研究生教材

数理统计

汪荣鑫 著

西安交通大学出版社

研究生教材

数理统计

汪荣鑫 编著

西安交通大学出版社

内 容 提 要

本书比较系统地介绍数理统计的基本概念，原理和方法。全书分五章，包括抽样和抽样分布，参数估计，假设检验，方差分析和正交试验设计，回归分析；书后并附有概率论基本知识。

本书基本概念叙述清晰，循序渐进，内容较为全面，应用性强，各章配有大量的例题和习题，便于教学和自学。

本书可作高等院校工科各专业研究生或高年级学生教材，也可供科学技术工作者阅读参考。

数 理 统 计

汪荣鑫 编著

西安交通大学出版社出版

（邮政编码 710049）

西安新华印刷厂印装

陕西省新华书店经销

开本 850×1168 1/32 印张 9.25 字数:230 千字

1986 年 10 月第 1 版　　1996 年 2 月第 6 次印刷

印数:27701—31700

ISBN 7—5605—0101—X/O·18　　定价:7.50 元

《研究生教材》总序

研究生教育是我国高等教育的最高层次，是为国家培养高层次的人才。他们必须在本门学科中掌握坚实的基础理论和系统的专门知识，以及从事科学研究工作或担负专门技术工作的能力。这些要求具体体现在研究生的学位课程和学位论文中。

认真建设好研究生学位课程是研究生培养中的重要环节。为此，我们组织出版这套《研究生教材》，以满足当前研究生教学，主要是公共课和一批新型的学位课程的教学需要。教材作者都是多年从事研究生教学工作，有着丰富教学和科学研究经验的教师。

这套教材首先着眼于研究生未来工作和高技术发展的需要，充分反映国内外的最新学术动态，使研究生学习之后，能迅速接近当代科技发展的前沿，以适应“四化”建设的要求；其次，也注意到研究生公共课程和学位课程应有它最稳定、最基本的内容，是研究生掌握坚实的基础理论和系统的专门知识所必要的，因此在研究生教材中仍应强调突出重点，突出基本原理和基本内容，以保持学位课程的相对稳定性和系统性，内容有足够的深度，而且对本门课程有较大的覆盖面。

这套《研究生教材》虽然从选题、大纲、组织编写到编辑出版，都经过了认真的调查论证和细致的定稿工作，但毕竟是第一次编辑这样的高层次教材系列，水平和经验都感不足，缺点与错误在所难免。希望通过反复的教学实践，广泛听取校内外专家学者和使用者的意见，使其不断改进和完善。

西安交通大学研究生院

西安交通大学出版社

1986年12月

前　言

概率论与数理统计是从数量上研究随机现象规律性（即统计规律）的数学学科。数理统计更着重于从试验数据出发来认识随机现象的规律。随着科学技术的发展，对随机现象需要认识，数理统计方法越来越广泛地被人们所采用。今天，在我国“四化”建设中，数理统计已应用于物理、化学、生物、工业、农业、医学、管理等各个领域。

目前，在高等院校中大部分工科专业的研究生都要学习数理统计，甚至对某些专业的高年级本科生也开设此课。

编者曾多次对工科各专业研究生讲授过数理统计。在讲稿的基础上，1985年编写了数理统计讲义，并在西安交大研究生中试用，后经修订和补充写成本书。

工科学生学习数理统计，要求正确理解基本概念和原理，能熟练运用统计方法。编者力图把概念讲得清晰，循序渐进。为了帮助学生理解概念，常用实例说明。在讲解统计原理时，尽量多作直观解释，删去较长的数学证明。

本书包括的数理统计内容较为全面。统计推断是数理统计的核心。本书讲解统计推断时，采用先讲参数点估计和区间估计，再讲假设检验的方式。这种方式的优点是两部分内容比较均衡。在内容的选择上，尽量选取工科各专业用得较多的项目，例如正交试验设计，一元非线性回归，多元线性回归等。另外，为了节省篇幅，把有些容易掌握的方法放在习题中，例如简化计算公式的推导和应用。为使读者便于阅读，书后附录中扼要介绍了概率论的基本知识。

本书应用性较强。为了培养学生运用统计方法解决实际问题

的能力，书中配有大量带应用性的例题与习题，并在讲解例题时指出所需注意事项。但是，也配备了一定数量对掌握统计原理有帮助的理论题。书后附有习题答案。

编者希望把本书写成一本便于教学也便于自学的教材。要求读者具备工科高等数学和概率论基本知识。

本书中图与表采用的编号是以章作区分的，如图 1-2 表示第一章第 2 图，表 2-3 表示第二章第 3 表。公式编号是以节作区分，如式(1.3)表示 §1 中第 3 式，而不指出所在的章。

