

Shuli Tongjixue Jiben Yuanli he Fangfa



数理统计学 基本原理和方法

吴梅村 编著



西南财经大学出版社
Southwestern University of Finance & Economics Press

0212
58

数理统计学 基本原理和方法

吴梅村 编著



西南财经大学出版社
Southwestern University of Finance & Economics Press

图书在版编目(CIP)数据

数理统计学基本原理和方法/吴梅村著. —成都:西南财经大学出版社, 2006. 1

ISBN 7-81088-301-1

I. 数... II. 吴... III. 数理统计 IV. 0212

中国版本图书馆 CIP 数据核字(2005)第 043435 号

数理统计学基本原理和方法

吴梅村 编著

责任印制:杨斌

责任编辑:李雪

封面设计:何东琳设计工作室

出版发行:	西南财经大学出版社(四川省成都市光华村街 55 号)
网 址:	http://press.swufe.edu.cn
电子邮件:	xcpress@mail.sc.cninfo.net
邮政编码:	610074
电 话:	028-87353785 87352368
印 刷:	成都科刊印务有限公司
成品尺寸:	148mm×210mm
印 张:	12.75
字 数:	320 千字
版 次:	2006 年 1 月第 1 版
印 次:	2006 年 1 月第 1 次印刷
印 数:	1—2000 册
书 号:	ISBN 7-81088-301-1/F·267
定 价:	19.80 元

1. 如有印刷、装订等差错,可向本社营销部调换。
2. 版权所有,翻印必究。
3. 本书封底无本社数码防伪标志,不得销售。

序 言

本书的出版是为了纪念作者吴梅村老教授。吴老师在完成《数理统计学基本原理和方法》一书的写作并即将修改完毕时因突发疾病,不幸逝世,时年 84 岁高龄。我们以出版吴教授遗作的方式来寄托我们的哀思,是因为老先生热爱党的教育事业,一生坚持“老老实实做人,认认真真教学”的原则。老先生以 70 高龄退休后,已是学术有成,桃李遍天下,本该安享晚年,但他仍以“经世济民,孜孜以求”的精神笔耕不止。除已于 1997 年公开出版的《相关关系分析的基本原理和方法》一书外,又开始了本书的写作和修改,不幸的是病魔夺去了令人尊敬的老先生的生命,致使本书的序言令读者不能亲见先生墨迹,只好令后生代序。

本书内容包括了数理统计学最基本的理论和方法,文字简明易懂,概念清楚明确,推证逻辑严谨,显示出先生深厚的统计理论底蕴。

本书分为三部分。第一部分是绪论(第一章)。该章论述了数理统计学的研究对象;阐述了数理统计学与概率论、数学和社会经济统计学的关系;说明了数理统计学在自然界、技术界和社会经济领域广泛的应用性。

第二部分是论述数理统计学的数理基础概率论理论(第二章至第六章)。该部分从概率论中的一些基本概念出发,论及概率分布。并着重介绍了二项式分布和泊松分布两种离散分布及最常用的连续分布——正态分布。该书还从正态分布出发对实践中实际观察资料存在的各种误差和各项指标,在大量观测下,如何趋近于理论总体的相应指标,详细而又严密地阐述和推论了极

限定理和中心极限定理,作为指导实践的应用。

第三部分是讲述数理统计的主要内容(第七章至第十二章)。这一部分首先介绍了抽样和推断的基本原理。其次,介绍了以样本推断总体的点估计、区间估计和假设检验等方法;较详细地论述了大样本分析和小样本分析方法;数学期望,方差的估计和假设检验的方法等内容。最后,在对随机现象的动态分析上,介绍了回归和相关分析方法和时间数列分析方法。在回归和相关分析方法中,介绍了单相关和复相关,一元回归和多元回归,直线回归和曲线回归等内容。在时间数列分析方法中,介绍了确定型时间数列的拟合方法;还引入了平稳随机过程的概念,深入论述了移动平均模型、自回归模型和自回归移动平均相结合模型的三种分析方法。

本书非常强调各种分布和方法的实际应用,特别是对工农业生产 and 科学试验中现象波动分析常用的方差分析,作者做了细微的论述。因此,此书可作为教学参考用书,也可作经济管理上进行预测决策分析的参考用书,还可作为工农业生产试验和科学试验的参考用书。

