

应用数理统计

周复恭 汪叔夜 黄运成 编



中央广播电视大学出版社

应用数理统计

周复恭 汪叔夜 黄运成 编

中央广播电视大学出版社

应用数理统计

周复恭 汪叔夜 黄运成 编

*

中央广播电视大学出版社出版

新华书店北京发行所发行

北京印刷一厂印装

*

开本787×1092 1/32 印张19.25 千字400

1987年6月第1版 1987年10月第1次印刷

印数 1—13000

定价3.35元

ISBN 7-304-00078-3/C·4

前 言

近年来广大社会科学研究人员和实际工作者在对现象进行定性研究时,强调了定量分析的重要意义。适应着这一客观需要,很多高等财经院校开设了作为进行定量分析重要工具的应用数理统计学课程。这无疑是可喜的现象。

这本教材是我们在教学实践基础上,为中央广播电视大学统计专业和经济类其他专业学生编写的,也可作为普通高等财经院校和广大统计工作者自学的教材。

在编写时,我们注意了以下几点:(1)强调数理统计方法的工具性和应用性。(2)考虑到现代统计学的科学发展,编写时在学时数的限制内,尽量保持教材在科学内容方面的完整性,把适用于社会科学研究方面重要的一般性数理统计方法,纳入本教材的体系内。(3)根据广播电视大学的教学要求和学生的特点,本教材编写时,减少了繁重的数学推导,力图深入浅出,从直观上和逻辑上讲解数理统计方法的内容,以便学生容易理解和掌握。(4)为帮助学生易于理解和掌握本教材的内容,教材中的例题多为有关社会经济现象方面的问题。各章的习题包括在另册出版的学习指导书中。

参加本教材编写的同志有周复恭、汪叔夜和黄运成。

第一章、第二章和第十一章由周复恭同志编写;第三、四、五、六、七章,第十二章和第十三章由汪叔夜同志编写;第八、九、十章和第十四章由黄运成同志编写。

由于编写时间的仓促和水平所限,难免有不妥与错误之处,诚恳地希望读者批评指正。

编 者

1987年3月

目 录

第一章 绪论	(1)
§ 1.1 统计学的研究对象和性质	(1)
§ 1.2 数理统计学的分类	(3)
§ 1.3 应用数理统计学的作用	(6)
§ 1.4 总体和样本	(8)
§ 1.5 统计误差	(10)
第二章 统计资料的搜集和描述	(11)
§ 2.1 统计资料的搜集——统计调查	(11)
§ 2.2 统计资料的整理和次数分配	(12)
§ 2.3 集中趋势的计量	(23)
§ 2.4 差异的计量	(33)
第三章 概率基础	(43)
§ 3.1 预备知识	(43)
§ 3.2 随机事件	(48)
§ 3.3 概率与概率运算	(51)
§ 3.4 条件概率与贝叶斯公式	(58)
第四章 随机变量及其分布	(66)
§ 4.1 离散型随机变量及其分布	(66)
§ 4.2 重要的离散型分布	(76)
§ 4.3 连续型随机变量及其分布	(88)
§ 4.4 重要的连续型分布	(91)
§ 4.5 二维随机变量及其分布	(103)
第五章 抽样分布	(114)
§ 5.1 简单随机样本与经验分布函数	(114)

§ 5.2	统计量与抽样分布	(119)
§ 5.3	样本均值及两个样本均值之差的抽样分布	(127)
§ 5.4	有关样本方差的抽样分布	(143)
第六章	参数估计	(154)
§ 6.1	估计量与估计方法	(154)
§ 6.2	评价估计量优劣的准则	(161)
§ 6.3	总体均值及两个总体均值之差的区间估计	(167)
§ 6.4	总体比例及两个总体比例之差的区间估计	(182)
§ 6.5	正态总体方差的区间估计	(186)
§ 6.6	确定必要样本容量	(189)
第七章	参数假设检验	(196)
§ 7.1	参数假设检验的概念和步骤	(196)
§ 7.2	总体均值及两个总体均值之差的假设检验	(204)
§ 7.3	总体比例及两个总体比例之差的假设检验	(216)
§ 7.4	正态总体方差及两个正态总体方差比的假设 检验	(218)
§ 7.5	β 错误与功效检验	(225)
第八章	方差分析	(238)
§ 8.1	引言	(238)
§ 8.2	方差分析的基本原理	(241)
§ 8.3	单因子方差分析	(250)
§ 8.4	两因子方差分析	(260)
第九章	简单回归与相关分析	(272)
§ 9.1	引言	(272)
§ 9.2	回归参数的估计	(276)
§ 9.3	线性关系的检验	(287)
§ 9.4	回归参数的检验	(305)
§ 9.5	回归预测问题	(310)
第十章	多元回归分析	(323)

