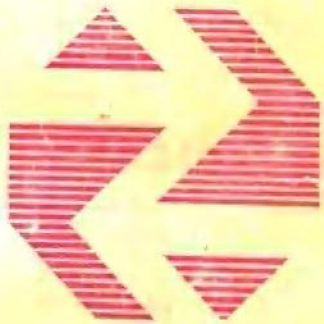
陈希孺　倪国熙　编著

数理统计学教程

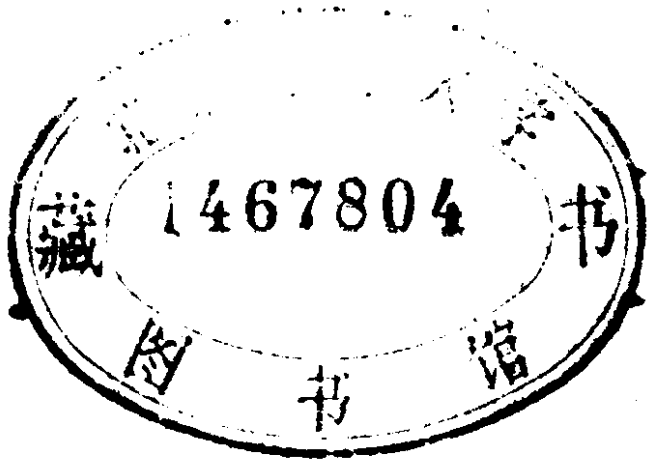


数理统计学是数学的一个分支，它的任务是研究怎样用有效的方法去收集和使用带随机性影响的数据。本书是数理统计学的基础教程，内容包括基本概念　点估计　假设检验、区间估计，Bayes统计与统计判决理论、线性统计模型和多元分析基础等。

上海科学技术出版社

数理统计学教程

陈希孺　倪国熙　编著



上海科学技术出版社

内 容 提 要

本书是数理统计学的基础教程，内容包括基本概念、点估计、假设检验、区间估计、Bayes统计与统计判决理论、线性统计模型和多元分析基础等，是为综合性大学及师范院校数学系以及数理统计专业大学生、研究生和教师进修班的数理统计基础课提供一种教材，也可供工科等非数学系作为此课程的教材或参考书，具备初等微积分、矩阵以及概率论基本知识的读者，均可使用本书。本书主要读者对象为理工科、经济、管理、师范院校等大学基础课师生及具有大学二年级数学程度的其他读者。

数理统计学教程

陈希孺　倪国熙　编著

上海科学技术出版社出版

（上海瑞金二路 450 号）

新华书店上海发行所发行　上海商务印刷厂印刷

开本 850×1156 1/32　印张 12　字数 318,000

1988 年 4 月第 1 版　1988 年 4 月第 1 次印刷

印数 1—7,000

ISBN 7-5323-0688-7/O·77

定价：3.60 元

序　言

本书是数理统计学的基础教程，要求读者具备初等微积分、矩阵以及概率论的基本知识，大体上相当于工科院校二年级学生的数学水平.

本书的主要目的，是为综合大学及师范院校数学系的数理统计课，以及高等院校数理统计专业大学生、研究生和研究生班、教师进修班的数理统计基础课，提供一种教材. 并可供非数学类学生作为此课程的参考书或课本. 本书内容是按一学期周四学时设计的，如果只有30学时左右的时间，可考虑选用第一章，§2.1，§3.2，§3.6，§4.1，§5.1的一部分，§6.1，§6.2与§6.3的一部分，§7.1与§7.4的一部分，以及其他各节适当一部分，未选入部分可作为学生的课外阅读材料. 本书也希望能对数理统计课教师在备课中有所助益. 对具备了前述数学基础的应用统计工作者，本书可作为自学读物.

