

普通高等教育“十二五”规划教材

公共基础课系列

统计学

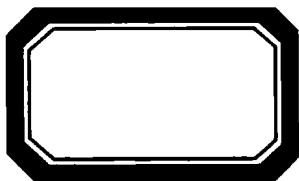
S t a t i s t i c s

王德发 主编

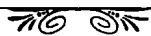


 上海财经大学出版社

普通



规划教材·公共基础课系列



统 计 学

王德发 主编

 上海财经大学出版社

图书在版编目(CIP)数据

统计学/王德发主编. —上海:上海财经大学出版社,2012. 10
(普通高等教育“十二五”规划教材·公共基础课系列)
ISBN 978-7-5642-1488-3/F·1488

I. ①统… II. ①王… III. ①统计学-高等学校-教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2012)第 226861 号

- ☐ 责任编辑 张美芳 吴晓群
- ☐ 封面设计 钱宇辰
- ☐ 投稿热线 apin001@163.com
- ☐ 订购电话 021—65904705
- ☐ 责任校对 廖沛昕 胡 芸

TONG JI XUE 统 计 学

王德发 主编

上海财经大学出版社出版发行
(上海市武东路 321 号乙 邮编 200434)
网 址: <http://www.sufep.com>
电子邮箱: webmaster@sufep.com

全国新华书店经销

上海崇明裕安印刷厂印刷装订
2012 年 10 月第 1 版 2012 年 10 月第 1 次印刷

787mm×1092mm 1/16 22.5 印张 665 千字
印数:0 001—4 000 定价:43.00 元

总 序

近年来,我国的经济、金融领域发展迅猛,规模迅速扩大,创新步伐加快,广度和深度也不断地得到拓展。建立一套内容新颖、结构合理、体系科学、切合实际的经济管理类专业系列本科教材,既是当前经济发展的必然要求,也是培养经济管理类专业人才所必需的。

上海财经大学浙江学院由上海财经大学和浙中教育集团合作举办,是一所按新机制和新模式运作的具有独立法人资格、举办全日制本科学历教育的大学。学院依托上海财经大学在经济管理学科领域的深厚积淀和财经人才培养方面的丰富经验,紧贴长三角、浙中城市群经济和社会发展的需要,培养能够融入国际社会、参与区域竞争和国际竞争的应用型、开拓型、外向型优秀人才。学院将紧密切合社会发展需要和市场需求,拓展交叉学科,发展综合性学科。专业设置以经济学、管理学为主,兼护理学、文学和法学等专业。

上海财经大学浙江学院以全面提高办学质量、办学效益和办学声誉为目标,树立“质量兴院、特色强院”的思想观念,以提高人才培养质量为核心,构建完善的人才培养体系,加强师资队伍建设,强化教学管理,抓好学科建设、专业建设、课程建设,积极引进国内外优质教学资源,提高教学和科研水平,坚持内涵建设和外延发展相协调、规范的制度体系与灵活的办学机制相协调、夯实内部管理基础与外联开放办学相结合。目前,学院主要学科负责人、专业核心课程教师由上海财经大学委派的富有深厚科研能力、良好教学经验的教授担任,贯彻“质量兴院、特色强院”的办学理念,采用先进的教育方法、教育手段,配置优良的教育资源,努力建设一批具有领先水平的特色专业。学院秉承上海财经大学重视基础课程教学的理念建设优质的经济管理类平台课程;另外,与上海财经大学相比,在重视基础理论的同时,更加强调实用性和可操作性。经过几年的摸索,2011年起,学院开始建设符合自己特色的院级精品课程,首批院级精品课程有会计学、基础会计、管理学、统计学、高等数学、计算机基础和英语听力7门。本套教材就是我院精品课建设过程中选取优秀的讲义纳入教材编写系列,由上海财经大学出版社负责出版,双方共同打造的符合上海财经大学浙江学院品牌定位和人才培养目标的系列精品教材。

在这套教材的编写中,希望能够体现以下特点:

◇ 在教材的选择上,主要考虑面向经济管理类本科专业,同时也要考虑其他各类专业的需求,力求选材能够“精”和“新”。

◇ 每本教材的内容选择上,注意广泛收集国内外优秀教材的成果,尤其注意吸收国外较新的优秀教材,力求在完整介绍基本理论、基本内容的基础上,能够介绍一些新的成熟内容,并且强调实用性和可操作性。

◇ 在教材的编写中注重计算机的应用,提高学生运用经济管理理论与方法和计算机技术解决实际问题的能力。在具体操作中,将根据教材的需要选择使用相应的软件。

在本套系列教材的酝酿和编写过程中,自始至终得到上海财经大学浙江学院和浙中教育集团的全力支持。上海财经大学浙江学院理事会理事长应恩民先生一直关心精品课程的建设进展,上海财