§ 10.1	引言	(323)
§ 10.2	回归参数的估计	(325)
§ 10.3	回归分析中的假设检验	(328)
§ 10.4	回归预测问题	(338)
§ 10.5	复相关分析和偏相关分析	(349)
§ 10.6	一般线性回归模型简介	(354)
§ 10.7	回归变量的选择	(365)
§ 10.8	曲线回归问题	(371)
第十一章	时间数列分析和指数	(375)
§ 11.1	时间数列及其构成	(375)
§ 11.2	原时间数列的调整	(377)
§ 11.3	长期趋势	(378)
§ 11.4	季节变动	(389)
§ 11.5	循环变动	(394)
§ 11.6	时间数列分析的限制性	(398)
§ 11.7	指数的概念	(400)
§ 11.8	简单几何平均数指数	(402)
§ 11.9	综合指数	(404)
§ 11.10	个体指数的平均数指数	(411)
§ 11.11	指数的基期变换和拼接	(415)
第十二章	非参数假设检验	(420)
§ 12.1	引言	(420)
§ 12.2	单样本非参数假设检验	(423)
§ 12.3	双样本非参数假设检验	(432)
§ 12.4	k 个样本的非参数假设检验	(451)
§ 12.5	等级相关系数	(460)
第十三章	抽样技术	(468)
§ 13.1	引言	(468)
§ 13.2	简单随机抽样	(472)

§ 13.3	分层随机抽样	(479)
§ 13.4	整群抽样	(493)
§ 13.5	系统抽样	(500)
§ 13.6	多阶段抽样	(508)
第十四章	统计决策基础	(514)
§ 14.1	引言	(514)
§ 14.2	统计决策中的基本概念	(517)
§ 14.3	不确定型决策	(524)
§ 14.4	风险型决策	(534)
§ 14.5	效用理论	(547)
§ 14.6	贝叶斯决策	(558)
参考文献		(572)
附表 1	二项分布表	(574)
附表 2	泊松分布表	(580)
附表 3	标准正态分布表	(582)
附表 4	指数分布表	(584)
附表 5	χ^2 分布表	(585)
附表 6	t 分布表	(587)
附表 7	F 分布表	(588)
附表 8	德宾——沃特森检验表	(598)
附表 9	游程检验的临界值表	(600)
附表 10	威尔科克森配对符号等级检验中 T 的临界值表	(602)
附表 11	柯尔莫哥洛夫——斯米尔诺夫双样本检验中 K_D 的临界值表 (小样本)	(603)
附表 12	柯尔莫哥洛夫——斯米尔诺夫双样本检验中 D 的临界值表 (大样本, 双侧检验)	(604)
附表 13	曼——惠特尼 U 检验临界值表	(605)
附表 14	克鲁斯卡尔——沃利斯单向等级方差分析中 H 的相伴概率 表	(607)
附表 15	斯皮尔曼等级相关系数 r_s 的临界值表	(609)
附表 16	肯达尔一致性系数 S 的临界值表	(610)

第一章 绪 论

近年来,人们越来越认识到,在研究问题时,除进行定性分析外,还必须进行定量分析。而统计学,即数理统计学则是进行定量分析的重要工具。近年来,在高等院校的财经系科先后开设了《数理统计学在经济中的应用》(我们认为改称《应用统计学》或《应用数理统计学》更为合适)。作为高等院校财经系科的学生,读者在学习这门学科时,如同学习任何学科一样,为对它有个完整而正确的理解,首先要了解这门学科的研究对象、性质及其作用。作为绪论,本章主要是阐明这些问题的。