数理统计基础课的内容，有基本原理和应用方法两方面. 由于有用的统计方法很多，本课程为学时所限，不能兼收并蓄. 本书选择了我们认为对各方面应用都较大的两个方面——线性模型和多元分析，各设专章. 另一些专题，如抽样方法，时间序列分析，可靠性统计等等，在有关的应用领域中也有其重要性. 这些内容可由任课者根据需要作适当补充. 在基本原理方面，第五章要特别交代一下. 数理统计学中的Bayes学派现已有很大影响，任何学习数理统计学的人，即使其主要兴趣在于应用，都有必要对此有所了解. 目前我国教本中对此似较少涉及，有时则将它作为一个纯方法性的内容去处理. 因此，本书花了一些篇幅去阐明这个学派的基本原理及所引起的一些争论问题. 统计判决函数已成为数理统计基础结构中的一个组成部分，在应用上也有其重要性. 本书着重阐明了它的基本观点，并注意不把它和统计推断混同起来. 另

外，本书中介绍了若干历史情况．作者认为，这些对于深刻地理解数理统计学中的一些问题，以及在培养一种用发展和批判的眼光看待这门学科的观点上，都可能是有益的．

由于本书花了较多篇幅在统计思想、观点和概念的阐述上，使书的篇幅略微加大了一些．我们认为这是值得和必要的．不论人们对数理统计学是否是数学的一部分这个问题持什么看法，都承认：把数理统计学，尤其是其基础部分，作为一门纯数学课去讲授是不可取的．一些同志的经验都表明，此课之所以难教难学，关键不在于数学推导上的困难，而在于初学者不易正确地把握住和深刻地理解有关的统计思想和概念．一旦这个问题处理好了，困难就会迎刃而解．另外，作者还有这样一个想法：数理统计基础课的目的，不应是纯技术性的，即教给学生一些现成使用的方法，还要起到培养学生树立用正确的统计观点去观察和研究事物的能力和习惯．要做到这一点，就必须在讲授中作出相应的努力．作者希望本教程对处理这个问题多少有一点帮助．由于要正确地理解和掌握统计的思想和观点，往往需要有一个反复的过程，因此在第一次阅读书中的一些理论文字时，如一时不得要领，不妨待学到一定程度后再回过头来仔细揣摩，自可豁然开通．

本书的写作与出版，是由于上海科学技术出版社徐福生同志的提议和鼓励，以及上海科学技术出版社的支持。在写作中，除了作者自己过去的讲义和笔记以外，参考了不少中外数理统计基础著作．另外，在多年来与统计界同志的交谈中，启发了作者思考一些问题．对以上有关同志和单位，谨在此表示衷心的感谢．由于作者才疏学浅，纰误之处，在所难免，希望同志们提出宝贵的批评和指教．

作者 1984 年 5 月于武汉

目　录

第一章 基本概念

§1.1 导 言

(一) 什么是数理统计学

关于数理统计学，现在已有了用各种文字出版的大量教科书和专著．在这些著作中，对数理统计学的性质、任务、应用等等，作了不少的论述．应该说，这些问题目前在统计学界并无原则性的分歧．但是，若试图用少量的文字对"数理统计学"这个学科下一个正式的定义，就会碰到不少困难．你很难找到一种说法是完全无懈可击的．况且，任何这样的定义，若不辅之以大量的解释，就无法使人理解．因此，在以下的叙述中，我们将致力于从一些方面把数理统计学的实质说清楚，而不着重于一个形式的定义．

当用观察和实验的方法去研究一个问题时，第一步就是通过观察或试验以收集必要的数据．这些数据受到偶然性即随机性因素的影响．下一步就是对所收集的数据进行分析，以对所研究的问题作出某种形式的结论．在这两个步骤中，都会碰到许多数学问题．为解决这些问题，发展了许多理论和方法．这些就构成数理统计学的内容．故一般地可以说，数理统计学是数学的一个分支，它的任务是研究怎样用有效的方法去收集和使用带随机性影响的数据．下面来作些解释．

