



北京市高等教育精品教材立项项目

北京大学数学教学系列丛书



本科生
数学基础课教材

普通统计学

谢衷洁 编著



北京大学出版社
PEKING UNIVERSITY PRESS



北京市高等教育精品教材立项项目

北京大学数学教学系列丛书

普通统计学

谢衷洁 编著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

普通统计学/谢衷洁编著. —北京:北京大学出版社, 2004. 5

(北京市高等教育精品教材立项项目)

(北京大学数学教学系列丛书)

ISBN 7-301-06858-1

I. 普… II. 谢… III. 统计学-高等学校-教材 IV. C8

中国版本图书馆 CIP 数据核字(2003)第 119249 号

书 名: 普通统计学

著作责任者: 谢衷洁 编著

责任编辑: 刘 勇

标准书号: ISBN 7-301-06858-1/O · 0585

出版发行: 北京大学出版社

地 址: 北京市海淀区中关村 北京大学校内 100871

网 址: <http://cbs.pku.edu.cn> 电子信箱: zpup@pup.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 理科编辑部 62752021

排版者: 北京高新特打字服务社 51736661

印刷者: 北京大学印刷厂

经 销 者: 新华书店

890×1240 A5 11.5 印张 340 千字 28 页插图

2004 年 5 月第 1 版 2004 年 5 月第 1 次印刷

印 数: 0001—4000 册

定 价: 25.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,翻版必究

统计学是关于收集和分析数据的 科学和艺术

——《不列颠百科全书》

内 容 简 介

本书是一部为高等院校大学生讲述统计学基础知识的入门性教科书。全书共分八章,内容包括:开篇引言,绪论,描述性统计学,概率论基础知识,随机变量的分布与数字表征,假设检验,相关分析与回归分析,抽样方法等;并配有各章的适量习题供教师和学生选用。本书作者长期从事数理统计的教学和科研工作,具有丰富的教学经验并取得了丰硕的研究成果。本书以社会上关心的热点和典型问题为开篇引言,以生动、精彩的实例为引线,深入浅出地讲述统计学的基本概念、思想和方法,教给读者如何去收集、分析随机数据,如何运用统计学的方法和技巧,并且基于统计分析的结果给出客观、公正和科学的结论。本书侧重于介绍统计学有关的概念和方法,强调统计学的分析与应用,因此本书中多选用典型实例和真实的数据(它们来自二十多个科学领域)作为统计方法的讨论对象,因而对统计学的教学和科研人员具有很好的参考价值。

本书可作为综合大学、高等师范院校、理工科大学非概率统计各专业大学生的“统计”课的入门教材,也可供从事数理统计工作的科技工作者、管理工作者阅读。本书不要求读者以学过微积分为先修课程,适合于文理科大学生阅读。

《北京大学数学教学系列丛书》编委会

名 誉 主 编：姜伯驹

主 编：张继平

副 主 编：李 忠

编 委：（按姓氏笔画为序）

王长平 刘张炬 陈大岳 何书元

张平文 郑志明

编委会秘书：方新贵

责 任 编 辑：刘 勇

作者简介

谢表洁 北京大学数学科学学院教授、博士生导师,1959年毕业于北京大学数学力学系,长期从事数理统计的教学和科研工作,取得了丰硕的成果。研究方向是时间序列分析及其应用,尤其是谱分析、建模及小波方法。其应用领域涉及天文学中天王星光环的发现,通讯卫星的交调分析,内分泌激素、脑诱发电位研究,海洋与大陆地球物理及宏观经济与汇率研究等多学科领域。至今已在国内外发表了80多篇论文,并有六部著作在国内外出版。历任北京大学统计实验室主任、数理统计研究所副所长;中国概率统计学会理事、副秘书长,时间序列专委会主任;是全国政协第八、九、十届委员,北京市第七、八届委员、常委。曾获国家自然科学奖三等奖、国家教委科技进步奖一等奖、北京市科技进步二等奖、中国人民解放军军级科技进步奖,并在国外获得第八届、第十四届维也纳控制论与系统研究大会的“最优论文奖”。

序 言

自 1995 年以来,在姜伯驹院士的主持下,北京大学数学科学学院根据国际数学发展的要求和北京大学数学教育的实际,创造性地贯彻教育部“加强基础,淡化专业,因材施教,分流培养”的办学方针,全面发挥我院学科门类齐全和师资力量雄厚的综合优势,在培养模式的转变、教学计划的修订、教学内容与方法的革新,以及教材建设等方面进行了全方位、大力度的改革,取得了显著的成效。2001 年,北京大学数学科学学院的这项改革成果荣获全国教学成果特等奖,在国内外产生很大反响。