经大学浙江学院院长陈晓教授对于精品课程建设和教材编写给予了大力的资助,使得我们的首批教材得以顺利完成。在上海财经大学出版社的热情帮助下,编写大纲和书稿都经过教材编写委员会的多次反复论证、认真讨论。感谢参与论证和编写的各位同行,希望我们辛勤的劳动成果能够得到国内外同行们的认可,获得同学们的欢迎。

王黎明

2012年8月

上海财经大学浙江学院

“统计学”是国家教委 1990 年确定的高等院校财经专业 10 门核心课程之一。为了探索独立学院财经专业统计学课程教学模式与创新能力培养,本课程教学被批准列入 2010 年浙江省新世纪高等教育教学改革项目——改革实践类项目,2011 年被批准列入上海财经大学浙江学院 2011 年度精品课程。本教材是精品课程的建设目标之一,同时也是上海财经大学浙江学院首次组编的系列教材之一。

本教材的编写从以下几个方面进行了尝试:

1. 统计是一面镜子,它摄入的都是现象总体的样本,如何依样本来刻画总体,需要加强统计思想的训练,为此,本教材适时在各章节对相关内容的形成渊源进行了介绍。恐怕这在全国是首例,但效果如何,尚待实践检验。

2. 统计学是一门学科,为适应我国经济、社会、科技和高等教育的发展,我国国务院学位委员会、教育部在 2011 年 2 月公布了新的《学位授予和人才培养学科目录(2011 年)》,其中统计学已被列为一级学科。考虑到我国经济体制和统计体制的改革,同时适应统计学界的主流观点、国际统计科学和统计实践发展的趋势,本教材在内容阐述上淡化了我国传统经济统计学教学中的一些传统概念,如标志、估计标准误差等;加强了有关概念与数理统计学的接轨,如总体的阐述、时间序列的概念与分类等。

3. 统计是一种数据处理的艺术,需要培养学生在某些约束条件下对实际数据进行分类整理,运用一定的分析工具和分析方法,结合自己的实践经验,进行分析和比较,通过讨论甚至辩论从若干可供选择的方案中作出抉择。为此,本教材还将有配套的教材——《统计学学习指导》,内容包括理论知识点、案例、练习、教材思考题解答提示和习题参考答案。此外,由于计算机科学的快速发展,各种统计学应用软件不断涌现,《统计学学习指导》还将主要针对 Microsoft Office 中的 Excel 统计模块的数据处理功能进行介绍,在加强统计学抽象概念直观化的基础上,提高学生的数据处理能力。

本教材共分十一章,以抽样为基础,以统计推断为主线,以社会经济现象为对象,按照由描述统计到推断统计、由简单到复杂、由单变量到多变量、由同质总体到异质总体的顺序,组织各章内容,形成严密的体系。其中,需要说明的是:第五章概率论基础是为适应课程内容体系列在其中,但实际上又把这些章节单独作为一门课程另行开设的,因此在课程中可不予讲述;除了教学大纲规定的内容外,有些章节如第八章假设检验部分中的非参数检验,第九章相关分析和回归分析中的非线性回归、第十章时间序列分析中的时间数列预测方法等,可根据课时进度布置学生自学。总之,不同专业可根据本专业的特点,选择其中相关内容进行重点讲述。

本教材由王德发主编,参编人员及其撰写章节分配如下:王德发(第一章)、洪铁松(第五章、

第六章)、刘小锋(第二章、第三章)、胡志明(第七章、第八章)、奚欢(第九章)、罗平(第十章、第十一章)、詹欢(第四章)。

本教材虽然完稿并交付出版,但还存在许多不完善甚至不足之处,期望得到读者和同行专家们的批评指正,以便在今后的修订中予以完善和调整。

最后,向本教材编写过程中予以支持和帮助的上海财经大学浙江学院的各位领导以及同事们表示诚挚的谢意!

