

应用数理统计学

周复恭 倪加勋 朱汉江 汪叔夜 黄运成编著

0213
8

应用数理统计学

周复恭 倪加勋 朱汉江 编著
汪叔夜 黄运成

中国人民大学出版社

应用数理统计学

周复恭 - 倪加勋等、编著

中国人民大学出版社出版发行

(北京西郊海淀路39号)

中国人民大学出版社印刷厂印刷

(北京鼓楼西大街新胡同61号)

新华书店经销

开本: 850×1168毫米 32开 印张: 21.25

1989年9月第1版 1989年9月第1次印刷

字数: 524 000 册数: 1—3 000

ISBN7-300-00552-7

C·34 定价: 7.15元

编写说明

本书是大专院校统计专业本科学生的应用数理统计学教材。它是在给学生讲授的讲稿基础上形成的。在编写过程中注意了以下的指导思想。

1. 正如本书的书名所示, 为应用数理统计学, 因此内容侧重于数理统计方法的介绍和应用, 而不是数学的证明。但考虑到有些读者想进一步了解有关公式的来源, 所以也作了一些必要的推导证明。

2. 在内容的安排上, 主要包括了数理统计学中一些常用的方法, 但是重点放在数理统计在社会经济中的应用。

3. 在叙述方式上, 力求把数理统计的思维方法告诉读者, 即从不肯定性或概率的角度来思考问题。本书尽量以具体的例子和直观、逻辑的方法来阐明数理统计的原理, 代替单纯的数学推导, 并试图以辩证唯物主义的思想来解释这些原理。

4. 教材尽可能适合多方面不同层次读者的需要, 既通俗, 叙述比较详细, 不使数学基础较差的读者看见一大堆数学公式望而生畏, 又照顾到大学本科的适当深度。对缺乏微积分和矩阵代数知识的读者可以把数学证明和有关章节略去而直接使用结论。故本书也可用于统计干部的培训及自学。其他专业的师生也可作为参考资料。

5. 本书吸取了目前西方应用统计学的一些内容和例子, 但尽量注意到洋为中用。

但由于我们的水平有限, 书中难免有不少错误, 希望读者批

评指正。有若干例题选自国内已出版书籍，不能一一列举，在这里一并表示感谢。

参加本书编写的有周复恭、倪加勋、朱汉江、黄运成和汪叔夜。第一、二章由周复恭同志执笔，第三、四、十三、十六、十七章由倪加勋同志执笔，第五、六、八、十四章由朱汉江同志执笔，第七、十五章由汪叔夜同志执笔，第九、十、十一、十二章由黄运成同志执笔。全书最后由倪加勋同志负责统一定稿。

中国统计出版社项威同志和人民出版社苏一针同志对本书提出了许多宝贵意见，付出了辛勤的劳动，特致谢忱。

目 录

第一章 绪论	1
§1.1 什么是数理统计学	1
§1.2 概率论与数理统计学	4
§1.3 数理统计学的分类	5
§1.4 数理统计学的发展简史	6
第二章 描述统计	10
§2.1 统计数列和次数分配	11
§2.2 集中趋势的计量	22
§2.3 差异的计量	32
§2.4 偏态和峰度的计量	39
§2.5 探索性资料分析	41
§2.6 指数	45
第三章 概率	54
§3.1 随机事件与样本空间	54
§3.2 概率的定义及其计算	58
§3.3 概率的运算规则	62
§3.4 全概率公式与贝叶斯公式	68
第四章 随机变量	73
§4.1 随机变量的概念	73
§4.2 概率分布及其作用	75
§4.3 离散的概率分布	76
§4.4 连续的概率分布	78

§4.5	联合概率分布	83
§4.6	数学期望	92
§4.7	期望值的一些定理	96
§4.8	矩	99
§4.9	方差及其性质	103
§4.10	中位数、分位数和众数	105
§4.11	随机变量的一些其他特征	111
§4.12	几个重要的不等式	113
第五章	某些离散随机变量的概率分布	117
§5.1	贝努里试验和二项式概率分布	117
§5.2	超几何分布	127
§5.3	普哇松分布	133
§5.4	几何分布	140
§5.5	负二项式分布	144
§5.6	离散型分布的小结	147
第六章	某些连续随机变量的概率分布	149
§6.1	正态分布	150
§6.2	均匀分布	164
§6.3	指数分布	166
§6.4	伽马 (Γ) 分布	170
§6.5	对数正态分布	174
§6.6	威布尔分布	178
§6.7	连续型概率分布的小结	181
第七章	抽样分布	183
§7.1	归纳推断	183
§7.2	总体和样本	184
§7.3	参数、统计量与抽样分布	188
§7.4	变量的变换	191

