。

§12.5　资金流量表和国民资产负债表

§12.5.1　资金流量表

资金流量表反映的是一国国民经济及各机构部门的资金来源和运用情况。在对生产领域的国民经济活动进行分析时，常常采用的是国民经济的部门分类。但在考察资金流量时，首先应该注意的是，使部门分类方法适应资金流量核算的需要。

在资金流量核算中，整个国民经济往往被分为六个大的部门，即非金融企业、金融机构、一般政府、居民户、私人非营利组织和国外。这种分类方法，称为国民经济的机构部门分类。之所以采取这样的分类，是因为这个表所要进行的是财务金融方面的分析，而进行财务金融活动的单位，往往是比生产某一种或某一类产品的基层单位或企业大得多的单位。例如，一个实业公司可以控制许多生产相同或不同产品的基层单位，这些基层单位可以划为相同的或不同的产业部门，而这个实业公司本身又可能是受一个巨型实业公司控制的许多实业公司中的一个。在这种情况下，就不可能按照产业部门对企业进行划分(因为一个企业往往属于不同的部门)，而必须根据各大单位在金融作用、行为和经验上的不同，将它们归入各个不同的机构部门。虽然在各国的统计实践中，对国民经济的机构部门分类略有不同，但总的来说，都是按照各个单位在财务金融活动中的决策作用而加以分类的。

资金的流动，就是通过上述各部门间的借贷关系而实现的。这些借贷关系，就是我们常说的金融债权。在西方国家，金融债权的主要形式有：黄金、外汇和特别提款权；通货和可转移存款；其他存款；证券：各种贷款；各种保险等。在社会主义国家，也存在着各种金融债权，如存款、商业信用等。这使得资金流量核算方法成为一种在各国都受到普遍欢迎的方法。

金融债权的形式是多种多样的，各个国家也不尽相同。下面我们回过头来研究资金的流量问题。每当储蓄同投资出现背离的情况，就必然会产生资金的流动。当一个部门的储蓄大于投资时，这个部门就可以运用占有金融债权的手段，如利用存款购买股票的方法，把多余的资金提供给其他部门。这些金融债权是债权人的金融资产。当一个部门的投资需要大子储蓄时，这个部门就会发生金融负债，如吸收存款，发行股票或债券，以取得需要的资金来实现投资。任何时候一笔金融债权业务的发生，对一些人是金融资产，同时必然对另一些人是金融负债。各机构部门之间都有这样的资金往来，金融机构则在调剂资金方面起着独特的作用，它是集中闲置资金以供应需要资金的部门的专门机构，所以也称金融中介。

表 12.7 列出的是一个资金流量表,我们可以利用它来观察整个国民经济和各个机构部门的资金流动情况。这里，总储蓄包括固定资本损耗，资本形成总额中也没有拍除这一部分。

资金流量表由两大部分组成。第一部分反映资金的来源和运用，第二部分反映金融资产和负债的变动。一个部门的储蓄和投资的差额，通过部门之间的借贷来平衡；整个国民经济储蓄和投资上的差额，通过同国外的惜贷来平衡。本表的结构是这样的：总储蓄等于净储蓄加上固定资本损耗，再加上资本转移支付：总投资等于资本形成加上土地矿产权，再加上金融投资净额；金融投资净额则等于金融资产与金融负债的差额；金融资产的总计等于金融负债的总计；各项金融债权的运用总计等于来源总计。

§12.5.2　国民资产负债表

资产负债表的编制对于企业来说，有着较长的历史，并且通常规定为企业必须提供的文件。而对国民资产负债表的研究却是比较晚近的事。国民资产负债表，反映的是有关一国的经济在某一时期终了时的资产、负债和净值的存量及国民财富的存量。它是生产与收支、储蓄与投资、贷入与借出的结果。一国经济在资产上的增加，是净产出的未被消费部分在存量上的投资，它通过一切部门的储蓄来筹措资金。但由于每一单位、每一部门的储蓄与其所投资而形成的有形资产往往是不相对应的，于是资金有盈余的单位或部门直接地、或是通过金融中介间接地将多余的资金贷给有所不足的单位或部门，由此而引起了金融债权债务的变动。结果，资金流量表上有形资产投资和金融投资的变动就反映到资产负债表的资产上，而借款和储蓄的变动则反映到资产负债表的负债与净值上。