1．数据必须带有随机性的影响，才能成为数理统计学的研究对象．例如，考虑一个国家的全面人口普查．假定人力物力时间允许我们对国内每一个人的状况进行调查，而这种调查又是准确无误的，则我们可利用普查所得数据，通过既定的方法，把所感兴趣的指标计算出来，例如，男性人口占全体人口的百分之多少，在所作假定之下这是准确无误的．这里不需要用到什么数理统计方

法. 又如要比较两个小麦品种甲、乙谁优(能有更高的产量). 若我们作一个不大现实的假定,即其他条件可以控制得如此严格(且这种条件也是日后大面积推广时所使用的),以致产量完全取决于品种,则我们只须在两块地上把甲、乙各种植一次,就可准确无误地判断其优劣. 在此数理统计方法也没有用武之地. 总之,是否假定数据有随机性,是区别数理统计方法和其他数据处理方法的根本点.

数据的随机性的来源有二: 一是问题中所涉及的研究对象为数很大,我们不可能对之全部加以研究,而只能用"一定的方式"(说详下)挑选其一部分去考察. 例如,一批产品有10,000件,其中含有废品 m 件,m 未知,因而废品率 $p=m/10000$ 也未知. 要确切地知道 p,必须对这10000件逐一加以检验. 这不仅是不经济的,且往往无法做到(如检验是破坏性的). 因此我们只能从其中挑出一部分,例如100件,根据对这100件的检验结果去估计 p. 在这里,随机性的影响就表现在:那100件被挑出是偶然的.

一般,在社会调查性质的问题中,问题的要求规定了调查的范围. 如问题是研究某一地区内以农户为单位的经济状况,则该地区的全体农户都是调查对象. 若这个数目太大,则我们只能挑一部分作实地调查. 这时,所得数据的随机性就来自被挑出的农户的随机性. 对这种数据作分析,就必须使用数理统计方法.

数据随机性的另一种来源是试验的随机误差,这是指那种在试验过程中未加控制、无法控制,甚至不了解的因素所引起的误差. 例如,设反应温度和压力是影响产品质量 Y 的重要因素,我们想通过一定的试验去考察这影响的程度,并挑选一个适当的温度和压力值以供在今后大批生产中使用. 但是,Y 除了与温度、压力有关外,还受到大量其他因素的影响. 例如,每次试验所用原材料略有差异,可能使用不同的仪器设备和操作者等等. 这些因素无法或不便加以完全的控制,而对试验结果(数据)产生随机性的影响. 这就带来一种不确定性. 例如,从试验数据上看,使用温度 t_2 比用 t_1 好. 但这个表现在数据上的优势究竟是本质的——即

有足够的理由可解释为是由于 t_2 确优于 t_1，还是只是随机误差的偶然性表现？这就需要用数理统计的方法去分析.

2. 所谓"用有效的方式收集数据"一语中，有效一词该如何解释. 归纳起来有两个方面：一是可以建立一个在数学上可以处理并尽可能简单方便的模型来描述所得数据，一是数据中要包含尽可能多的、与所研究的问题有关的信息.

例如，在考察某地区共 10,000 农户的经济状况的问题中，我们前面说挑出 100 户作实际调查. 100 这个数字是否恰当？太大了则费用过大，太小了则代表性不够. 要决定一个较好的数字，须权衡这两个方面，并用得着统计方法. 其次，假定我们选择了 100 这个数字. 这 100 户如何挑选？假设你只在该地区最富裕的那部分去挑，这样得到的数据就没有代表性，也谈不上有效了. 反之，你如果用一种纯随机化的方法，即设法使这 10000 户中的每一户有同等的机会被挑出，则所得数据就有一定的代表性，我们也不难建立一个简单的模型来描述它. 在一些情况下，我们还可以设计出更有效的方法. 举一个简单情况. 若该地区分成平原和山区两部分，前者较富裕且占全体农户的 70%，则我们可规定，在预定要考察的 100 户中，有 70 户从平原地区挑，30 户从山区挑，而在各自的范围内则用纯随机化的方式挑. 直观上我们觉得，这样得到的数据，比在全体 10000 户中用随机化方式挑选得到的数据更有代表性，因而也更"有效". 数理统计的理论证明确是如此.

