



Ying Yong Shuli  
**Tongjifangfa**



**应用数理  
统计方法**

# 应用数理 统计方法

李俊德 编



河南科学技术出版社

## 应用数理统计方法

李俊德 编

责任编辑 范云操

河南科学技术出版社出版

河南省新乡市印刷厂印刷

河南省新华书店发行

850×1168 32开 14.25印张 328千字

1985年11月第1版 1985年11月第1次印刷

印数：1—5,000

统一书号4245·3 定价2.35元

## 前 言

本书的主要目的是介绍数理统计的常用方法及其在实践中的应用。数理统计是一门理论完整、应用性很强的数学学科。其发展非常迅速，渗透到许多科学分支，应用到国民经济的各个部门。因而，数理统计，不论在理论研究还是在联系实际方面，都是当前数学中最活跃的分支之一。

全书共十三章，可分为两部分：第一部分，介绍概率论与数理统计的基础知识（第1～3章）；第二部分，介绍各种数理统计方法（第4～13章）。

第一部分内容包括：总体和样本，事件和概率，随机变量和概率分布，样本统计和频率分布等；为本书以后各章的学习奠定必要的理论基础。

第二部分介绍数理统计的主要内容。实际上，数理统计就是研究各种统计方法的理论背景及其在实际中的应用。本书所介绍的统计方法依其特征可分为：统计推断，包括点估计、区间估计和统计假设检验等；统计预测，包括回归分析、判别分析以及平稳时间序列和链状相关时间序列等；样本设计，包括正交试验设计和抽样方法等。

除第一章外，每章之后均附有一定量的习题，因为演练习题是学习掌握各种数理统计方法的重要措施之一。当然，能应用所学的知识解决本职工作中的一两个实际问题，那将是学习数理统

计的最好方法。

本书重在应用。为此,在讲述各种方法时,尽量做到先以实际问题引入,然后概括出一般的方法步骤及应用范围,最后举例说明。为使读者了解各种方法的来龙去脉,对其主要的理论依据亦加以扼要的介绍,但不作繁琐的推导证明,以免涉及高深的数学知识。因而,具有中学以上文化水平的读者皆可阅读。

本书内容是笔者在应用和推广数理统计的实践中逐步编写成的,其间得到很多同志的大力支持和帮助。书中有些例题和习题选自战斗在生产、科研第一线的同志们的研究成果。另外,在编写中,还参考了一些兄弟单位的交流资料。谨此一并致谢。

由于理论水平所限,实践经验不足,对书中缺点和错误,恳切希望广大读者批评指正。

编 者

## 内 容 提 要

全书共十三章，内容包括：概率统计基础知识，假设检验，统计预测，样本数据的收集，时间序列法的应用。该书主要讲解各种方法的实施步骤和应用范围，并扼要介绍其主要数学理论依据，但不作繁琐的推导证明。语言通俗。

本书供具有中学以上文化水平的从事统计工作的有关技术人员和管理人员阅读，亦可供理工农医大、中专院校学生和中学教师参考。

# 目 录

第一章	绪论	( 1 )
§1.1	什么是数理统计方法	( 1 )
§1.2	数理统计的起源和发展	( 2 )
§1.3	总体和样本	( 8 )
第二章	概率分布	( 10 )
§2.1	随机事件	( 10 )
§2.2	事件的概率	( 14 )
§2.3	离散型随机变量的概率分布	( 17 )
§2.4	正态分布	( 29 )
§2.5	随机变量的数字特征	( 40 )
§2.6	大数定理和中心极限定理	( 47 )
第三章	总体特性的估计	( 57 )
§3.1	频率分布表	( 57 )
§3.2	频率分布图	( 60 )
§3.3	样本平均数	( 64 )
§3.4	样本方差	( 71 )
第四章	大样本理论和统计假设检验	( 82 )
§4.1	样本平均数的分布	( 82 )
§4.2	两个样本平均数之差的分布	( 87 )
§4.3	点估计	( 89 )

§4.4	区间估计(I)	( 96 )
§4.5	统计假设检验的基本思想	( 102 )
§4.6	U检验法	( 106 )
§4.7	两类错误	( 111 )
<b>第五章</b>	<b>小样本理论和统计假设检验</b>	<b>( 117 )</b>
§5.1	t分布	( 118 )
§5.2	t检验法	( 123 )
§5.3	$\chi^2$ 分布	( 129 )
§5.4	$\chi^2$ 检验法	( 132 )
§5.5	F分布	( 135 )
§5.6	F检验法	( 142 )
§5.7	区间估计(II)	( 145 )
<b>第六章</b>	<b>非参数统计假设检验及<math>\chi^2</math>分布的应用</b>	<b>( 157 )</b>
§6.1	符号检验法	( 158 )
§6.2	秩和检验法	( 162 )
§6.3	联列表中的独立性检验	( 169 )
§6.4	拟合优度检验	( 176 )
<b>第七章</b>	<b>方差分析</b>	<b>( 183 )</b>
§7.1	引言	( 183 )
§7.2	一元方差分析	( 185 )
§7.3	二元方差分析	( 197 )
§7.4	随机化区组试验的方差分析	( 211 )
<b>第八章</b>	<b>回归分析</b>	<b>( 219 )</b>
§8.1	引言	( 219 )
§8.2	预报因子检验法	( 222 )
§8.3	一元线性回归分析	( 232 )