表 12.8 列出的就是一个期初资产负债表(或者说是上个时期的期末资产负债表)，它的部门分类采用了和表 12.7 相同的机构部门分类，所以能和资金流量表相互衔接。

更进一步，还可以把期初情况和本期的资金流量结合起来，编制期末的国民资产负债表。从理论上说，期末的资产和负债，应该是期初的资产或负债加上本期资金流量表中的资产或负债计算出来的。但事实上，期末资产或负债同期初在价值上的差别，不能完全由资金流量表上的流量反映出来。因为在这期间有的资产本身的价值会发生变动，如某些资产和负债所依据的市场价格的变动；另外还有一些事先没有估计到的事件，如新工艺使一些机器提前报废；或者计划折旧同一些实际的正常磨损不同等等，这些都可能改变期末价值。所以，为了如实反映期末资产和负债的当时价值，有必要设置一个重估价项目，作为本期资产负债增减变化之外的一个连接期末和期初价值的项目。这样，期末同期初资产(负债)关系的公式为

$$\text{期初资产(负债)} = \text{期初资产(负债)} + \text{本期资产(负债)的变动} + \text{重估价}$$

这里，对重估价项目要作出一点说明，对金融资产和负债中的通货和存款，不存在重估价的问题，因为这类资产和负债的价值将永远按它们的帐面(票面)价值估价。

表 12.9 列出的是一个期初期末资产负债表的例子,从中可以看出所有机构部门期初和期末资产负债表之间的关系。

表 12.9　　　　　　　期初和期末国民资产负债表

	期初资产	资产的获得	资产重估价	期末资产		期初负债和净值	负债的发行	负债的重估价	期末负债和净值
通货和存款	275	12	%	287	**通货和存款**	296	14		310
证券	457	3	—25	435	**证券**	442	8	—	424
								26	
其它金融债权	517	43	4	564	**其它金融债权**	479	37	3	519
有形资产净额	661	28	42	731	**净值**	693	27	44	764
总计	1910	86	21	2107	**总计**	1910	86	21	201 7

§12.6　从五种核算到国民经济核算体系

§12.6.1　联合国的国民经济核算体系的基本原理

联合国的国民经济核算体系的结构是比较复杂的，但在对五种核算方法进行了比较详细的讨论之后，要讲清楚它的基本原理说不会很困难了。在本节中，我们将通过一个简单的例子来说明它的原理，至于这个体系本身的结构，将是一个应该进一步研究的问题。

假定整个国民经济包括十项大的流量，它们分别为 1.增加值 255；2.进口品购买额 54；3.消费品销售额 210；4.资本物销售额 47；5.出口品销售额 52；6.储蓄 27；7.向国外现期转移净额 4；8.固定资本损耗准备 19；9.来自国外的已分配要素收入净额 5；10.向国外贷出净额—1。

从以上项目可以看出，它们分别发生在生产、消费、积累和对外往来领域。下面我们就用帐户形式来反映这些活动。先看生产领域的情况。我们可以用一个生产帐户或国内生产帐户来反映(见表 12.10)。

表 12.10　　生产(国内生产帐户)

支　出		收　入	
1.总收入支付额（增加值）(9)	255	3.消费品销售额（6）	210
2.进口品额买额（18）	54	4.资本物销售额（12）	47
	52	5.出口品销售额（16）	
总　计	309	总　计	309

表 12.10 这个帐户的意思是很明白的，左边表示生产部门的支出，构成了其他领域和国外的收入，并由此得到了可供社会使用和出口的商品。将这些商品售往不同的领域，如消费品售往消费领域，资本物售往积累领域，出口品售往国外，得到的就是生产领域的收入，记在右栏。右边的收入和左边的支出正好相等。

再看消费领域的情况。我们把它反映在一个消费帐户(或称为收入支出帐户)上(见表 12.11)。

表 12.11　　消费(收入和支出帐户)