§ 1.1 统计学的研究对象和性质

要了解统计学,即数理统计学的研究对象和性质,先要了解随机现象。所谓随机现象指的是,在一定的条件下,可能出现也可能不出现,可能这样出现,也可能那样出现的一类现象。对这类现象说,在基本条件不变的情况下,进行试验或观察,会得到不同的结果,而这种结果在试验或观察之前是未知的,因而是不确定的。这类现象自然界中有,社会现象中也有,遍于自然和社会之中。

作为随机现象的例子有:

在同一个公共汽车站,候车的人数有时很多,非常拥挤,有时又很少;

同一种农作物,即使在相同的管理条件下,其产量也有高有低;

同一个工人在同一台机器上操作,由他生产的产品质量有时好,也有时差;

新生婴儿的性别,有的为男,有的为女;

就个别家庭的收入和支出看,有的多些,有的少些;

如此等等,此类例子不胜枚举。

此类现象虽属于不同的科学领域,彼此不同,甚至有极大的差别,但它们都具有这样的共同特点,这就是:就个别试验或观察的结果看是不同的,不确定的,乍看起来是杂乱无章的;但就个别结果的集体看,通过大量观察和综合平均,却呈现出一定的数量规则性。例如,新生婴儿的性比例,就个别家庭观察,参差不齐,但作为集体看,新生婴儿的性比例却为14:13,男婴略多于女婴,也有一定的数量规律性。这种规律性称为统计规律性。从科学研究领域的划分看,正是这一类随机现象,成为统计学,即数理统计学的研究对象,可以说,统计学,即数理统计学研究随机现象数量规律性,是应用数学的分支之一。

数理统计学是适应着人类社会实践的需要而产生和发展起来的。它是在属于纯数学的概率论的基础上发展起来的,而又以概率论作为自己的理论基础。数理统计学和概率论的关系是这样的密切,以致可以把前者视为后者的增广。就是说,数理统计学的完整结构从广义上说,可以认为是包括概率

论的。这样就容易理解,为什么人们常把二者合在一起,而称之为概率论与数理统计的原因。

§ 1.2 数理统计学的分类

统计学,即数理统计学从它产生起,经历了一个不断发展和不断完善的过程。在这个过程中,它形成了两种分类,即理论统计学或统计理论和应用统计学与描述统计学和推论统计学。下面对这两种分类,分别加以说明。

1. 理论统计学和应用统计学

统计学,即数理统计学在其不断发展和不断完善中,按照统计学的数学理论和统计方法,分成理论统计学和应用统计学。理论统计学阐明统计学的数学理论。它以我们所观察到的规则性的最本质的和最基本的特征作为我们的出发点,用简单而理想化的形式把它们表述成数学命题作为理论的基本公理。然后由这些公理,应用纯逻辑推理,不再求助于经验,就得到各个命题。这种在公理基础上构成的,在逻辑推理又没有矛盾的命题系统就组成数学理论。这种理论就成为理论统计学。理论统计学研究的是抽象的随机现象的数量规律性。从事理论统计学研究的统计学家总是致力于发展和扩展统计学的数学理论,为各个科学领域的研究工作者创造工具和方法。由于理论统计学的深入发展,统计方法的应用才有可能得到广泛的进展。

应用统计学侧重于统计方法的应用。它把理论统计学的研究成果作为统计方法,为各门科学的研究人员,在研究各自

领域内的具体对象时,提供有用而必要的研究方法或工具。事实上,应用统计学早已广泛应用于各门科学领域,如物理学、天文学、生物学、农学、医学以及社会学、经济学、心理学等等。可见它的应用范围是十分广泛的。统计方法用于处理各门科学领域内大量数据资料或统计资料,帮助人们研究各门科学领域的具体的数量规律性。上面举的新生婴儿的性比例为14:13,便是一个明显的例子。它反映了人口学所关心的问题的具体数量规律性。

关于应用统计学的定义,在其历史发展过程中有很多,因人而异。但在统计方法的应用这一点却是共同的。我们把应用统计学定义为:它是关于在不确定情况下进行决策时,如何搜集、整理、分析和解释统计资料的一套理论和方法。这套理论和方法既可用于研究自然,又可用于研究社会。因之应用统计学是一门通用的方法论性质的科学。