§8.4	一元非线性回归分析	( 250 )
§8.5	二元线性回归分析	( 257 )
§8.6	多元线性回归分析	( 266 )
<b>第九章</b>	<b>判别分析</b>	<b>( 278 )</b>
§9.1	数学模型	( 278 )
§9.2	二级判别分析	( 283 )
§9.3	逐步二级判别分析	( 291 )
§9.4	距离法	( 303 )
<b>第十章</b>	<b>平稳时间序列法</b>	<b>( 309 )</b>
§10.1	概述	( 309 )
§10.2	平稳时间序列的线性外推法	( 312 )
§10.3	显著性相关函数值预报法	( 320 )
§10.4	时间序列的平稳性检验	( 325 )
<b>第十一章</b>	<b>链状相关时间序列法</b>	<b>( 330 )</b>
§11.1	概述	( 330 )
§11.2	一重链状相关预报法	( 332 )
§11.3	二重链状相关预报法	( 336 )
§11.4	链状相关综合预报法	( 340 )
<b>第十二章</b>	<b>正交试验设计</b>	<b>( 347 )</b>
§12.1	试验方案的设计	( 347 )
§12.2	试验设计的直观分析	( 353 )
§12.3	交互作用	( 360 )
§12.4	试验设计的方差分析	( 368 )
§12.5	重复试验	( 376 )
§12.6	在较佳生产工艺条件下指标的估计	( 383 )
<b>第十三章</b>	<b>抽样方法</b>	<b>( 390 )</b>

§13.1	基本概念	( 391 )
§13.2	简单随机抽样	( 396 )
§13.3	计数抽样方法	( 399 )
§13.4	计量抽样方法	( 403 )
§13.5	分层随机抽样	( 412 )
附表		( 420 )
附表1	随机数表	( 420 )
附表2	标准正态分布函数数值表	( 424 )
附表3	t 分布双侧临界限表	( 425 )
附表4	$\chi^2$ 分布上侧临界限表	( 427 )
附表5	F分布上侧临界限表	( 428 )
附表6	符号检验表	( 432 )
附表7	秩和检验表	( 433 )
附表8	检验相关系数 $\rho=0$ 的临界值( $r_0$ )表	( 434 )
附表9	$r$ 到 $z$ 的变换数值表( $z = \frac{1}{2} \ln \frac{1+r}{1-r}$ )	( 435 )
附表10	常用正交表	( 436 )

# 第一章 绪 论

## §1.1 什么是数理统计方法

在生产斗争和科学实验中，我们经常搜集到许多数据。这些数据是科学研究的宝贵财富，它为人们提供许多有用的信息。但这些信息往往并非一目了然，而是蕴藏在大量的数据之中。那么，我们应如何从这些大量的数据中获得更多的信息呢？数理统计为我们提供了许多有效的方法。

实际上，数理统计就是专门研究关于数据的搜集、整理、分析和判断的一门数学学科。

数据的收集，是通过对研究对象的观察、测量或计算而获得数据的过程。数据，是数理统计的研究对象，正象点、直线和平面是几何学的研究对象一样。数据的获得比较直观，但它是数理统计的重要内容之一。因为，如果数据搜集得不准确，自然代表性不强，那么，任何再好的数理统计方法也无济于事。为了获得可靠性强、有代表性的数据，往往通过随机抽样，获得样本数据（这部分内容，我们将在本书第十三章里详细介绍）。

数据的整理，是通过计算、列表或作图的方式，对搜集到的数据进行归纳、压缩、整编和简化的过程。数据的整理（也称数据的简化）是对数据的初步分析，通常是从整体上对数据所反映的特性，进行粗略的定性估计。

数据的分析，往往通过样本函数（即统计量），如样本平均数 $\bar{X}$ 、样本方差 $S^2$ 等，较深入地揭露数据所反映的总体的特征性质，对数据进一步加工和简化。数据的分析比数据的整理前进了。前者，侧重对总体的特性、参数进行定量的研究；后者，着重对总体特性的定性探讨。