支　出		收　入	
6.消费器购买额（3）	210	9.自国生产得到的总收入（1）	255
7.储蓄（15）	27	10.减：固定资本损耗（13）	—19
8.向国外现期转移净额（19）	4	11.来自国外的已分配 要素收入净额（17）	5
总　计	241	总　计	241

从表 12.11 可以看到，生产领域支付了 255 的总收入，这笔支出现在构成了消费帐户的收入，表中括号内的(1)，说明了这笔收入和生产帐户中第 1 项内容是相对应的。在这笔收入中，有 19 的固定资本损耗将不在消费领域中分配，所以必须把它扣除。另外，居民还从国外获得了已分配要素收入净额 5。把这几项合并起来，得到的就是这个经济按市价计算的国民收入(即按要素成本计算的国民收入再加上间接税)241。这些收入都用到什么地方去了呢？从左栏中可以看出，购买了 210 的消费品，向国外现期转移净额为 4，

剩下的转为储蓄，为 27，支出也是 241。

再看积累帐户(或称为资本交易帐户)的情况(见表 12.12)。

表 12.12　　积累(资本交易帐户)

支　　出		收　　入	
12.资本物购买额（4）	47	15.储蓄（7）	27
13.减：固定资本损耗（10）	—19		
14.向国外贷出净额（20）	—1		
总　计	27	总　计	27

表 12.12 的右栏记录的是这个经济的储蓄情况，在消费帐户中它表现为支出，正好和这一帐户表示的收入相对应。储蓄 27，加上固定资本损耗 19，再加上由国外获得的 1，总数为 47，正好等于购买的资本物 47。在这一帐户中，固定资本损耗之所以用负债记在支出端，是因为在这一帐户中，积累是按净值处理的，同时也考虑了它和收入支出帐户的相互关系。

最后一个帐户是国外帐户，或称为国际收支平衡帐户(见表 12.13)。

表 12.13　　国外(国际收支平衡帐户)

支　　出		收　　入	
16.出口品购买额（5）	52	18.进口品额售额（2）	54
17.已分配要素的支付净额（11）	5	19.现期转移净额（8）	4
		20.借入净额（14）	—1
总　计	57	总　计	57

在表 12.13 这个帐户中，收入和支出的主体都是国外。可以看出，它的核算对象实际上就是国际收支平衡表的核算对象。所不同的是，为了适应整个体系的要求，它的收支方向和前面讨论过的表式的收支方向正好相反。国外帐户的右边反映的是国外的收入情况(即本国的支出情况)，国内进口了 54 的商品，也就相当于国外获得了 54 的收入。出口是国内获得了收入，但却是国外所付的支出，所以记国外支出 52。表 12.13 中其他项目的意思都很清楚，不再说明。

从这四个帐户可以看出，对于国民经济中的十项流量，可以通过把它们记人不同帐户的方法，反映它们之间的相互联系。由于每一个流量都被记录了两次，在一个帐户中反映为收入，另一个帐户则表现为支出。这样就可以通过收入和支出的关系，观察某一个领域中经济活动的发生过程；在不同的帐户中，则可以通过同一笔项目在不同帐户中的位置，观察不同领域间的相互关系。

我们还可以看到，在这四个帐户中，第一个帐户和第二个帐户进行的实际上是国民收入在生产、使用、分配上的情况；第二个帐户反映的则是资金流量表的内容；第四个帐户实际上就是一个国际收支平衡表。如果把这四大帐户加以一定的扩展，使之包括中间消耗的内容，再把国民资产负债表的内容反映进来，就能形成一个完整的国民经济核算体系。

但是，用这套帐户表现各个领域间的联系，也存在着一个缺点，这就是不同领域间的联系不容易反映得很清楚。当然，现在的帐户中只有十项登录，这样做问题还不大，但如果与数以千计的登录打交道，情况就大不相同了。

于是，我们需要考虑能否用一个更好的形式来反映这种联系。可以考虑使用一个矩阵，在这个矩阵中，行表示收入，列表示支出，每一个帐户都由一行和一列所组成。这样，每一笔流量只用一个数值就能表示，它的性质则由它所在的位置来表明。表 12.14 就是这样一个由矩阵表示的核算体系。