§7.5	矩母函数	200
§7.6	中心极限定理	205
§7.7	一些常用的抽样分布	212
§7.8	从正态分布导出的几个重要分布	223
§7.9	其他一些样本特征的分 布	234
第八章	参数估计	238
§8.1	参数估计的基本原理	239
§8.2	总体平均数的估计	252
§8.3	两个平均数之差的估计	257
§8.4	总体比例的估计	263
§8.5	样本容量的确定	269
§8.6	总体标准差及方差的估计	274
§8.7	两个正态总体方差之比的估计	278
第九章	参数的假设检验	281
§9.1	假设检验的基本原理	281
§9.2	总体平均数的检验	288
§9.3	两个总体平均数的检验	292
§9.4	总体比例的假设检验	298
§9.5	总体方差的假设检验	303
§9.6	假设检验中的两类错误	309
第十章	方差分析	317
§10.1	方差分析的基本原理	317
§10.2	单因素方差分析	323
§10.3	两因素方差分析	335
§10.4	两因素有交互作用的方差分析	344
第十一章	简单回归分析和相关分析	354
§11.1	引言	354
§11.2	简单线性回归模型及其假设	356

§11.3	简单线性回归方程的估计	360
§11.4	线性回归的显著性检验	367
§11.5	回归分析的预测推断	372
§11.6	相关模型	378
§11.7	回归与相关分析中应当注意的问题	389
第十二章	复回归和复相关分析	392
§12.1	复线性回归模型及其假设	392
§12.2	线性复回归方程的估计	393
§12.3	复回归方程的评价	399
§12.4	复回归方程的预计推断	406
§12.5	复相关模型的假设及相关系数的计算	408
§12.6	复回归分析中的有关问题	415
§12.7	变量间非直线关系的回归	422
第十三章	非参数统计方法	430
§13.1	χ^2 分布的应用	430
§13.2	哥尔莫哥洛夫-斯米尔诺夫检验	441
§13.3	威尔科克森 W 检验及曼-惠特尼 U 检验	446
§13.4	成对比较检验	453
§13.5	等级相关	459
§13.6	克罗斯考尔-瓦里斯检验	462
第十四章	时间序列分析及预测	468
§14.1	古典时间序列模型	468
§14.2	长期趋势分析	470
§14.3	季节变动分析	490
§14.4	指数平滑(修匀)法	496
第十五章	抽样调查	501
§15.1	抽样调查的一些基本问题	501
§15.2	纯随机抽样	509

§15.3	分类(类型)抽样	519
§15.4	整群抽样	531
§15.5	等距抽样	542
§15.6	多阶段抽样	547
第十六章	统计质量管理	557
§16.1	统计质量管理的内容	557
§16.2	工序控制	558
§16.3	产品的抽样检验	579
第十七章	统计决策	600
§17.1	决策的含义及其种类	600
§17.2	风险型决策	603
§17.3	连续变量的风险型决策	613
§17.4	风险型决策中的其他标准	616
§17.5	不确定情况下的决策	622
§17.6	决策树	627
§17.7	贝叶斯决策方法	632
附录		
附录1	正态分布的面积和纵坐标表	645
附录2	卡方(χ^2)分布表	647
附录3	t 的临界值	648
附录4	威尔科克森 T 值	649
附录5	二项分布	650
附录6	普哇松分布	656
附录7	控制图的系数	658
附录8	曼·惠特尼检验, U 的临界值	659
附录9	斯皮尔曼等级相关系数: ρ 的临界值	660
附录10	F 分布: 显著水准, 5%和10%的临界值	661
附录11	哥尔莫诺夫、斯米尔诺夫两个样本检验中 D 的临界值(大样本, 双尾检验)	663

附录12	随机数字表	664
附录13	多重比较中的 q 表	665
附录14	检验相关系数 $\rho = 0$ 的临界值(r_α)表	667
附录15	r 与 z 的换算表	668
附录16	普哇松 (Poisson) 分布参数 μ 的置信区间表	669

第一章 绪 论

随着科学的发展，不论是在日常生活中，还是在科学研究中，对客观现象进行计量及定量分析显得越来越重要。统计方法作为定量分析的一种重要工具已得到广泛应用，成为认识客观世界的一种重要手段。

统计是一种古老的计量的手段，古代国家为了征兵、征税，就已经开始进行人口和土地等统计活动。资本主义的发展和生产社会化，对统计提出了新的要求，促进了统计工作的发展，并使统计成为一门独立的学科。特别是自17世纪以来，以概率论为基础的数理统计学发展很快，它一开始应用于生物和田间试验，以后又发展到工程、医学等其他领域。数理统计在社会经济领域也得到广泛应用。它已成为企业管理人员进行决策的重要工具。尤其是近代科学技术的飞速发展和电子计算技术及现代通讯传输设备的广泛应用，数理统计在现代化管理中的作用变得愈来愈重要，应用的范围也愈来愈广泛。