本书是老先生尚未完全修改完的遗作,难免可能存在一些不足,如果给老先生更多的时间,相信遗憾更少些。

西南财经大学统计学院

2005年12月

目 录

第一章 绪论	(1)
第一节 数理统计学的研究对象	(1)
第二节 数理统计学的性质	(2)
第三节 数理统计学与数学、概率论和社会经济统计学 的关系	(4)
第四节 数理统计学的基本内容和作用	(6)
第二章 随机事件和概率	(9)
第一节 随机事件	(9)
第二节 概率的定义和性质	(17)
第三节 概率的基本运算	(24)
第四节 全概率公式和逆概率公式	(32)
第三章 随机变量及其概率分布的一般性质	(36)
第一节 随机变量	(36)
第二节 随机变量的概率分布	(38)
第三节 概率分布的数字特征	(58)
第四节 矩和矩母函数	(74)
第四章 主要离散分布	(85)
第一节 理论分布的意义	(85)
第二节 二项分布	(86)

2 数理统计学基本原理和方法

第三节	泊松(poisson)分布	(97)
第四节	其他离散型分布	(105)
第五章	主要连续分布——正态分布	(112)
第一节	正态分布的意义和函数式	(112)
第二节	正态分布密度函数的推导	(114)
第三节	正态密度函数的性质	(118)
第四节	正态分布曲线的图示和概率积分	(120)
第五节	正态分布的期望值、方差、矩和矩母函数 ..	(132)
第六节	正态分布的应用	(139)
第六章	极限定理	(149)
第一节	基本概念	(149)
第二节	大数定理	(151)
第三节	中心极限定理	(158)
第七章	统计抽样和推断的基本原理	(169)
第一节	抽样和推断的基本概念和方法	(169)
第二节	点估计	(177)
第三节	区间估计	(188)
第四节	假设检验	(193)
第八章	大样本分析	(202)
第一节	矩母函数在大样本指标分布中的推断方法 ..	(202)
第二节	样本平均数的分布及其应用	(209)
第三节	样本成数的分布及其应用	(218)
第四节	样本方差和均方差的分布	(222)

第九章 小样本分析	(228)
第一节 Γ (Gamma)分布	(228)
第二节 χ^2 分布	(234)
第三节 t 分布及其应用	(247)
第四节 F 分布及其应用	(258)
第十章 方差分析	(267)
第一节 方差分析的基本概念和种类	(267)
第二节 单因素相等重复试验方差分析	(268)
第三节 单因素不等重复试验方差分析	(278)
第四节 双因素不重复试验方差分析	(285)
第五节 双因素有交错作用的方差分析	(290)
第六节 系统分组的方差分析	(300)
第七节 方差分析中的多重比较	(306)
第十一章 回归和相关分析	(310)
第一节 回归和相关分析的基本概念	(310)
第二节 简单线性回归分析	(312)
第三节 简单非线性回归分析	(340)
第四节 简单相关分析	(353)
第五节 多元线性回归分析	(363)
第六节 多元相关分析	(382)
参考书目	(397)

第一章 绪论

第一节 数理统计学的研究对象

数理统计学常与概率论相提并论,因为它们都是以随机(或偶然)现象的数量(或统计)规律性作为自己研究对象的科学。

随机现象普遍存在于现实世界中。无论是自然界、技术界还是社会经济领域,事物的发展变化,都表现为偶然现象和必然现象互相交织的复杂过程。辩证唯物论认为:事物的必然性和偶然性都是客观存在的,是对立统一的;必然性是在事物的内部联系中必然要产生的东西,是事物的本质;偶然性则与事物的本质没有直接联系,而是事物发展变化所表现出来的不确定性现象。总之,偶然性是必然性的表现形式和补充,必然性必须通过偶然性呈现出来,两者是密不可分的。要排除偶然性对必然性的掩盖和干扰,才有可能使人们认识客观世界,进而改造客观世界。

数理统计学不是研究事物本质的客观规律的科学,它是透过随机现象去研究事物的统计规律性和应用统计规律性的方法。所谓统计规律性,简而言之,就是通过对随机现象的大量观察,在数量呈现出来的有关事物的集体性规律,又称为大数恒静定律。事物的集体性规律不同于事物本质规律:事物的本质规律毫无例外地适用于事物包含的任何个体,而集体性规律则由事物的集体性质所决定,它与事物的任何单一个体的性质没有什么联系。但这两种不同性质的规律,在完成对事物的全面研究任务上,则是相