又如，在产品质量与反应温度和压力的关系的例中，怎样用有效的方式收集数据，问题更多. 若可以考虑的温度在 t_1 和 t_2 之间，压力在 p_1 和 p_2 之间. 首先，我们当然只能取有限个温度和压力值去做试验. 取多少个值好？这里也有与上例中一样的问题：太多了费用太大，太少了不说明问题. 在定下了一个数目，例如四个温度值和四个压力值去做试验，则这些值是否均匀地取在相应的区间中好？另外，若把这些值所有可能的搭配都做试验，则至少需做 16 次. 也许条件不允许做这么多，而只能做一部分，则这一部分如何挑选？这些问题解决得好，试验数据就有一种平衡或对称的

结构，不仅更富于代表性，且可建立一种简单而便于分析的模型。

用有效的方式收集数据的问题的研究，构成了数理统计学中的两个分支，其一叫**抽样理论**，其二叫**试验设计**，它们分别处理相当于上面讨论过的两个例子中的那种类型的数据收集问题。

3．现在来解释“有效地使用数据”一语的意义。获取数据的目的，是提供与所研究的问题有关的信息。但这种信息并非是一目了然地表现出来，而需要用“有效”的方式去集中、提取，进而利用之以对所研究的问题作出一定的结论。这种“结论”，在统计上叫做“推断”。在§1.3中我们将仔细解释统计推断的意义，这里只指出：所作的推断应是对所提出的问题的一个回答，而不只限于所得数据的范围内。有效地使用数据，就是要使用有效的方法，去集中和提取试验数据中的有关信息，以对所研究的问题作出尽可能精确和可靠的推断。其所以只能做到“尽可能”而非绝对地精确和可靠，是因为数据受到随机性因素的影响。这种影响可以通过统计方法去估计或缩小其干扰作用，但不可能完全消除。

为有效地使用数据以进行统计推断，涉及很多的数学问题。需要建立一定的数学模型，并给定某些准则，才有可能去评价和比较种种统计推断方法的优劣。例如，为估计一物体的重量 a，把它在天平上秤九次，得到数据 $x_1, \cdots, x_9$，它们都受到随机性因素的影响（影响大小反映天平的精密度）。我们可以用这九个值的算术平均 $\bar{x}=\frac{1}{9}(x_1+\cdots+x_9)$ 去估计 a，也可以考虑下述方法：把 $x_1, \cdots, x_9$ 按大小依次排列成 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(9)}$，而取正中间的一个，即 $x_{(5)}$，去估计 a。甚至也可以用两个极端值的平均，即 $w=\frac{1}{2}(x_{(1)}+x_{(9)})$。你可能在直观上会认为：作为 a 的估计，$\bar{x}$ 优于 $x_{(5)}$，而 $x_{(5)}$ 又优于 w。但是为什么？这是不是对？在什么意义下对？在什么条件下对？这些问题就不容易回答。事实上，对这些问题的研究，正是数理统计学的中心内容，要使用大量的数学和概率论的工具。以后我们将看到：在一定的情况（取决于随机性影响的概率结构，即统计模型）和一定的意义（即衡量优越性的指标）之下，上述三个估

计方法中的任一个都可能成为最优的.

4. 最后一点,就是数理统计学只处理在收集和使用带随机性影响的数据中的数学问题,因而是一个数学分支.