表 12.14　　用矩阵形式表现的国民经济四大帐户

	1	2	3	4	总计
1.生产		210	47	52	309
2.消费	255		—10	5	241
3.积累		27			27
4.国外	54	4	—1		57
总　计	309	241	27	57	

表 12.14 的第一行和第 1 列反映的就是表 12.9 中的生产帐户。可以看出，除了对每项交易的说明在这里不象表 12.9 中那么清晰之外，反映的内容则是完全一致的。但从反映帐户间的关系上看，这个形式却显得清晰得多。把生产帐户的五项交易(或者说五个流量)和其他帐户间的联系，都系统地表现了出来。将四个帐户的十项交易结合在一起，就得到了一个反映整个国民经济十项流量的国民经济核算体系。它虽然比较简单，却很好地说明了国民经济核算体系的基本原理。

表 12.14 可以进一步延伸，利用资金流量表和国民资产负债表的关系把国民经济的各种存量也包括进来。以行表示资产，以列表示负债，将矩阵的帐户扩展为 7 个。同时，再在第 2 行与第 2 列交叉的地方加上一项由生产帐户对生产帐户的支付，表示在生产过程中的中间消耗，这样就得到了一个新的用矩阵形式表现的国民经济核算体系(见表 12.15)。

表 12.15　　用矩阵形式表现的国民经济核算体系

	1	2	3	4	5	6	7
1.期初资产净额			693				
2.生产		300	210	47	52		
3消费		255		—19	5		
4.积累	693		27			44	764
5.国外		54	4	—1			
6.重估价				44			
7.期末资产净额				764			

在表 12.15 中,第 2 行和第 2 列反映的是国民收入核算和投入产出核算：第 3 行和第 3 列反映的是收入支出的核算；第 4 行和第 4 列反映的是资金流量核算与国民资产负债核算 ;第 5 行和第 5 列反映的则是国际收支平衡核算。当然，这里表现的只是一个最简单的例子。实际上这里的每一对帐户都可以进一步细分，使之反映更为纷繁复杂的经济联系。例如，在生产领域，可以将各种流量按国民经济部门进行分类，对于消费品，可以按使用去向分类；对于资产和负债，可以按性质分类，等等。这样，矩阵表 12.14 就可以扩展为很大。从理论上说，可以无限细分。联合国《国民经济核算体系》一书中 88x88 的矩阵帐户，就是这种扩展的结果。

当然，在实际工作中，要列出这样的矩阵表是不方便的。如果我们得到了一个 500 × 500 的矩阵，把它印出来就很困难。即使印出来了，阅读起来也不方便。所以，各国在公布资料时，常常还是使用复合帐户、统计表、分部门表的形式表现国民核算资料，但从总体上来说，都是以这样一个矩阵帐户为思路来搜集、整理和公布统计资料的。

搜集和整理国民核算资料，当然是为了应用。在联合国的《国民经济核算体系》一书中，列出了七个帐户可进行的经济分析和政策研究(见表12.16)。

在表 12.16 中,最左边的一列数字 1—7 和最上方的一行相应数字分别是“期初资产(纵列为负债)”、“生产”、“消费”、“积累”、“国外”、“估价调整”和“期末资产(纵列为负债)”这七大帐户的代号。而各行与各列之交的方块则说明利用相应帐户的资料所能进行的经济分析或政策研究。例如，在第 2 行与第 2 列相交的方格上就列出了两项内容，投入产出分析和生产率分析，这说明利用第 2 行和第 2 列所给出的生产帐户的资料，就可以进行这两项的研究。又例如，在第 2 行与第 3 列相交的方格上，列出了消费者需求分析和政府支出研究这两项内容，这说明将生产帐户中的横行内容与消费帐户中纵列的内容结合起来，可以进行这两项的研究。在全表的 24 个方格上，共列出了 34 项经济分析和政策研究的内容，这都是对一个国家的经济发展非常重要的研究。