关于数据的整理和分析，将在第三章里详细介绍。

数据的判断，这是数理统计研究的主要内容。它是根据数据所反映的信息，应用概率论的理论，对数据所来自的总体的有关特性进行预测、估计和推断的过程。这方面的内容相当广泛，如参数估计、显著性检验、回归分析、判别分析、方差分析等等。实际上，数理统计的主要内容就是，研究如何由样本的特性来估计、推断总体特性的各种方法及其理论背景。

综上所述，数理统计方法，就是对数据进行搜集、整理、分析和判断的各种方法的总称。

## §1.2 数理统计的起源和发展

任何一门学科的产生和发展，都离不开实践的需要，离不开已建立的其他邻近学科。对数理统计这门学科来说，尤其是这样。因而在谈到数理统计的起源和发展时，必须介绍概率论的产生和形成，因为概率论是数理统计的理论基础。

### I 起源

数理统计与概率论的关系，可以用测地学与几何学的关系来比拟。几何学产生于土地的测量，这是众所周知的。概率论，也是通过人们观察大量的随机现象，搜集大量的数据，进行归纳、

分析，而逐步产生出来的。所以，从某种意义上说，概率论的创立，与初等统计是有密切关系的。

目前我们所说的数理统计，它是研究样本数据的搜集、整理、分析和判断的各种统计方法及其理论背景的一门纯数学学科。通常人们所说的统计学就其内容来说，大致可分为两部分：一是描述性的统计学（或称初等统计学）；一是分析性的统计学（或称高等统计学，即数理统计）。前者，重点研究数据的搜集和整理，对数据进行直观的描述；后者，重点对数据进行分析 and 推断。

描述性的统计学的产生，可以追溯到很远。

公元前2250年，大禹治水，根据山川土质，人力和物力的多寡，分全国为九州；殷周时代实行井田制，按人口分地，已进行了土地与户口的统计；春秋时代常以兵车多寡论诸侯实力，可见已进行了军事调查和比较；汉代全国户口与年龄的统计数字有据可查；明初编制黄册与鱼鳞册两种，黄册乃全国户口名册，鱼鳞册系全国土地图籍，绘有地形，完全具有现代统计图表的性质。可见我国历代对统计工作还是很重视的，只是缺乏系统的研究，未形成专门的著作。

在西方各国，统计工作开始于公元前3050年，埃及建造金字塔，为征收建筑费用，对全国人口进行普查和统计。到了亚里斯多德（Aristotle）时代，统计工作开始往理性演变。这时，统计在卫生、保险、国内外贸易、军事和行政管理等方面的应用，都有详细的记载。统计（Statistics）一词，就是从意大利文Statisti（意指国家、政治）逐步演变而成。

到了15世纪，意大利进入了文艺复兴时期。一些随机博弈盛行，有的赌徒为了获胜，终日冥思苦想，作了大量的试验和统计工作，从中发现一些解释不了的现象，便去请教当时著名的数学

家、天文学家吉里埃(Galileo 1564~1642)。吉里埃研究了赌徒们提出的问题后,得出了关于概率论的一些简要而有价值的定理。这些定理,为数理统计的发展奠定了根基。

到了16、17世纪,各种娱乐和赌钱的方法越来越复杂,这样,有些人又提出了一些新的问题,需要专家们来解释。如当时法国的一位叫梅耳(Me're')的著名赌徒,他曾向当时的哲学家和数学家巴士加(B. Pascal 1623~1662)提出如下问题:掷一粒骰子,4次中至少出现一个6的机会,要比掷两粒骰子4次中至少出现一对6的机会更多些,这是否成立?这一问题,引起了巴士加和他的朋友——另一位数学家费尔马(Fermat 1601~1665)的兴趣。两人经几次通信,认真研究,证明梅耳的猜想是正确的。因为

$$1 - \left(\frac{5}{6}\right)^4 = 0.5177$$

$$1 - \left(\frac{35}{36}\right)^4 = 0.4914$$

后来,巴士加将梅耳提出的问题推广到一般随机现象中去,得到一般化的解法,写出了关于现代概率论的原始形式和组合分析等专著。

## I 发展

回顾数理统计的发展历史,大致可分为如下三个阶段:

古典时期(19世纪末以前):主要是描述性的统计学的形成和发展,也可以说是数理统计的萌芽时期。

近代时期(19世纪末至1945年):小样本理论形成,数理统

计的主要分支建立，可以说是数理统计的形成时期。

现代时期（1945年以后）：这个时期，由于计算机的发明和使用，推动着数理统计在理论研究和应用方面都不断地向纵深发展，并产生一些新的分支和边缘性的新学科，如最优设计和非参数统计推断等。

1. 古典时期。瑞士数学家贝努里(J. Bernoulli 1654—1705)系统地论证了大数定律。贝努里的后代D.贝努里(1700—1782)将概率论的理论应用到医学和人类保险事业上。德谟斐尔(De Moivre 1667—1754)用半生时间为贵族解答赌博中的数学难题；他同时也很爱好纯数学的研究，于1733年首先发现了正态分布的密度函数的公式