全书讲授约需 40 学时。对高年级本科生，可以删去打星号的章节，约需 32 学时。

本教材由吴云江同志帮助选配习题，并作出全部习题的答案。

本书由西北工业大学朱言堂同志审阅，并提出了很多宝贵意见。在编写过程中我校张文修同志提出了不少有益的建议和意见，周家良同志看了部分章节并提出意见。本书的出版得到西安交大研究生院和西安交大出版社的热情支持和帮助。谨此一并致谢。

本书的姐妹篇——随机过程，不久即能问世与读者见面。

由于编者水平所限，错误之处在所难免，恳请读者批判指正。

编　者

1986 年 7 月

引　言

在日常生活中，"统计"一词通常是指收集资料，登记数据，画成图表，进行一些简单的计算。例如，某汽车厂的统计员为了统计汽车日产量，可以把某月的日产量画成图表，进而计算累计数、百分率、平均数，从而得到月产量、各天日产量所占百分率、平均日产量。这是带有全局性的统计。如果研究的对象比较庞大，全局统计比较困难，可作局部性统计，即可先简缩数据，然后象上面一样，对于需要的指标作一些计算。

下面考察一种局部性统计。例如，为了对一大批产品进行质量检查，将每个产品分为一等、二等、次品。如果采用全部逐个检查的方法，费时费工不合算。为了节省时间和费用，可以采用抽样检查的方法，即从中随意地抽取一部分进行检验，然后根据这一部分产品的质量情况，分析、推断整批产品的质量。特别是对于试验具有破坏性的质量检查，如灯泡寿命试验、炮弹射程试验等，必须采用抽样检查的方法。抽样检查法要求用较少的数据比较合理正确地推测整体情况。从整体中随意地考察一个局部，而由此局部分析、推断整体情况，是数理统计讨论的主题。

数理统计包括两个方面内容：一个是怎样合理地搜集数据——抽样方法，试验设计；另一个是由收集到的局部数据怎样比较正确地分析、推断整体情况——统计推断。当然，对不同的抽样方法、试验设计，采用的统计推断方法是不同的。本书主要讲述统计推断。

鉴于数理统计中局部数据是从整体中随意地抽得的，带有随机性，所以数理统计以概率论作为理论基础。然而，概率论中一些基本量(如随机变量的概率分布、数学期望、方差等)怎样用试

验值来确定，虽然在概率论教材中已有所涉及，但对试验次数应取多大、精确性如何等更深入一步的问题没有进行讨论，这些问题在数理统计中将会得到解决。

数理统计是一门应用性很强的数学学科。它已被广泛而深入地应用到自然科学和工程技术的各个领域，如物理学、力学、化学、生物、机械加工、无线电通讯、计算机、管理、冶金、地质、气象、农业、医学等方面，在我国四化建设中它将起积极的作用。

目　　录

第一章 抽样和抽样分布

本章主要介绍数理统计中一些基本术语和基本概念（如母体、子样、抽样、统计量等），以及一些重要的统计量分布——抽样分布。

§1 母体和子样

1.1 母体及其分布

所研究对象的全体元素组成的集合，称为**母体**或**总体**。母体中每一个元素称为**个体**。

例1 有一批产品共1000个，每个产品可区分为一等、二等、次品。我们要研究这批产品的质量，1000个产品的等级构成一个母体，每个产品的等级是个体。

例2 为考察在某种工艺条件下织出的一批布疋的疵点数，共取5000疋布。那末这5000疋布中每疋布疵点数的全体构成一个母体，每疋布的各自疵点数是个体。

例3 在检查某军工厂生产的一大批炮弹的质量时，若只考察炮弹的射程，那末，这批炮弹中每一颗的射程的全体构成一个母体，每颗炮弹的各自射程是个体。

从例2、例3可见，母体中的元素常常不是指元素本身，而是指元素的某种数量指标。在例2中，母体中元素指每疋布的疵点数，在例3中，母体中元素指每颗炮弹的射程。在例1中，如果一等品用“1”表示，二等品用“2”表示，次品用“0”表示，母体中元素是指每个产品的等级指标，同样，母体可看成数“1”、“2”、“0”的集合。从三个例子可以看出，数量指标取同一值的