要指出的是,应用统计学应用于不同科学领域时,其具体的应用对象虽有不同,但其应用的原理却是相同的。在其应用于不同的对象时,所不同的主要是它所强调的特殊技术。例如,应用于工商业和经济学时,发展出指数和时间数列;应用于社会学和教育学时,发展出因素分析;而应用于生物学时,发展出实验设计和方差分析;如此等等。作为应用统计学,本教材是为中央广播电视大学统计及经济类其它专业的学生编写的,也可以作为普通高等院校财经系科的学生的教材。因之,本教材所举例子多系社会经济现象,这是自然的。

统计学在其发展过程中,虽然分为理论统计学和应用统计学,但它们之间的关系却是十分密切的。理论统计学的研

究成果,为应用统计学的应用创造工具,使之不断扩大其应用范围。应用统计学在应用中遇到理论上的难题时,又要求理论统计学予以解决。正是这种密切关系推动着统计学的不断发展和完善。

· 附带说明,上面用的是应用统计学这一名称。由于目前在我国统计学界存在“是一门还是两门统计学”的争论,为避免人们的误解,我们把它改称为应用数理统计学。这就是为什么本教材冠以应用数理统计学和在以后的阐述中采用这一名称的原因所在。

2. 描述统计学和推论统计学

统计学,即数理统计学从其应用目的看,分为描述统计学和推论统计学。前者阐述的是关于如何搜集资料以及如何简缩和描述这些资料的一套统计方法。资料简缩应该这样安排,使搜集到的统计资料所含信息尽可能多地被选定的一组描述性的特征数反映出来。后者阐述的是关于如何根据部分数据推论出更为一般情况的一套方法。用统计术语说,就是如何根据样本提供的信息,推断出样本来自的总体的相应情况的一套统计方法。这套统计方法可以使我们由个别到一般,获得新的知识,加深我们对客观事物的认识。

描述统计学和推论统计学这一分类,也是在统计学的发展过程中形成的。约在二十年代之前,它主要是描述统计学。但随着统计学的不断发展和完善,推论统计学越来越重要,成为统计学的核心部分,在内容上越来越占有较大的比重。即使如此,也没有影响描述统计学在统计学中的重要地位。可以说描述统计学是统计学的基础部分。没有这个基础部分,

推论统计也无法进行。因为没有对样本的描述，就无法根据抽样的结果对总体进行推论。所以，统计学正是由这两部分构成的。那种认为只有推论统计学才是统计学，即数理统计学的观点是不正确的。

本教材正是按照统计学，即数理统计学的基础部分和核心部分的顺序编排的。第一章是绪论，第二章属于描述统计。第六到第十章是推论统计。进行统计推论先要了解什么是概率，常用的概率分布有那些，样本的抽样分布又如何。只有了解了这些内容，才能理解推论统计的具体内容。所以，第三章到第五章分别讲了概率、概率分布和抽样分布。可以把它们看成是由描述统计进到推论统计的桥梁。时间数列分析和指数，是统计方法应用于研究经济现象时产生的特有方法。考虑到教材篇幅的限制和它们二者之间也存在某种联系，我们把它们合并成一章，即第十二章。下余的三章分别是非参数统计、抽样技术和统计决策基础。这是由于近几十年来，统计学有了很大的发展，非参数统计和统计决策是这种发展中的新成果。所以，应该列为本教材的内容。推论统计是由样本推论总体。但样本是怎样抽选出来，在以前各章中未予交待。所以，第十三章专门介绍样本的抽选方法和估计方法。这便是本教材的基本内容。它反映了作为科学的应用数理统计的最基本的内容。

§ 1.3 应用数理统计学的作用

理论统计学以抽象的形式研究随机现象的数量规律性。

但它的科研成果与客观存在的各门科学领域中的具体的或特殊的规律性有什么关系以及如何被应用于各门科学领域中的具体对象上去，它未予以回答也不予以回答。因为这不是它的科学任务。这个科学任务要由应用数理统计学来完成。