§ 1.1 什么是数理统计学

数理统计学是研究大量随机现象数量规律性的一门科学。所谓随机现象，从广义来说也称为不确定现象或偶然现象。确切地说，随机现象是指在基本条件不变的情况下进行试验或观察能得到不同结果的那种现象。典型的随机现象常以掷一枚硬币，可能出现正面，也可能出现反面，和掷一枚均匀的骰子，可能出现一

点、二点、……、六点为例。其结果在试验或观察之前是未知的，因而是不可确定的。

在日常生活中，随机现象也是普遍存在的：

如：在同一个公共汽车站，候车的人数有时很多，有时却很少；

同一种农作物，在同样管理的条件下，产量不同，有的高，有的低；

同一个工人在同一台机器操作，其产品质量有好有坏；

新生的婴儿有男有女；

某一地点、在同一季节的气温有时高，有时低；

在同一市场上，某种蔬菜的价格有高有低等等。

从所举的例子看，虽然它们属于不同的领域，彼此之间各不相同，但却有一个共同的特点。这就是：在个别试验或观察中，其结果是不同的，似乎没有什么规律性。但通过大量观察，其结果就会呈现出一定的规律性。例如，新生婴儿的性别比例，就个别家庭观察，参差不齐；但如果进行大量观察，就会发现男女的比例接近 1 : 1，男婴略多于女婴。又如汽车站的候车人数，若对其进行长期的观察，也会发现一定的规律性。如这个车站每天大致有多少人上车，其流量会在一定的水平上下波动；从每天不同的时间看也有规律，即什么时间是高峰等等。汽车公司正是根据这些规律来调度车辆的。这种从大量观察中得到的规律性，在数理统计上称为统计规律性。所以数理统计学正是研究如何发现客观事物统计规律性的方法论科学。

为什么通过大量观察可以发现事物的规律性呢？这可以从辩证唯物主义的偶然性和必然性关系的原理得到解释。因为辩证唯物主义认为，客观世界中既不存在纯粹的必然性，也不存在纯粹的偶然性，任何事物的发生都是偶然和必然的对立统一。所谓必然性是客观事物发展过程中一定发生的、不可避免的趋势，它是由事物内部的根本原因引起的，它反映事物的本质联系。而偶然

性是指客观事物发展过程中可以出现也可以不出现，可以这样出现，也可以那样出现的现象，它是由事物外部或事物内部次要的、非根本性原因引起的。我们可以结合上面的例子作进一步分析。以公共汽车站候车人数这一现象来说，它决定于许许多多因素，我们可以把它大致分为两类。一类是主要因素，起着根本的、决定性的作用。例如，住在这个车站附近的居民而又在比较远的地方工作，需要每天乘车去上班；附近中学的走读生要乘公共汽车去上学；附近电影院散场后人们要在这里候车等等。这类原因对在这个车站候车的人数具有较大的影响，或者说与候车人数有着本质的或必然的联系。它决定了候车人数的总水平。这类原因在数理统计上称为系统性原因。另一类是次要的原因，虽然它对候车人数的多少也有一定的影响，但不经常发生，而且影响较小。例如，某人由于朋友来信约他今天去会面，因而平时不常乘车的人今天也来候车；有人生病了，原来天天要乘车去上班而今天不来了等等。这些原因对乘车的人数起着次要的、非本质的或偶然的作用，使候车的人数或多或少产生波动。这类原因在数理统计上称为偶然原因或随机原因。然而在现实世界中，主要原因和次要原因总是结合在一起共同发生作用的。其结果使候车的人数围绕一定的水平上下波动，使每一次观察到的结果各不相同。科学的任务就是研究某一特定领域的事物之间的联系和规律性，也就是要发现合乎规律的、必然的东西。然而那些非本质的联系往往影响了我们对客观事物本质的认识。因此要求我们透过现象去认识本质。恩格斯说过：“在表面上是偶然性在起作用的地方，这种偶然性始终是受内部的隐藏着的规律支配的，而问题只是在于发现这些规律。”^①数理统计就是从大量的个别偶然现

① 《马克思恩格斯全集》第21卷，人民出版社1965年9月第1版，第341页。

象中发现规律性的一种方法，但它是从数量的角度研究随机现象的规律性。根据大数定律，当我们研究的总体是由大量相互独立的随机因素构成的，当每个因素对总体只产生有限的影响时，这些个别影响会相互抵销，使总体具有稳定的性质。因此通过大量现象的观察，这些由非本质的个别因素形成的差异将相互抵销，从而呈现出客观事物间的本质联系来。由此可见，数理统计是科学研究的一种重要工具，也是认识客观世界的一种重要手段。