辅相成、互相促进、互为补充的。当某门科学,例如生物学、医学或社会经济学等,在研究其事物本质的规律时,进展迟缓或遇到一定困难不能取得显著成果时,若从事物的集体性规律方面进行研究,或可创造性的提出新的数理统计理论和方法,既可推进对事物本质研究的进展和新成果,也丰富和发展了数理统计学的内容,这在数理统计学的发展史上,是不乏先例的。

第二节 数理统计学的性质

数理统计学,顾名思义,既包含了数学的理论和方法,又包含了统计的理论和方法;但数学的理论和方法是贯穿在统计的理论和方法中的,它不是单纯的数学,重点在统计。数理统计学在研究和应用统计规律的发展过程中,产生了相应的理论和方法,如随机试验、抽样观察、分组(群)整理、综合指标,以及根据试验或观察资料数据推断更为广泛的情况等的理论和方法,它是统计特有的理论和方法,是数理统计学的重要内容。

数理统计学应用数学的理论和形式概括地表述随机现象的统计规律性,不同的数学形式作为模型来表述不同的统计规律性。数理统计学从事物的量的规律性方面进行定量分析,采用一定的数学模型,应该是没有疑义的。关键的问题是要选用恰当而准确的数学模型,即所选数学模型必须适合于所研究的事物的性质,才能产生所期望的效果,作出正确的结论。在数理统计学的发展史上,由于对这个问题的认识不足或错误,曾有过两种不同的偏向:一种偏向是不管事物的性质如何,滥用数学模型和数理统计方法,对观察资料进行数字游戏式的演算和分析,甚至把数学模型所表述的事物的数量规律性,错误地视为事物本质的规律;另一种偏向是,由于对数理统计学对事物非确定性测度的理解不

够,甚至认为它是资产阶级唯心主义的花招,将它摒弃于社会经济领域实用范围之外。我们在研究和应用数理统计学时,应吸取教训,端正认识,纠正这两种偏向。

数理统计学是方法论的科学。数理统计学虽然有自己的研究对象,但它的研究对象不是某一种物质世界的客观实体,而是各种自然、技术和社会经济领域普遍存在的不确定的随机现象的统计规律性。数理统计学对这些普遍存在的随机现象,提供了一系列观察、测度、估计、推断、预测和抉择方法,这些方法广泛应用于各种科学技术领域的研究和实践。这些数理统计方法,不是孤立的、简单方法堆集,而是有严格的数学理论和逻辑贯穿其间,以数学理论为基础的系统科学方法。由于科学技术和生产实践的不断进步和发展,数理统计学在由数学理论到广泛深入具体应用的方法之间,逐渐形成了完整的原理性质的体系,因而形成了独立的方法论的科学。

数理统计学一般分为描述统计和推断统计两部分。前者研究事物数据资料的搜集、整理和指标计算等方法,用以尽可能的全面准确地表述所研究事物的实际情况。我国自新中国成立以来所流通的社会经济统计学,基本上都是单纯的描述统计,是整个统计科学的基础部分。推断统计虽然离不开对事物的数量描述,但描述不是最后目的,它是以对部分观测资料的表述为依据,用以推断事物更为全面的情况或预测事物未来的发展变化,提供更为全面可靠的信息。因为在实际工作中,要研究的某种现象的单位个数通常是很庞大的或是无限的,不必要也不可能对所有单位进行全部观测,只须任意观测部分单位数;统计上把全部单位数称为总体,把任意观测的部分单位数称为样本。数理统计学就是根据样本观测资料所提供的信息,运用一定的数学模型和方法,对样本资料所提供信息的不确定性予以测度、分析和归纳,用以对总体一般情况进行有一定可信度的推断。简而言之,数理统计学

就是根据样本资料,应用数学模型,去推断总体的科学。

第三节 数理统计学与数学、概率论 和社会经济统计学的关系

由于科学技术的日益发展,一方面不断产生新的科学体系,对事物的某个方面进行更加深入的探讨和研究;另一方面又由于事物间的复杂性和联系性,为了适应研究的需要,不断出现边缘性的和综合性的科学。数理统计学由于其研究对象的广泛性,它与许多科学都有联系;特别是近代发展起来的一些综合性科学,如计量经济学、信息系统工程学等,它们都与数理统计学有着深刻的联系。从数理统计学的对象、方法和内容来看,它与数学、概率论和社会经济统计学的关系尤为密切。对这些关系我们不能不有一个明确的概念。