§12.6.2 我国国民经济核算体系概况

我国目前新设计的国民经济核算体系，是按照社会再生产过

表12.16 与国民经济账户有关的经济分析和政策研究领域

	1	2	3	4	5	6	7
1				国民财富研究；生产率分析			
2		投入产出分析；生产率分析	消费者需求分析；政府支出研究	建立储备和固定资本形成模型；投资政策	出口需求分析		
3		生产函数；生产率分析；要素份额分析		折旧分析；投资补助	外国投资收益研究；双重税收政策		
4	净值研究		储蓄行为分析	货币正策和流动性偏好	国际金融和清偿能力；长期外援政策	酱收益和损失研究；资本收益税政策	净值研究
5		进口需求分析	短期外援政策	国际金融和清偿能力；长期外援政策	国际贸易收支平衡分析		
6				资本重估价			
7				国民财富研究；生产率分析			

程设计的，它包括以下六个组成部分：

第一部分是社会再生产条件的核算，包括人口、劳动力、周民财产和自然资源。这部分核算共有“人口平衡表”、“劳动力平衡表”、“资产负债表(一)(实物资产)”、“资产负债表(二)(金融资产)”、“自然资源”五张表。

第二部分是生产与使用的核算，即生产成果及其使用的核算，主要是以国民生产总值反映全社会各行各业的劳动成果，包括物质生产和芳务活动增加的价值，同时还反映国民生产总值使用于消费、投资、净出口的情况。这部分核算包括“国内生产总值及其使用”、“物质产品生产与使用”、“投入产出”、“物质产品与劳务综合价格指数”四张表。

第三部分是分配与流通的核算，包括“资金流量表(一)(实物交易部分)”、“资金流量表(二)(金融交易部分)”、“主要商品资源与使用平衡表”三张表。

第四部分是积累与消费的核算，包括“固定资产投资来源与使用”、“总消费”两张表，主要反映投资资金来源与固定资产构成，消费基金用向等情况。

第五部分是国际收支核算，这实际上是国际收支平衡表的核算。

第六部分是国民经济循环的核算，即以矩阵表现的表，它是关于前五部分的价值流量和存量的一个总表。

该体系的设计者认为，这个新体系同目前国际上实行的两大宏观经济核算体系相比较，具有下列特征：

1.SNA 是建立在西方经济学的理论基础之上的。按照这种理论，劳动、资本、土地三要素共同创造社会产品和国民收入：劳动创造工资，资本创造利润，土地创造地租。MPS 的理论基础是马克思的再生产理论。按照这种理论，只有从事物质资料生产者的劳动才能创造国民收入，从事非物质生产胁劳动者得到的收入以及利润与地租，是在国民收入的再分配中形成的。我国新设计的国民经济核算体系所依据的是，在社会主义制度下，在社会生产过程中，尽管同样存在着生产三要素，甚至加上管理，成为四要素，但创造价值的只有活劳动这一个要素。在社会经济活动中，按照等价交换、按劳分配的原则，租用土地要支付租金，借用资金要支付利息，使用劳动力要支付工资，但这同土地、资本、劳动共同创造阶值，各自创造收入的理论有着本质的区别。如同 c、v、m 由于生产关系不同可以作出不同的解释一样。在资本主义经济里，c 是不变资本，v 是可变资本，m 是剩余价值。可是，在社会主义经济里，c 反映的是物化劳动——固定资产和流动资产的消耗，v 反映的是必要活劳动的消耗，m 反映的是剩余劳动创造的剩余产品的价值。c 是转移价值，v 和 m 是活劳动新创造的价值。活劳动是价值的唯一源泉。这是进行国民经济核算时必须坚持的马克思主义的基本原理，但是，这并不意味着提供劳务的劳动不创造价值，不创造自己的收入。

一切经济活动、科技活动、社会活动，不论物质生产活动，还是非物质生产活动，都需要投入劳动，才能有产出，取得成果。在商品、货币条件下，对于投入的劳动，通过货币这个同度量测量出来，就可以赋予它以具体的价值量。生产物质产品的劳动创造价值，创造自己的收入；提供劳务的劳动也创造价值，也创造自己的收入。所不同的是，前一种劳动凝固在物质产品上，后一种劳动体现为活劳动，从事前一种劳动的劳动者获得的收入，形成狭义的国民收入，而包括两种劳动者的劳动所创造价值在内的收入，将形成广义