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad -\infty < x < \infty \quad (1.1)$$

并计算出该曲线在各种不同区间内的概率。这样为整个大样本理论奠定了基础。德国数学家高斯(Gauss 1777—1855)在研究观测误差理论时，也独立发现了正态曲线的方程式(1.1)；他同时还创立了最小二乘法以及随机观测误差的理论。法国数学家拉普拉斯(Laplace 1749~1827)也独立地发现了正态曲线方程式，并使概率统计的理论应用到天文学方面的研究。凯尔穆(Kramp)于1799年首先编制了正态密度函数数值表。贝叶司(Bayes 1702—1761)是第一个将数理统计知识应用到工业上的人。欧拉(Euler 1707—1783)首先用希腊字母“ $\Sigma$ ”表示求和的缩写符号，应用于数理统计，进行数据的简化运算。阿莱木本特(Alembert 1717—1783)第一个用数理统计的方法，来处理气象方面的数据。

在1835—1870期间，比利时统计学家寇特莱特(Quetelet

1796—1874) 在概率论和数理统计的理论与应用方面都做出了重要贡献。他不仅将数理统计应用到生物学方面,而且应用到教育学和心理学的研究,并且详细地论证了数理统计应用的广泛性。他曾预言:“统计方法,可应用于各种学科的各种部门。”

在19世纪最后15年里,英国人类学者哥尔顿(Galton 1822—1911),通过生物统计的研究,开辟了生物学研究的新领域。哥尔顿和他的继承人英国统计学家K.皮尔逊(K. Pearson 1857—1936),通过生物统计的研究,于1889年发表了《自然的遗传》一书。在该书里,他首先提出了回归分析问题,并给出了计算简单相关系数和复相关系数的计算公式。K.皮尔逊在研究样本误差的效应时,首先提出了 $\chi^2$ 检验问题。他还引入了样本方差和标准差的概念。

在这个时期,概率论学科的主要研究内容已基本形成;而数理统计的研究还处于感性阶段,它的许多分支,象回归分析、方差分析、假设检验等仍处于萌芽时期,而且更多的分支还未被涉及到。

2.近代时期。如果说古典时期是概率论的形成时期,那么,近代时期则是数理统计的形成时期。在本世纪初,由于概率论的发展从理论上接近完备,加上工农业生产的迫切需要,推动着数理统计这门学科的蓬勃发展。一个个小样本分布的理论接踵而起,名家辈出,百花争艳,一系列带有根本性的重要概念和方法被提出,数理统计的许多分支在这个时期纷纷形成,使数理统计形成了一门纯数学性的数学的独立分支。

在1900年,K.皮尔逊首先提出了 $\chi^2$ 分布。这是数理统计发展史上第一个出现的小样本分布。

1908年是数理统计发展史上极其重要的一年。英国的统计学



家傲赛特(Gosset 1876—1937),以笔名“学生”(Student)在该年的生物统计杂志(Biometrika)上发表论文,创立了小样本检验代替大样本检验的理论和方法(即 $t$ 分布和 $t$ 检验法),为以后小样本的统计分析和推断开创了新纪元。 $t$ 分布的发现,其意义远远超过其结果本身的含义。它是研究小样本函数的精确理论分布中一系列重要结论的开端,为数理统计的另一分支——多元分析奠定了理论基础。

傲赛特的研究成果,当时并未引起人们的重视,约历经十年,英国剑桥大学教授、统计学家费歇(A. Fisher 1890—1962)应用、推广了傲赛特的 $t$ 检验法,才被大多数人们所接受。同时,费歇发展了显著检验及估计理论。另外,他还从事农业试验的安排及试验数据分析的研究,创立了正交试验设计和方差分析等数理统计的新分支。

奈曼(Neyman 1894—1981)和S.皮尔逊(S. Pearson)在1928年开始了关于统计理论的研究工作,明确地提出:假设检验问题,是一个数学上的最优化问题;区间估计问题也可看成一种数学上的最优解问题。

威尔得(A. Wald 1902—1950)的两本著作《序贯分析》和《统计决策函数》,被认为是数理统计上的很大成就。

北京大学许宝录(1910—1970)教授在极限理论、马氏过程、多元分析、正交设计、过程设计和判别函数等许多方面都有突出的贡献,他的许多研究成果都达到了世界先进水平。

至此,数理统计的一些重要分支如假设检验、回归分析、方差分析、正交设计、判别分析、序贯统计等,这个时期已有了决定其面貌的内容和理论。数理统计已成为理论完整、内容丰富、应用广泛、方法独特的一门纯数学学科。K.皮尔逊、傲赛特和费