在本科教育改革方面,我们按照加强基础、淡化专业的要求,对教学各主要环节进行了调整,使数学科学学院的全体学生在数学分析、高等代数、几何学、计算机等主干基础课程上,接受学时充分、强度足够的严格训练;在对学生分流培养阶段,我们在课程内容上坚决贯彻“少而精”的原则,大力压缩后续课程中多年逐步形成的过窄、过深和过繁的教学内容,为新的培养方向、实践性教学环节,以及为培养学生的创新能力所进行的基础科研训练争取到了必要的学时和空间。这样既使学生打下宽广、坚实的基础,又充分照顾到每个人的不同特长、爱好和发展取向。与上述改革相适应,积极而慎重地进行教学计划的修订,适当压缩常微、复变、偏微、实变、微分几何、抽象代数、泛函分析等后续课程的周学时。并增加了数学模型和计算机的相关课程,使学生有更大的选课余地。

在研究生教育中,在注重专题课程的同时,我们制定了 30 多门研究生普选基础课程(其中数学系 18 门),重点拓宽学生的专业基础和加强学生对数学整体发展及最新进展的了解。

教材建设是教学成果的一个重要体现。与修订的教学计划相

配合,我们进行了有组织的教材建设。计划自 1999 年起用 8 年的时间修订、编写和出版 40 余种教材。这就是将陆续呈现在大家面前的《北京大学数学教学系列丛书》。这套丛书凝聚了我们近十年在人才培养方面的思考,记录了我们教学实践的足迹,体现了我们教学改革的成果,反映了我们对新世纪人才培养的理念,代表了我们新时期的数学教学水平。

经过 20 世纪的空前发展,数学的基本理论更加深入和完善,而计算机技术的发展使得数学的应用更加直接和广泛,而且活跃于生产第一线,促进着技术和经济的发展,所有这些都正在改变着人们对数学的传统认识。同时也促使数学研究的方式发生巨大变化。作为整个科学技术基础的数学,正突破传统的范围而向人类一切知识领域渗透。作为一种文化,数学科学已成为推动人类文明进化、知识创新的重要因素,将更深刻地改变着客观现实的面貌和人们对世界的认识。数学素质已成为今天培养高层次创新人才的重要基础。数学的理论和应用的巨大发展必然引起数学教育的深刻变革。我们现在的改革还是初步的。教学改革无禁区,但要十分稳重和积极;人才培养无止境,既要遵循基本规律,更要不断创新。我们现在推出这套丛书,目的是向大家学习。让我们大家携起手来,为提高中国数学教育水平和建设世界一流数学强国而共同努力。

张继平

2002 年 5 月 18 日

于北京大学蓝旗营

前 言

20 世纪 90 年代中期的一次北京大学教授会上,当时任副校长的王义道教授指出:目前我校可供文科学生选修的理科课程太少,因为大部分理科课程都是要求学生学过高等数学为前提,而许多文科学生是没学过高等数学的;希望有科研和教学经验的老教授设法开一些不要求高等数学为先修课程,文科学生能接受的高质量的理科课程。在概率统计系同仁们的热情鼓励和帮助下,一门取名为“普通统计学”的全校通选课由我来执教。未曾想到开设之后受到全校文、理科学生的热烈欢迎,最多时 150 人的教室有 300 多人听课,学生来自 20 多个系。一个最大的难题就是听课学生中有数学学得很多的物理系学生,也有只具备中学初等数学的英语、俄语、历史系等数学基础较弱的学生,因此,如何将这门课的讲授,从教材到教学手段都能够吸引住数学背景有很大差异的众多学生就成为一个非常具有挑战性的课题。几年来的实践证明我们的教学是成功的,这首先是本课程始终受到文、理科学生的欢迎,上课的积极性和出勤率很高,那怕是在 SARS 期间学校通知学生可以不来课堂上课的情况下仍有 2/3 以上的学生来听课。

我们对本课程教学的指导思想是:

1. 侧重点是统计学的思想和方法,而不是严谨的数学论述和推导。但是书中介绍的统计学的思想和方法却是有扎实的理论基础的。

2. 统计学是和现实生活非常贴近的学科,切忌把本课程变成一门数学课。本书大力收集来自自然科学、工程技术、社会科学、医学、经济与金融等多种学科的实际数据作为提出统计学中

不同问题的实例,进而讲解统计学是如何解决这些问题的。由此,学生不仅感受到统计学的实用性,许多学生还会联想到本专业的问题,促进了他们的学习和研究兴趣。

3. 根据文、理科学生的不同需求,在练习和课外阅读材料上适当区别对待。本书有一些带*的段落是不属于课堂教学范围的,它们或者是属于数学推导或者属于扩大知识面的内容。有些学生不满足于课堂教学内容的,可在教师指导下阅读这些内容并做一些较难的练习。