元素可以有几个，也就是每一个值可以重复。母体是一个可重复的（即允许相同）数的集合。在例1的1000个产品中，值为“1”的有721个，值为“2”的有213个，值为“0”的有66个，因此“1”占$\frac{721}{1000}$，“2”占$\frac{213}{1000}$，“0”占$\frac{66}{1000}$。从数学角度说，**母体**是指所研究的数量指标可能取的各种不同数值的全体，而各种不同数值含有一定的比率。**母体分布**是指数量指标取不同数值比率的分布。

母体的数量指标用$\mathscr{X}$表示。从母体中随意地取得的一个个体是随机变量，记为X。显然，随机变量X所有可能取得的数值就是$\mathscr{X}$可能取的不同值的全体。X的概率分布与母体分布有什么关系呢？以例1为例，随机变量X的概率分布列为

X	1	2	0
p	$\frac{721}{1000}$	$\frac{213}{1000}$	$\frac{66}{1000}$

与$\mathscr{X}$取各种不同值的比率相同，即X的概率分布与$\mathscr{X}$的母体分布相同。这个结论具有普遍性。以后母体数量指标与相应的随机变量都用X表示，并不严加区分。**母体分布**指相应随机变量X的概率分布，可用分布列、分布密度、分布函数具体表示出来。**母体分布的数字特征**指的是相应随机变量的数字特征。

为方便起见，母体数量指标X有时简称为母体X。母体X的分布和数字特征采用概率论中随机变量的相应量的记号。母体X的分布函数、分布列和分布密度分别用$F(x)$、$P(x)$和$f(x)$表示。需要指出的是，分布列$P(x)$中的x只能取X所有可能取的数值$x^{(1)}, x^{(2)}\cdots\cdots$（有限个或可列无限多个）。母体$X$的平均数（亦称平均、均值、数学期望）、方差和标准差分别用EX、DX和$\sqrt{DX}$（或$\sigma[X]$）表示。对母体同样地可以定义矩。

上面是从母体得到随机变量。反之，从随机变量亦可得到母

体。例如，扔一颗骰子出现的点数是随机变量，它可能取得的不同值的全体"1"、"2"、"3"、"4"、"5"、"6"构成一个母体，它的分布是随机变量的概率分布。

1.2 子样

从母体中取得一部分个体，母体中的这一部分个体称为**子样**或**样本**。取得子样的过程称为**抽样**。一个子样中每一个个体称为**样品**。子样中个体的个数称为**子样容量**。表示子样时通常用小括号括起来。如在本节例1的母体中抽取一个子样(0，1，1，2，0，2，0，0，1，0)，容量为10。

在数理统计中，采用的抽样方法是**随机抽样法**，即子样中每一个个体（样品）是从母体中被随意地取出来的。随机抽样分重复抽样与非重复抽样二种。以例1为例，从1000个产品中抽取一个容量为10的子样，如果随机地取一个产品检查后放回，再随机地取一个检查后又放回，直至取得10个个体为止，这种方法称为**重复(或返回)抽样**。如果每取一个检查后不再放回，直至取得10个个体为止，或者一次抽取10个，这种方法称为**非重复（或无返回）抽样**。需要指出，随机抽样得到的子样，所含样品是有一定次序的，通常按它被抽到的先后顺序排列。

从母体X随机抽样得到的子样可以用n维随机变量$(X_1, X_2, \cdots\cdots, X_n)$表示。现在考察它的概率分布。在重复抽样情形，由于每取出一个个体检查后要放回，母体成分不变(母体分布不变)，所以$X_1, X_2, \cdots\cdots, X_n$是独立同分布的，并且每一个随机变量的分布与母体分布相同。对于非重复抽样，则分二种情形：在有限母体（即母体中个体总数有限）情形，因每取出一个个体后改变了母体的成分，所以随机变量$X_1, X_2, \cdots\cdots, X_n$不相互独立；在无限母体（即母体中个体总数是无限的）情形，每取出一个个体后并不改变母体的成分，所以随机变量仍然是独立同分布的，并且每一随机变量的概率分布都是母体分布。

在实际情况中，我们有时遇到的是有限母体，而采用的是无返

回抽样。此时，如果子样容量 n 相对于母体容量 N（母体中个体总数）很小，实用上要求 $\frac{n}{N}\leqslant 0.1$，可以把 $X_1,X_2\cdots\cdots,X_n$ 近似地看成独立同分布，而每个随机变量的分布都是母体分布。