已经说明，应用数理统计学告诉我们如何把统计方法用于处理各门科学领域中的大量数据资料或统计资料。这些资料总是与各门科学领域中的具体现象密切联系而又不可分割的。作为大量数量差异资料或统计资料，它们是必然性和偶然性综合作用的结果。辩证唯物主义告诉我们，必然性和偶然性是对立统一的。就是说在客观世界中，没有脱离偶然性的纯粹的必然性。也不存在脱离必然性的纯粹的偶然性。

根据辩证唯物主义关于必然性和偶然性对立统一的观点，把统计方法应用于各门科学领域的具体对象时，通过大量观察和综合平均，就可以抵销掉那些由偶然性作用引起的数量差异，从而可以揭示出由必然性作用形成的数量特征。可以这样说，应用数理统计学可以用来帮助人们探索各门科学领域中的具体的或特殊的数量规律性。这就是在科学意义上所说的应用数理统计学或数理统计方法的作用。正是由于它具有这种作用，它就成为各门科学在研究各自领域内的具体的或特殊的数量规律性时不可缺少的重要工具。也正是由于它具有这种作用，它在发展过程中，应用范围不断扩大，形成为既可用于自然又可用于社会的通用的方法论性质的科学。

综上所述，可以说应用数理统计学是我们进行科学研究、科学试验和科学管理的必要工具。把统计方法应用于社会经

济问题的研究时,无论是进行科学研究,还是进行宏观控制或对微观经济作出生产经营决策,它都大有用武之地。问题不是在研究这些问题时,它能不能被应用,而是我们善不善于应用它。这个问题只有在不断的应用实践和总结中,才能得到妥善的解决。

善不善于应用应用数理统计学,除必须熟练掌握和恰当地选用它所提供的各种统计方法外,还必须具备与我们所研究的问题有关的理论知识、专业经验和科学的判断。要充分发挥应用数理统计学的作用,除熟练掌握和恰当选用各种统计方法外,具备与所研究的问题有关的理论知识、专业经验和科学判断是极其重要的。那种认为仅仅掌握了一些应用数理统计方法,而不问被研究问题的具体性质,就自然而然地可以揭示被研究问题的具体的或特殊的数量规律性的观点,显然是一种误解。

§ 1.4 总体和样本

总体和样本是应用数理统计中最常用的统计术语,为今后叙述的方便,先在此加以说明。

总体是指所有可能的个体的集合。而个体则是组成总体的基本单位,又叫总体单位。总体随研究的目的和性质的不同而不同。例如,研究某厂某天生产的日光灯管的使用寿命,则该厂该日生产的所有的日光灯管的使用寿命构成总体,而每支日光灯管的使用寿命则为个体,即总体单位。如果研究的是要了解该厂近三个月中每天生产的日光灯管的平均寿

命，那么该厂近三个月中所有每天生产的日光灯管的平均寿命，就构成了总体；所说的某日生产的日光灯管的平均寿命，便成为个体，即总体单位。根据举例的说明，人们常常定义总体为所有观察值的集合。

总体有有限总体和无限总体之分。就上例而言，该厂某日生产的日光灯管的使用寿命无论怎样多，总是有限个个体，由它构成的总体是有限总体。如果该厂的日光灯管一直连续生产下去，而没有具体时间的限制，这时由日光灯管使用寿命构成的总体就是无限总体。总体可以是实际的，也可以是假设的或概念上的。前者象所举日光灯管使用寿命，后者可以投掷一枚硬币为例，给出最简单的说明。在这一情况下，我们说我们的总体包括所有各次投掷，而这些投掷可以用一枚硬币在指定的情况下来进行。我们可以清楚地想象出这一总体，可以清楚地从这一总体得出各次投掷的结果。这样的总体就是假设的或概念上的总体。

与总体相对的是样本。由于各种原因，诸如总体无限，总体虽有限但无必要，带有破坏性的试验以及经济上节约的考虑等等，只能观察总体中一部分个体。由这一部分个体构成的集合，便称为样本。样本在推论统计学中具有特殊的意义。因为没有样本，就谈不上统计推论。关于总体和样本的问题，我们将在后面的章节中进一步讨论。

§ 1.5 统计误差

应用数理统计学总是与被研究现象的不确定性相联系