一个问题的研究,涉及到问题所在领域的专门知识. 数理统计学不以任何一种专门领域为研究对象:不论你的问题是物理学的、化学的、生物学的或工程技术方面的,只要在安排试验和处理试验数据中涉及到一些一般性的、共同的数学问题,就可以用统计方法. 例如,不论作那种试验,都有一个试验规模的问题,即试验须重复多少次,才能把随机误差的影响控制在必要的限度内. 这是一个与专业知识无关的带共性的问题,一组试验数据,只要对其所受的随机性影响作了明确的规定(如服从正态分布),则可以用相应的统计方法去分析之,而不管这些数据的实际含义如何. 这种带共性的问题既然从专门的知识领域中超脱出来,就可以用纯数学的方法去研究,这就是数理统计学的对象. 我们这样说,并不意味着一个数理统计学者可以不过问其他专门领域的知识.相反,如果他要将统计方法用于实际问题,他必须对所论问题的专门知识有一定的了解. 这不仅可以帮助他选定恰当的统计模型和统计方法,而且,用数理统计方法分析随机性数据所得结论的恰当解释,离不开所论问题的专门知识. 例如,数理统计方法对数量遗传学很有用,但一个对遗传学一无所知的统计学家,就难于在这个领域中有所作为.

统计方法的应用很广泛,所以许多学习其他专业的人,都需要一些这方面的知识. 幸好,统计方法的具体使用并不需要很高深的数学知识. 相反,这些方法的理论根据,不具备较多较深的数学知识就说不清楚. 因此,在一些统计方法得到广泛应用的国家,例如在美国,出版了大量专供各领域的应用者使用的著作. 这种著作介绍统计方法及其应用,但不涉及或很少涉及这些方法的理论根据. 这种著作被列入"统计方法"或"应用统计"的范畴内,而只有那种用严格的数学去论证统计方法的理论根据的书,才称为数理统计著作. 这显示在这些国家中,"数理统计学"一词是给以一

种狭义的解释，即只包括统计学中的数学基础部分．在我国，数理统计学一词则是与作为一门社会科学的统计学相对而言的．粗略地说，在我国，数理统计学与西方的统计学相当，而具有较广泛的含义．明白这个差别就可以避免一些误解．

（二）数理统计学的应用

数理统计方法的应用很广泛，几乎在人类活动的一切领域中都能程度不同地找到它的应用．这是因为，实验是科学研究的根本方法，而随机性因素对试验结果的影响是无所不在的．反过来，应用上的需要又是统计方法发展的动力．例如，现代数理统计的奠基人、英国著名学者 R. A. Fisher 和 K. Pearson 在本世纪初期大力从事这方面的研究，就是出于在生物学、数量遗传学、优生学和农业科学方面的需要．

在工农业生产中，一个常见的问题是：有一个（或多个）我们感兴趣的指标，如工业产品的某项质量指标，农业中的单位面积产量等．有一些因素对这个指标可能有影响，例如，工业生产中所用设备、原材料、配方和温度、压力及反应时间等工艺因素．农业生产中所用种子品种，肥料类型和施放数量，以及耕作方法等方面的因素．为了得到最大的经济效益，需要了解这些因素对所感兴趣的指标起影响的具体情况：那些因素是主要的，其影响有多大，因素与指标之间是否可建立某种数量上的联系等等．弄清楚了这些问题，就可确定一组较好的生产条件．这些都要通过做试验，就是把有关因素固定在某些水平上做试验，去观察感兴趣的指标值．试验要经过精心的设计，所得试验结果必然受到大量随机因素的影响，而必须用统计方法分析．因此，随着近几十年来工农业生产的规模愈来愈大，数理统计学在这方面的应用也与日俱增．在历史上说，试验设计的基本思想，以及分析试验数据的一种极重要的方法——方差分析方法，就是 R. A. Fisher 等在 1923～1926 年期间，在进行田间试验中开始发展起来的．Fisher 提出的思想和方法，在四十年代以来经过发展，日益广泛地用于工业生产中．目前

最常用的正交设计，就是一个有代表性的例子。

数理统计方法应用于工业的另一个重要方面是：现代工业生产多有大批量及要求很高的可靠度的特点。需要在连续生产的过程中进行工序控制，成批的产品在交付使用前要进行验收，这种验收一般不能是全面检验，而只能是抽样验收。需要根据数理统计学的原理，去制定适合种种要求的抽样验收方案。还有，一个大型设备往往包含成千上万个元件。由于元件数目很大，它们的寿命可视为随机的而服从一定的概率分布规律。整个设备的可靠性，与设备的结构及这种分布规律有关，因而可以用统计方法去估计之。为解决上面这些问题，发展了一系列的统计方法，目前常提到的"统计质量管理"，就是由这些方法构成的。