4. 鼓励学生应用统计软件和电脑来解题,但前提条件是必须搞清楚有关的统计概念、背景并学会正确的统计方法的使用。当今社会上各种统计软件已相当流行,如 Excel, SAS, SPSS, S-plus 等以及如 TI 公司出品的 TI-89、92、V-200 等都是学习统计和将统计应用于研究的强有力的工具。我们的学生毕业后,在工作岗位上大部分也可以使用这些 IT 产品。在正确掌握统计的使用方法的基础上,使用这些手段可以大大节约计算时间。学生可以多做题,而且很多图形对学统计的人理解统计的内容和结果是很有帮助的。

本教材就是根据以上的指导思想、在经过前后五年的教学实践的基础上编写成的,为适应不同的教学对象,内容上有适当的扩充。但整体教材的绝大部分内容可在一个学期每周三学时的教学计划中完成。如果教学对象是文科学生,则概率论基础知识部分可以介绍得精练一点。这部分的理论和练习对他们来讲一般都会感到比较吃力。

本教材得到国家自然科学基金研究项目(No. 10171005)及北京市精品教材项目的资助,在此表示谢意。作者还要感谢日本学术振兴会(JSPS)的资助和日本早稻田大学有关方面的大力支持,使我有机会于 2002 年夏在早稻田大学充分利用他们的图书、资料的优势,参阅和收集许多优秀的统计学的多种教科书和数据并利用其提供的良好条件写出了本书的一些章节,有些内容在国内的图

书中是未曾见到的。

最后,作者还要衷心地感谢陈家鼎教授极其细心地阅读了全书,非常准确并富有建设性地指出了原书稿中的一些不足和欠妥之处。没有他的帮助,本书就不会以现在的面貌和读者见面。

然而,限于作者的水平及对陈家鼎教授修改意见理解的不足,本书的缺陷和错误在所难免,恳请同仁们及广大读者给予批评、指正,以便将来再版时给予更正。

谢表洁

2004年3月于北京大学

目 录

第〇章 开篇引言	(1)
第一章 绪论	(13)
1.1 引言	(13)
1.2 关于随机现象及其统计规律性	(14)
1.3 若干统计学中的名词	(19)
第二章 描述性统计学	(24)
2.1 引言	(24)
2.2 随机抽样及其频数	(24)
2.2.1 简单的随机抽样	(24)
2.2.2 数据的简单整理	(26)
2.3 直方图、频数密度分布图及频数多边形图	(31)
2.4 茎叶图	(38)
2.5 数字特征	(44)
2.5.1 刻画总体代表性的数值	(44)
2.5.2 刻画分散度的数值	(49)
2.5.3 均值、中位数与方差的简单性质	(53)
2.6 分位点和 Box-Whisker 图	(55)
2.6.1 Q_1 与 Q_3 (1/4 与 3/4 分位点)	(55)
2.6.2 Box-Whisker 图	(59)
2.7 本章的回顾和复习要点	(61)
第三章 概率论基础知识	(62)
3.1 古典概型	(62)
3.1.1 古典概型及其概率意义	(62)
3.1.2 若干组合分析中的定理	(64)
3.1.3 古典概型问题的直接计算	(67)

3.2 随机事件的运算及概率的公理化	(76)
3.2.1 随机事件的运算	(76)
* 3.2.2 概率空间($\Omega, \mathcal{F}, P$)	(79)
3.2.3 若干常用的概率公式	(81)
3.3 条件概率、全概公式与贝叶斯公式	(83)
3.4 本章的回顾和复习要点	(90)
第四章 随机变量的分布与数字表征	(92)
4.1 引言	(92)
4.2 离散型随机变量及其分布	(92)
4.2.1 二项独立实验序列概型	(94)
4.2.2 离散型随机变量的典型分布	(104)
4.3 连续型随机变量及一般的概率公式	(112)
4.3.1 连续型随机变量的定义及典型分布	(112)
4.3.2 一般的概率公式	(120)
* 4.3.3 随机变量的分布函数	(122)
4.4 随机变量的数字表征	(125)
4.4.1 数学期望	(126)
4.4.2 方差	(129)
4.4.3 有关期望与方差的性质,切比雪夫不等式	(132)
* 4.5 样本期望与方差的收敛性	(135)
4.6 几种分布之间的关系与中心极限定理	(136)
* 4.7 附录:有关期望与方差定理的证明	(140)
4.8 本章的回顾和复习要点	(144)
第五章 假设检验(I)	(146)
5.1 假设检验的基本思想	(146)
5.2 正态单总体的均值检验	(154)
5.2.1 正态单总体、方差未知时对均值的检验	(155)
5.2.2 正态单总体、方差已知时对均值的检验	(164)
5.2.3 大样本条件下,分布未知时对均值的检验	(166)
5.3 均值和概率的区间估计	(169)