在经济管理中，常常会遇到一系列的问题需要我们作出判断和决策。例如，一批产品是否合格，工厂改变了工艺以后，产品质量是否提高了等等。如果从个别的产品看，往往具有随机的性质。通过大量观察就可以作出判断。又如全国当年粮食产量有多少？较上年增产还是减产？若只调查个别地块就会受偶然因素的影响，如果应用数理统计方法，抽取足够大量的数据，就能作出较正确的估计。在这个意义上，数理统计方法又是经济管理的一种重要手段。

§ 1.2 概率论与数理统计学

大家知道概率论也是研究随机现象规律性的一门科学。那么概率论与数理统计学又有什么不同呢？从研究对象来说两者是一致的，很难截然划分。但是它们的研究方法和侧重点有所不同。一般说来，概率论侧重于数学理论的研究。它对一些典型的随机现象的规律性用数学模型加以描述，称为概率模型。概率论对这些理想化的概率模型进行研究，找出其内在的性质和联系。而数理统计学则是以概率论为基础，创造一套如何对客观现象进行搜集资料、整理分析以及推断的方法。其次，概率论用的是演绎法，从一般到具体。如它通过对一些典型概率模型的研究，应用于客观实际，从而对一些具体的客观随机现象可能出现的结果作出估计。而数理统计用的是归纳法，它通过对客观现象部分资料

的观察、搜集和整理分析，根据样本推断总体，从具体到一般。二者的关系是：数理统计学以概率论为理论基础，而概率论则需要数理统计在实践中搜集资料加以验证，二者互为补充，所以数理统计也称为概率统计。

前面我们讲到统计规律性要通过大量观察才能被发现。但是在客观上有时只允许我们对随机现象进行有限次的观察，因此推断总体就会产生误差。我们需要对误差作出估计，否则这种推断也就没有意义。运用概率论的原理，根据部分与整体之间的内在联系进行分析与推断，就可以使归纳推断所产生的不确定性得到度量和控制。

综上所述，数理统计学的定义可以简要归纳如下：数理统计学是研究大量随机现象数量规律性的一门科学。它通过一定的设计来搜集数据和进行整理分析，是以部分资料来推断总体的一种方法论。它以概率论为基础，使得归纳推断所产生的不确定性得到度量和控制。

§ 1.3 数理统计学的分类

数理统计学在其发展过程中形成了理论统计学和应用统计学两个方面。理论统计学研究统计方法的数学基础，如各种估计的原理，假设检验的原理以及一般的决策原理。这些原理可以看成是概率论公理的推广，它与概率论之间有着更为密切的联系。理论统计学家根据实践中提出的问题致力于发展和扩充统计学的理论，统计理论也是概率论的自然的引伸。应用统计学则是把统计方法应用于各个领域，如有生物统计学、医药统计学、工程统计学、管理统计学、工商业及经济统计学等等。它们之间既有共同的一些方法，也有各个领域一些特殊的问题。应用统计学的产生一方面在于统计方法应用的广泛性，当它应用于许多不同的领

域时提出了许多特殊的问题；另一方面也在于科学研究的发展，各种调查、观察和实验越来越复杂，统计工具也相应地复杂起来，以至于在许多领域中研究工作者想要熟悉所有有用的工具已不可能。况且由于工具更加专门化了，它们就缺乏通用性，为了适合某一特殊的调查（观察）和实验，往往就要改进工具。这就要求具备统计的理论和研究的客观现象本身这两方面的知识。无论是理论统计学或某一领域的科学都不能很好地完成这一任务。适应这种发展的需要，应用统计学应运而生，它已经发展成为一个新的领域。本书的内容主要是应用统计学，叙述一般的统计方法，但侧重于在社会经济领域中的应用。虽然为了使读者能应用好这些工具，懂得一些公式的来源，也作某些数学推导，但我们的主要目的在于说明如何应用。

数理统计学的另一种划分是将其分为描述统计学和推论统计学。顾名思义，描述统计学是研究如何搜集统计资料以及将搜集到的资料用图、表以及一些概括性的数字加以表示的方法，其中也包括一些曲线的拟合等。描述统计的目的是使统计资料更加简明扼要，便于理解掌握，这是统计学的基础部分。而推论统计则是根据样本的信息来推断总体，也称作归纳统计。正如我们前面提到的，在我们日常生活中只能作部分的和有限的观察，要达到对整体的认识就要依赖于推断，因而统计推论部分是数理统计学的核心。从统计学的发展看也是从描述到推论。本书的内容侧重于推论。从章节的安排看，第二章是描述统计，第三章至第七章是有关概率论的内容，为统计推论作准备，第八章以后是一些具体的统计方法，把描述和推论结合在一起，以推论为主体。

§ 1.4 数理统计学的发展简史

数理统计学的发展有两个来源：一是对客观社会进行调查、