如果子样 $(X_1,X_2,\cdots\cdots,X_n)$ 中各个体独立同分布，且每一随机变量的概率分布是母体分布，则称它为简单**随机子样**。这种子样数学上比较容易处理。在本书中，除特别指出外，子样都是指简单随机子样。

子样 $(X_1,X_2,\cdots\cdots,X_n)$ 是 n 维随机变量，这是对具体进行一次抽样前而言。在抽样后获得它的一组观察值 $(x_1,x_2,\cdots\cdots,x_n)$，称为**子样值**。为方便起见，有的时候子样与子样值亦可统称为子样。

设母体 X 的分布函数是 $F(x)$，则子样 $(X_1,X_2\cdots\cdots,X_n)$ 的概率分布函数

$$F_n(x_1,x_2,\cdots\cdots,x_n)=F(x_1)F(x_2)\cdots\cdots F(x_n)$$

在母体离散分布情形，设母体分布列为 $P(x^{(i)})=P\{X=x^{(i)}\}$，$i=1,2,\cdots\cdots$（有限个或可列多个），则子样的概率分布列

$$\begin{aligned}P_n(x_1,x_2,\cdots\cdots,x_n)&=P\{X_1=x_1,X_2=x_2,\cdots\cdots,X_n=x_n\}\\&=P(x_1)P(x_2)\cdots\cdots P(x_n)\end{aligned}$$

其中 $x_1,x_2,\cdots\cdots,x_n$ 每一个值都是在 X 所有可能取的值 $x^{(1)},x^{(2)},\cdots\cdots$ 之中。

在母体连续分布情形，设母体分布密度为 $f(x)$，则子样的概率密度

$$f_n(x_1,x_2,\cdots\cdots,x_n)=f(x_1)f(x_2)\cdots\cdots f(x_n)$$

对应于随机变量 X 有一个母体，如何讲抽样与子样呢？如果对随机变量独立重复地进行 n 次试验，所得观察值为一个简单随机子样，那么，进行 n 次试验观察相当于进行一次抽样，而且这是重复抽样。例如，对靶射击一次得到的环数是一个随机变量，

今独立重复地对靶射击7次，可以看作进行一次重复抽样，所得7个环数构成一个子样。

1.3 子样分布

子样分布刻划子样中数据的分布情况。它的定义方式类似于母体分布。通常有三种形式：频数分布和频率分布；经验分布函数；直方图。

一、子样频数分布和频率分布。先举二个例子：

例1 从织布车间抽取7疋布，检查每疋的疵点数，得到子样(0，3，2，1，1，0，1)。把7个数从小到大依次排列，相同的数合并，获得下列频数表：

X	0	1	2	3
频数	2	3	1	1

表 1-1

称它为子样频数分布。频数是子样中各个不相同数值出现的次数，而频率是频数除以子样容量。因此，子样的频率分布可以用下表给出：

X	0	1	2	3
频率	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{1}{7}$	$\frac{1}{7}$

表 1-2

例2 某工厂生产一批铆钉。从中抽取200个，测得其直径(单位：毫米)。所得数据从小到大依次排列，相同的合并，可以得到子样频数分布和子样频率分布如下：

X	13.13	13.14	13.18	13.20	13.23	13.24	13.25	13.26
频数	1	1	1	3	2	3	1	4
频率	0.005	0.005	0.005	0.015	0.010	0.015	0.005	0.020
X	13.27	13.28	13.29	13.30	13.31	13.32	13.33	13.34
频数	1	5	6	2	5	7	6	7
频率	0.005	0.025	0.030	0.010	0.025	0.035	0.030	0.035
X	13.35	13.36	13.37	13.38	13.39	13.40	13.41	13.42
频数	4	3	6	10	7	12	4	6
频率	0.020	0.015	0.030	0.050	0.035	0.060	0.020	0.030
X	13.43	13.44	13.45	13.46	13.47	13.48	13.49	13.50
频数	9	6	7	7	3	9	1	6
频率	0.045	0.030	0.035	0.035	0.015	0.045	0.005	0.030
X	13.51	13.52	13.53	13.54	13.55	13.56	13.57	13.58
频数	6	5	4	4	3	3	4	5
频率	0.030	0.025	0.020	0.020	0.015	0.015	0.020	0.025
X	13.59	13.60	13.61	13.62	13.63	13.64	13.66	13.69
频数	2	1	1	3	1	1	1	1
频率	0.010	0.005	0.005	0.015	0.005	0.005	0.005	0.005

表注　子样频数分布取X与频数二栏，子样频率分布取X与频率二栏。

表 1-3