5.4	正态双总体的均值检验	(174)
5.4.1	正态双总体、方差已知时对均值的检验	(174)
5.4.2	正态双总体、方差未知而相等时对均值的检验	(178)
5.4.3	正态双总体、方差不等且未知时对均值的检验	(183)
5.5	正态相关两个总体的均值检验	(184)
5.6	若干与检验问题有关的概念	(190)
5.7	附录: 关于 t 分布的概率密度函数及其期望与方差	(198)
5.8	本章的回顾和复习要点	(198)
第六章	假设检验(II)	(200)
6.1	引言	(200)
6.2	方差的检验和区间估计	(200)
6.2.1	正态单总体的方差检验	(200)
6.2.2	独立双总体的方差检验	(205)
6.3	类别的相关性检验——列联表分析	(207)
6.4	分布的 χ^2 拟合优度检验	(215)
6.5	连续型分布的柯尔莫哥洛夫检验	(219)
6.6	双总体的秩和检验	(223)
6.7	附录: 关于 χ_n^2 分布和 $F(k_1, k_2)$ 分布的概率密度函数	(227)
6.8	本章的回顾和复习要点	(227)
第七章	相关分析与回归分析	(229)
7.1	引言	(229)
7.2	联合分布、独立性与相关性的度量	(232)
7.2.1	两个随机变量的联合分布与相关性	(232)
7.2.2	相关系数与样本相关系数	(237)
7.3	线性相关性的统计检验	(240)
7.3.1	关于线性相关性的检验	(240)
7.3.2	相关系数的区间估计及综合性实例	(244)
7.3.3	秩的相关系数及其检验	(249)
7.3.4	相关分析中应注意的若干问题	(252)
7.4	线性回归分析	(259)

7.4.1 线性回归分析	(260)
7.4.2 残差分析	(268)
7.4.3 回归分析的统计检验与区间估计	(271)
7.4.4 案例分析	(275)
7.5 非线性回归分析	(279)
7.5.1 可线性化的回归分析	(279)
7.5.2 一个宏观经济学中的实例	(283)
7.5.3 潜在周期分析	(286)
7.6 本章的回顾和复习要点	(287)
第八章 抽样方法简介	(289)
8.1 引言	(289)
8.2 简单抽样	(290)
8.3 基于均值估计的抽样方法	(291)
8.4 分类抽样与多阶抽样	(295)
8.5 基于假设检验为目的的抽样方法	(299)
8.6 本章的回顾和复习要点	(302)
附录 本书各章配置的练习题	(303)
附表 I 标准正态分布的概率	(345)
附表 II 0~9 的 4000 个随机数	(348)
附表 III 二项概率 $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	(351)
附表 IV 标准正态分布临界值	(356)
附表 V t 分布的临界值	(357)
附表 VI F 分布的临界值	(359)
附表 VII χ^2 分布的临界值	(366)
附表 VIII 秩和检验表	(368)
附表 IX 正态概率纸	(369)
本书常用的数学符号	(370)
内容索引	(373)
参考书籍与文献	(377)

第〇章 开篇引言

本书是一部介绍统计学基本知识和方法的入门性的书籍。通过本书，读者将学到概率论的基础知识和统计学中的常用方法。通过本书，读者将了解到统计学与我们现实社会生活及科学技术的方方面面的关系是非常密切的，如以下问题就是本书关心的典型问题。

1. 王蒙先生关于命运的数学公式的思考

王蒙先生是中国当代著名的作家，他在《王蒙自述：我的人生哲学》(人民文学出版社，2003)一书中，在命运的数学公式一节中谈到在北戴河海滨遇到一个骗人的游戏，即经营游戏者放四种不同颜色的玻璃球在口袋里，每种颜色的球都是5个，然后让人从口袋里摸10个球，并规定了不同出球的比例下的不同的奖惩方法，其详细规则我们以下将给予介绍。王蒙先生冷眼旁观，发现十之八九摸出来的都是3322(以下解释)，此时玩者要罚款5元。他说他回家后用扑克牌或麻将牌也试过，同样是十之八九是3322(十之一二是4321或4330)。

王蒙先生从中思索了良久，他想这就是命运，这就是机会，“**命运是数学的公式和规律，数学就是上帝就是主**”。你想占有一切好运，或者你埋怨一切霉头都降临于你，这就与声称自己总是得到5500(头等大奖)一样，不是完全不可能，但机会极少，几率极低。“想明白了这一点，我们可以少一点怨天尤人，少一点愤愤不平”。王蒙先生还由此联想到“脱离了数学几率的公平公正公开轨道”的一些现行社会中的一些问题。由机会数学而思索人生哲学，从中可看出作者的深刻思想内涵。

我们关心的是上述游戏中的概率论问题：为什么5500“机会极少”而十之八九是3322？首先介绍该经营者的奖惩规则(见王蒙书)：