统计方法在医药卫生中有广泛的应用。例如，治疗一种疾病的种种药物和治疗方法的效果，常引用统计资料来说明。这种材料的可信性，依赖于其数据的取得方法与使用的统计方法。其他，如分析某种疾病的发生是否与特定因素有关（一个著名的例子是吸烟与患癌症的关系），关系大小，在污染大气的许多有害成分中，那些成分对人体有何种程度的影响，这类问题常用数理统计方法去研究，取得了不少有用的成果。

数理统计方法在气象预报、地震和地质探矿等方面有一些应用。在这类领域中，人们对事物的规律性认识尚不充分，使用统计分析方法可能有助于获得一些对潜在的规律性的认识，而用以指导人们的行动。不过，在人们对事物的规律性认识很不充分的情况下，一些起比较大的作用的系统性因素，只好当作随机性因素来处理，这样，统计分析的精度或可靠性就较差。

自然科学的任务是揭示自然界的规律性。一般是先根据若干观察或试验资料提出某种初步理论或假说，然后再从种种途径通过实验去验证之。在这里统计方法起相当的作用。一个好的统计方法有助于提取观察和试验数据中带根本性的东西，因而有助于提出较正确的理论或假说。在有了一定的理论或假说后，统计方法可以指导学者如何去安排进一步的观察或试验，以使所得数据

更有助于判定理论或假说是否正确. 数理统计学也提供了一些理论上健全的方法,以估量观察或试验数据与理论的符合程度如何。一个著名的例子是遗传学中的 Mendal 定律. 这个根据观察资料提出的定律, 经历了严格的统计检验. 数量遗传学的基本定律——Hardy-Weinberg 平衡定律,也是属于这种性质.

数理统计方法在社会、经济领域中也有很多应用. 在某些国家中,统计方法在这些方面的应用,比其在自然科学和技术领域中的应用更为显著. 统计方法在社会领域中的一项重要应用是抽样调查. 经验证明,经过精心设计和组织的抽样调查,其效果可以达到以至超过全面调查的水平. 另外, 对社会现象的研究有向定量化发展的趋势, 例如人口学, 确定一个合适的人口发展动态模型, 需要掌握大量的观察资料, 并使用包括统计方法在内的一些科学分析方法. 在经济科学中, 定量化的趋势比其他社会科学部门更早更深. 早在本世纪二、三十年代,时间序列的统计分析方法就用于市场预测. 现在,一系列的统计方法,从简单的到很艰深的,都在数量经济学和数理经济学中找到了应用.

(三) 简单历史

下面我们简单地介绍一下数理统计学这门学科的简单历史. 这必然是极为"粗线条"的,因为叙述一门学科的历史,离不开这门学科的具体内容. 另外,关于数理统计学的早期发展情况,学者们很少论述,可资征引的文献很少.

数理统计学是一门较年青的学科, 它主要的发展是从本世纪初开始. 在早期发展中, 起领导作用的是以 R. A. Fisher 和 K. Pearson 为首的英国学派. 特别是 Fisher, 在本学科的发展中起了独特的作用. 目前许多常用的统计方法以及教科书中的内容, 都与他的名字有关. 其他一些著名的学者, 如 W. S. Gosset (Student)、J. Neyman、E. S. Pearson (K. Pearson 的儿子)、A. Wald 以及我国的许宝騄教授等, 都作出了根本性的贡献. 他们的工作奠定了许多统计分支的基础, 提出了一系列有重要应用