的国民收入。

2.SNA 不区分物质生产与非物质生产，MPS 区分物质生产与非物质生产，但非物质生产的收人只是国民收入的再分配。我国新设计的国民经济核算体系，对物质生产与非物质生产、既分别核算它们的所得与所费，也综合核算它们的所得与所费。

3.SNA 偏重于价值量的核算，MPS 偏重于实物量的核算，我国新设计的国民经济核算体系，对实物量与价值量、物流与钱流的核算，既分别进行，又相互挂钩。

4.SNA 编制投入产出表，主要用于编制模型，开展预测。MPS 编制投入产出表，主要用于分析部门之间的经济技术联系，调整产业结构，也利用它来进行预测。我国新设计的国民经济核算体系中的投入产出表，除了上述用途以外，更重要的是要分析如何少投入，多产出，以提高经济效益。

5.SNA 主要采用帐户归集资料，MPS 主要采用平衡表归集资料，我国新设计的国民经济核算体系是根据实际情况来确定归集资料的方式适宜采用平衡表归集资料的，如人口、劳动力，就采用平衡表的方式；适宜采用帐户归集资料的，如收支流量、资产负债，就采用帐户的方式。

习　　题

1.某企业报告月生产成果如表 1。

表 1　　　　　　　　　　　　　　　　　　　　　　　　（计算单位：千元）

	产值	用于本月进一步加工	销出厂外
生铁	2500	2400	60
钢	3100	3000	
钢材	3900		3900

并知该月生产耗用原料价值为 10000，试计算该企业本月总产值、增加值。

2.根据以下资料按中央计划经济国家所使用的口径计算国民收入及其构成。

(1)工业总产值 270 亿元；

(2)农业总产值 10 亿元；

(3)建筑安装工程产值 50 亿元；

(4)房屋及其他建筑物的大修理价值 10 亿元；

(5)交通运输业收入总额 25 亿元，其中货运占 80%，

(6)邮电业总收入 10 亿元，其中为生产服务的占 75%；

(7)商业、公共饮食业等销售收益 60 亿元，其中：支付贷运费 7 亿元，邮电费 3 亿元；

(8)物质财富的生产性消费(物质消耗)：

工业 180 亿元；

农业 3 亿元；

建筑业 30 亿元；

货运邮电业 9.5 亿元，

商业公共饮食业等 14 亿元。

(9)农村经济有关资料：

农、林、牧、副、渔产值合计 190 亿元；

工业产值 10 亿元；

建筑业产值 5 亿元；

货物运输业产值 2 亿元；

生产性费用(物耗)：工业 4 亿元，农业 51 亿元，建筑业 2.5 亿元，运输业 0.4 亿元；

管理费用 2.5 亿元，其中物品购买支出 2 亿元，根据调查，农业占 80%，工业占 10%，建筑业占 5%，运输业占 5%。

3.试根据下列资料计算美国 1976 年国民生产总额、国民支出总额、国民收入分配总额(计量单位：10 亿美元)。

(1)工资和薪金 890.4；(2)辅助工资和薪金 138；(3)农业企业主纯收入 22.8；(4)非农业企业主纯收入 73.8； (5)纳税前公司利润 117.8；(6)公司上缴利润税 64.4；(7)租金收入 23.5；(8)利息收入 82；(9)商业转让支付 7.1；(10)间接税 149.7； (11)津贴减政府企业盈余 1.2； (12)固定资本损耗

179.8； (13)国外要素收入净额 13.4；(14)出口 163.3；(15)进口 159.6。

4.根据表 2 的资料，计算直接消耗系数矩阵和列昂惕夫逆阵。若 A 部门最终需求增加 5%，B 部门增加 3%，这两个部门的总产出应为多少？

表 2 投入产出表

		中间消耗		最终需求	总产出
		A	B		
中间投入	A	30	40	20	90
	B	25	35	25	85
最初投入		35	10	45	
总投入		90	85		175