首先,数理统计学与数学的关系。由于数理统计学贯串着许多数学理论,因此有人把它视为数学的一个分支。这里所指的数学是一般概念的数学,它与数理统计学在性质上是不同的。一般说来,数学是抽象地研究事物的纯数量关系,是确定性的,例如各种函数关系和几何关系等。数理统计学研究的数量关系,虽然也是抽象的,但这些数学关系是非确定性的,即具有随机性质,例如各种相关关系和因果关系等。由于数理统计学对数量关系的研究,比较接近实际,因此也有人把它视为一门应用数学。这种看法看重了它的数学的共同性,而忽略了其统计的特殊性。

其次,数理统计学与概率论的关系。数理统计学与概率论都是以大量随机现象的数量规律性为研究对象,这就决定了它们间有密切不可分的关系,很难有明确标准把它们完全分开。实践中既然将它们作为两门科学,它们间总存在着一些区别。这些区别

大体可以概括地理解为:概率论是从纯理论的角度研究大量随机现象的数量规律性,属于纯数学的范畴,是数学的一个分支;数理统计学则是将概率论所建立的公理化概念、模型和运算法则等理论具体应用到实践活动中去,如应用到各种自然科学、技术科学、物质生产部门以及社会经济领域中去,与各部门的具体情况相结合,产生具有该部门特殊性的统计理论和方法。同时又从各个部门的特殊统计理论和方法中,将其具有共同性的东西抽象出来,这就形成了日益发展成熟的一门独立的数理统计学,同时也不断地丰富和发展了概率论。

数理统计学与社会经济统计学之间,存在许多的共同点:首先它们都是从事物的量的规律性方面进行分析研究;其次分析研究的方法上是基本相同的,如大量观察法、分组整理法和综合指标法等;再次分析研究的目标也是一致的,都是根据观察的部分个体单位的样本资料来估计或推断全部个体单位数总体情况。但它们也是有区别的,社会经济统计学顾名思义,它是以社会经济领域的现象为研究范围,社会经济现象有更大的随机变化的特殊性,很难把它控制在一定的固定条件下进行重复观测来获取样本资料,它所体现的统计规律性,不可避免的蕴涵着社会因素的影响。为了保证一定的估计精度,常常需要采用观察容量更大的样本资料,甚至可能需要采用普查方式对总体单位数进行全部观察。因此,一般的社会经济统计学大都着重描述性统计的研究。

数理统计学的研究范围,包括社会经济现象在内的遍及所有领域(部门)的随机现象,只不过在自然界或技术领域内的随机现象。一般是在一定的条件下,采用重复试验或观测的方法,获取单位数不太多的样本资料,因为它尽可能避免了人为因素或其他无关因素的影响,使得所分析研究的数量规律具有更多的客观性。欲用观测的单位数常常不是太多的样本资料,据以估计或代表总体情况。为了保证估计的精度和可靠性,因此,数理统计学在

样本数据基础上,着重推断统计的研究,在实践中可收到事半功倍的效果。

第四节 数理统计学的基本内容和作用

由于现代科学技术的突飞猛进,现代社会已由资本化为中心的时代,发展到信息化时代,大量而纷繁的信息资料数量化处理,有了高速先进的计算机,计算已不再有什么困难;但对大量信息的检测、分析,提出有高度可靠性的结论,必须依赖数理统计学广泛深入的应用。科学技术愈进步,数理统计学的重要作用也就越明显。

由于客观世界随机现象普遍存在,数理统计学在各个领域(部门)的深入应用,结合各领域的特点,就产生了门类众多的相对独立的应用统计学,在自然界产生了生物统计学、气象统计学、医学统计学、工业和农业试验统计学等;在社会经济领域中,除了社会经济统计学外,还产生了人口统计学、宏观和微观经济管理统计学等。

数理统计学在实际应用中,派生了出众多门类统计学,它们之间必然存在联系,实质上它们就是一门完整的数理统计学的核心内容。一般包括如下几部分:

1. 为取得样本数据进行随机观察或试验设计

被研究现象的总体单位数通常是庞大,或是无限的,如何从中抽取少量的个体单位用于被研究,必须事前精心设计,以保证严格遵守“随机原则”。所谓“随机原则”:就是在同一的条件下,使每个单位有同等被抽中的机会,被研究标志可出现或不出现,也可这样出现或那样出现,实地观测前是不知道的。只有严格遵守随机原则,才能充分发挥大数法则的作用,保证数量规律的客观