价值的统计方法，和一系列的基本概念和重要理论问题．有一种意见认为，瑞典统计学家 H. Cramer 在 1946 年发表的著作«Mathematical Methods of Statistics»标志了这门学科达到成熟的地步．这样说也许并不过分，因为虽则在此以前已出现了一些重要的统计著作，特别是 R. A. Fisher 的«Experimental Design»和«Statistical Methods for Research Workers»．但第一次用严整的数学方法总结到那时为止数理统计学的主要成就的，还是要推 Cramer 的上述著作．

收集和记录种种数据的活动，在人类历史上很久远．翻开我国的二十四史，可以看到上面有很多关于钱粮人口及地震洪水等自然灾害的记录。在西方，Statistics(统计学)一词源出于 State(国家)，意指国家收集的国情资料．也有不少人为研究特定问题而进行观察试验，收集资料．但这些情况，终究还不能认为是数理统计这门学科已经成立的标志．因为有许多工作，只停留在收集数据或对之进行一些简单的加工整理．即使有时也作出了某种越出已有数据范围之外的推断，也只是基于一种朴素的直观想法，而未能把问题模型化使之带有普遍意义，更谈不上建立必要的基本概念和理论了．这种情况延续了许多年，这是因为，没有一定的数学工具特别是概率论的发展，无法建立现代意义下的数理统计学．也因为应用方面的要求还没有达到那么迫切，足以构成一股强大的推动力．到上世纪后半期直至本世纪初，情况才起了较大的变化．是否可以举出某一个时间或某一部著作，足以作为数理统计学正式诞生的标志？学者们提出过一些意见，但尚无定论．有的认为这个时间"不早于 1850 年"，有的将它定在 Fisher 诞生的那一年——1890 年．有的认为本世纪初 K. Pearson 关于 χ^2 统计量的极限分布的论文可以作为一个标志，也有人认为，直到 1922 年 Fisher 关于统计学的数学基础的那篇著名的论文发表，数理统计学才正式诞生．这个时间可能失之过晚，不过，Fisher 的这篇论文首次概括了统计理论的现状和存在的问题，并提出了数理统计学的三个任务．文中的观点的主要部分到现在仍基本有效．因之，

Fisher 这项工作无疑是数理统计学建立过程中的一个里程碑.

综合以上所述,我们可否试探性地下这样一个粗线条的结论:收集和整理乃至使用观测和试验数据的工作由来已久,这类活动对于数理统计这门学科的产生,可算是一个源头. 上世纪特别是上世纪后半期以来发展速度加快,且有了质的变化. 在上世纪末期到本世纪初期这一段,出现了一系列的重要工作. 无论如何,至迟到本世纪二十年代,这门学科已稳稳地站住了跟脚. 本世纪前四十多年有了迅速而全面的发展,到四十年代时,已形成为一个成熟的数学分支.

本世纪前四十多年是数理统计学辉煌发展的时期. 但战后以来这几十年,数理统计学的发展也很显著. 许多在战前开始成形的统计分支,在战后得到纵深的发展. 数学上的深度比以前大大加强了. 也出现了若干带根本性的新发展,如 Wald 的统计判决理论与 Bayes 学派的兴起(均见第五章),在数理统计的应用方面,也给人深刻的印象. 这不仅是由于战后工农业和科技等方面迅速发展所提出的要求,也由于电子计算机这一有力工具的出现. 许多统计方法的实施都涉及很大量的计算. 在电子计算机得到广泛应用以前,这些方法的威力就难于发挥. 例如在五十年代,当电子计算机还很稀少时,用电动计算机计算一个包含十余个自变量的线性回归,得用几十个人花成月的时间,现在在大型计算机上,所花时间则只以秒为单位计. 以此之故,有些在战前就已得到充分发展的统计方法,真正在应用上发挥作用还是在战后.

§1.2 样本和样本分布

在上节中,我们对"什么是数理统计学"这个问题,作了一番定性式的描述. 其所以是定性的,因为我们没有引进必要的数学概念及使用严密的数学语言. 在一个意义上说,本章其余这几节是继续上节的工作,对"什么是数理统计学"这个问题给以更细致而严密的回答. 在这个过程中,也就自然地引进了一些重要的基本