参 考 文 献

[1] 胡孝绳著：《统计学》，木屋学社出版社 1976 年版。

[2] 高庆丰著：《欧美统计学史》，中国统计出版社 1987 年版。

[3] 陈希孺。倪国熙编著：《数理统计学教程》，上海科学技术出版社 1988 年版。

[4] 张尧庭编：《概率统计》，中央广播电视大学出版社 1987 年版。

[5] 李茂年、周兆麟主编：《数理统计学》，天津人民出版社 1983 年版。

[6] 汪荣鑫编：《数理统计》，西安交通大学出版社 1987 年版。

[7] 复旦大学编：《概率论，第二分册：数理统计》，高等教育出版社 1979 年版。

[8] 成平、陈希孺等：《参数估计》，上海科学技术出版社 1985 年版。

[9] 中山大学数学力学系：《概率论与数理统计》，高等教育出版社 1980 年版。

[10] 周复恭、汪叔夜、黄运成：《应用数理统计》，中央广播电视大学出版社 1987 年版。

[11] 薛天栋：《数量经济学》，华中工学院出版社 1986 年版。[12] 茆诗松、丁元、周纪芗、吕乃刚：《回归分析及其试验设计》，华东师范大学出版社 1981 年版。

[13] 项静恬、杜金观、史久恩：《动态数据处理——时间序列分析》，气象出版社 1986 年版。

[14] 社会经济统计学原理教科书编写组：《社会经济统计学原理教科书》，中国统计出版社 1984 年版。

[15] 袁寿庄主编：《社会经济统计学概要》，中国人民大学出版社 1987 年版。

[16] 陈锡康主编：《现代科学管理方法基础》，科学出版社 1989 年版。

[17] 杨纪珂、孙长鸣编著：《生产中的数理统计》，科学出版社 1986 年版。

[18] 周复恭、倪加勋、朱汉江、汪叔夜、黄运成：《应用数理统计学》，中国人民大学出版社 1989 年版。

[19] 戴世光主编：《世界经济统计概论》，人民出版社 1987 年版。

[20] 闵庆全：《简明国民经济核算体系》，经济科学出版社 1987 年版。

[21] 易丹辉编著：《统计预测——方法与应用》，中国人民大学出版社 1990 年版。

[22] 袁卫著：《统计推断思想》，中国统计出版社 1990 年版。

[23] [日]田口玄一著：《质量工程学概论》，中国对外翻译出版公司 1985 年版。

[24] [美]W.G.科克伦著：《抽样技术》，中国统计出版社 1987 年版。

[25] [日]山根太郎著：《统计学》，福建人民出版社 1983 年版。

[26] [美]梅森著：《工商业和经济学中用的统计方法》，中国人民大学出版社 1984 年版。

[27] [澳]P. H. 卡梅尔、M. 波拉赛克著：《应用经济统计学》，中国统计出版社 1988 年版。

[28] [加]J.Neter，W. Wasserman，G. A. whitmore:《Applied Statistics》，Allyn and Bacon, Inc.,1982。

[29] [美]J.W. Tukey：《ExploratoryDataAnalysis》，MA：Addison——Wesley, 1977。

[30] [美]M.H.DeGroot：《ProbabilityandStatistis》，Addison-Wesley Publishing Compa-ny, Inc. ，1975。

[31] [美]P.J.BickelandK. A. Doksum：《Mathematical Statistics－BasicIdeasand SelectedTopics》，Holden-Day, Inc. ，1977。

[32] [美]S.Makridakis,S.C,Wheelwzight,V. McGee :《Forecasting：methodsandapplications》，JohnWileyandSons, 1983。

[33] [美]JeanDickinsonGibbons：《Nonparametric Methods for Quantitative Analysis》，Hoet,RinehartandWinston, 1976。

[34] [美]W.W.Danlel :《AppliedNonparametricStatistics》,Houghton Mifflin Company (1978) 。

[35] [美]R.E.Walpole andR.H.Myers：《Probability and Statistics for Engineers and Scientists》，Macmillan Publishing co.，Inc.1978。

[36] [美]R.I.Scheaffer, W. Mendenhall and L.ott：《Elementary Survey Sampling》，Duxbury Press 1986。