性,最后获得有高度可靠性的结论;不然就会使得观测或试验的可能误差过大,使推断结论达不到所要求的可靠程度。

2. 使用数学模型描述数量的规律性

各种性质不同事物的随机现象的数量化,在大量重复试验或观察之下,其数值的变化都会呈现出比较稳定的分布形态,对这种分布形态,可用恰当的数学模型予以描述,这就抽象掉了个别事物随机现象的特殊性,呈现它们共同的规律性。概率论和数理统计学,根据不同的分布形态,建立了不同类型的数学模型予以表述。它们是对大量随机现象进行统计估计、检测和分析,得出合理结论,不可缺少的工具。

3. 对总体情况进行估计或推断

根据样本资料提供的信息,对总体情况进行推断,是数理统计的主要作用,也是重要内容。例如根据样本数据的水平指标,推断总体水平指标;根据样本中两件事物的差异,判断两事物总体是否有实质性差异;或对某种情况所作的统计假设,判断其是否可信等等,这些都是对总体情况的推断。这种推断是有科学依据的,对推断结论的准确性提供一定的可信度,即可以通过多次的重复试验可以验证它是无误的。这种科学的推断或估计,有广泛的实际意义。

4. 分析影响事物发展变化的因素

一种事物的发展或变化,总是受着多种因素综合影响。这些因素,不可能都是同等重要的,必然有些是主要,有些是次要的;而且各个因素的影响作用,也不是固定不变的。数理统计学对这些影响因素的试验分析,可以判断哪些因素有主要的作用,哪些因素仅有次要的作用,哪些因素根本没有作用。这种分析在工业试验、农业试验、或其他各科学技术试验中,有极为重要作用。

5. 分析事物间的相互关系

在客观世界中,任一事物在时间或空间上的发展变化,总是

与其他多种事物在同一时间或空间上的发展变化,既有密切关联,又互相制约,孤立的发展变化,是不可能存在的。这种相互关系,在某一些事物之间有程度不同的内在联系,在另一些事物之间不存在内在联系,只有互相依存的关系;总的说来,这种相互间的关系不是一种因果关系,在统计上称之为相关关系。这种相关关系发展变化的数量规律性,根据事物发展变化的性质和条件,可选用合理的数学模型予以表述。在实践中,应用这些数学模型所表述的事物相互关系发展变化的趋势,可以对被选定为研究目标的事物,在观测的空间和时间之外的近期内进行控制和预测,这在信息时代有十分重要的作用。如自然界的天气预报、地震预报等;在社会经济领域里,各种宏观和微观的经济预测等,为经济管理和决策,提供重要的信息和科学依据。

综上所述,数理统计学是具有系统理论基础和广泛实用价值的科学。国内有关翻译和编著已出版不少,各有不同的论述角度、目的要求和特点,深浅详实程度各不相同。本书拟将概率作为数理统计学最基本的理论,比较系统地加以介绍和论述,尽可能避免用数学语言论述,采用通俗易懂语言加以诠释。在适用范围上只限于社会经济领域和工农业生产技术的试验或检验。希望对高等财经院校统计专业本科生和研究生,以及从事统计实践和研究的工作者,能予以帮助。

第二章 随机事件和概率

第一节 随机事件

一、随机事件的概念

数理统计学的研究对象包含,客观世界中存在着对立统一的偶然现象和必然现象。对这两个概念及其对立统一的关系如何理解和体现,还需涉及一些有关的概念,才能逐步深入展开讨论。

研究任何事物都离不开其所在的一定的客观环境和条件。在一定条件下,对一事物进行观察、测量、检验等活动,称为试验;对试验的结果,统称为事件。现实生活中,最常见的事件有下述三种。

1. 必然事件

在一定条件下,必然要出现的事件就叫必然事件,也就是必然现象。必然事件具有确定性,即在重复试验中,只要条件不变,每次试验的结果肯定都是相同的。例如:“在标准大气压下,水加热到摄氏 100°C ,必定沸腾”,“在同样的原材料、生产设备和工艺制造过程下,必然生产出同种类型的产品”等。通过这两个日常所见的例子,即可概括说明前述概念:烧一次开水和生产一件产品,分别就是一次试验;标准大气压和加热到 100°C ,是每烧一次水的固定条件;同样的原材料、生产设备和工艺制造过程,是每生产一件产品的固定条件;水沸腾和产出同类型产品,就是必然事