王德发

2012年8月15日于浙江金华

总序/1

前言/1

第一章 绪论/1

第一节 “统计”一词的渊源及其含义/1

第二节 统计学的产生与发展/2

第三节 统计学的性质和特点/11

第四节 统计学的内容和一些基本概念/13

关键术语/16

思考与练习/17

第二章 统计调查/18

第一节 统计调查的概念与分类/18

第二节 我国统计数据调查的组织形式/20

第三节 统计调查误差/26

第四节 统计调查方案/27

关键术语/31

思考与练习/32

第三章 统计数据处理/34

第一节 统计数据的含义及其层次尺度/34

第二节 统计数据处理的概念和内容/40

第三节 统计数据分组/42

第四节 频数分布/49

第五节 统计表与统计图/57

关键术语/60

思考与练习/61

第四章 数据描述性分析/63

第一节 数据总量描述/63

第二节 数据相对量描述/66

第三节 数据分布集中趋势的特征描述/72

第四节 数据分布离散趋势的特征描述/85

第五节 数据分布偏度和峰度的特征描述/90

关键术语/95

思考与练习/95

第五章 概率论基础/99

第一节 随机试验和随机事件/99

第二节 概率/101

第三节 随机变量/106

第四节 随机变量的分布/110

第五节 大数定律和中心极限定理/119

关键术语/123

思考与练习/124

第六章 抽样分布/125

第一节 有关抽样分布的一些基本概念/125

第二节 与正态分布有关几个重要抽样分布/129

第三节 几个重要统计量的分布/131

关键术语/136

思考与练习/137

第七章 参数估计/139

第一节 点估计/139

第二节 区间估计/146

第三节 样本容量的确定/159

关键术语/161

思考与练习/161

第八章 假设检验/164

第一节 假设检验的基本原理/165

第二节 一个总体的假设检验/171

第三节 关于两个总体的假设检验/175

第四节 方差分析/179

第五节 非参数检验/185

关键术语/203

思考与练习/204

第九章 相关分析和回归分析/208

第一节 相关的概念和二元概率分布/209

第二节 一元线性相关/219

第三节 一元线性回归分析/226

第四节 多元线性回归分析/251

第五节 回归分析应注意的问题/261

关键术语/268

思考与练习/269

第十章 时间序列分析/275

第一节 时间序列分析概述/275

第二节 时间序列的数据分析/278

第三节 时间序列的构成分析/286

第四节 时间数列预测/303

关键术语/311

思考与练习/312

第十一章 指数分析/315

第一节 统计指数的概述/315

第二节 综合指数/317

第三节 平均指数/323

第四节 几种常见的经济指数/327

关键术语/335

思考与练习/336

附录 常用统计表/338

附表1 正态分布概率表/338

附表2 t 分布临界值表/340

附表3 χ^2 分布临界值表/341

附表4 F 分布临界值表/342

附表5 二项分布临界值表/344

附表6 秩和检验表/345

附表7 游程检验 R 临界值表/346

附表8 斯皮尔曼等级相关系数 r_s 的上临界值 r_{α} 表/347

附表9 威尔科克森带符号等级检验/348

附表10 曼-惠特尼 U 检验/349

参考文献/350

第一章

绪 论

【引言】 人们经常会面对一些问题,如假定杭州西湖平均水深1米,你敢走到对岸吗?注意:必须“走”到对岸,不能用其他方式。谨慎的人会感到疑惑。答案应该是否定的。因为平均水深1米意味着:有的地方水深不到1米,有的地方水深不止1米。更进一步提问:如果水深在1~1.5米之间的可能性为99%,敢不敢走过去?答案还是否定的。因为水深低于或高于1.5米的可能性达1%。其实,这里涉及统计问题,前者是描述统计问题,后者是推断统计问题。

第一节 “统计”一词的渊源及其含义

“统计”(statistics)一词最早出现于德语“statistik”,由德国马尔堡和哥丁根大学教授阿痕瓦尔(Gottfried Achenwall, 1719~1772)提出,它源于中世界拉丁语的“status”,意思是各种现象的状态和状况。这一语根组成意大利语“stato”,表示“国家”的概念,也含有国家结构和国情知识的意思,后发展为“statista”,本意为“国民”(statesman)。1787年,英国博士齐默尔曼(E. A. W. Zimmerman)在其《欧洲现状的政治调查》(Political Survey of the Present State of Europe)中,根据“statistik”语音译成英语“statistics”,后来在各国广泛传播,最终形成“统计”一词,意指“国家显著事项的比较和记述”或“国势学”,认为统计是关于国家应注意事项的学问。此后,各国相继沿用“统计”这个词,并把这个词译成各国文字,法国译为“statistique”,意大利译为“statistica”,英国译为“statistics”。19世纪后半叶,传到日本,由日本统计学创始人杉亨二根据“statistik”的译音用日语假名直译,表示“政表”、“政算”、“国势”、“形势”等含义,直到1880年在太政官中设立了统计院,才根据意译形成日语汉字“统计”。

我国古代就有“统计”一词,但并非现在的含义,而是“总计、合计”之意。具有现代意义的“统计”一词,是20世纪初随着大批留日学生回国,从日本流传过来的。1903年(清光绪廿九年)由钮永建、林卓南等翻译了杉亨二的学生横山雅南所著的《统计讲义录》一书,才把“统计”这个词从日本传到我国。1907年(清光绪卅三年),彭祖植编写的《统计学》在日本出版,同时在国内发行,这是我国最早的一本“统计学”书籍。当时中文汉字“统计”一词意为:记述国家和社会状况的数量关系的总称。

现代的“统计”一词通常包括统计活动、统计资料和统计科学三种含义。

(1)统计活动是指人们根据一定的目的、采用相应的统计方法收集资料、整理资料和分析资料的工作过程。

现在各国的政府统计,包括经济、社会和科技等方面的统计;联合国统计机构的活动,如世界经济发展和环境统计、各国经济比较等;以至工矿企业或经营服务部门对生产经营过程中的投入和产出所进行的各种计量行为都是统计活动。

(2)统计资料是统计活动的最终成果,如某报报道:“2011年某市实现国内生产总值1 543.63亿元,按可比价格计算,比去年同期增长13.7%,增幅比去年同期高1.6个百分点,比全国平均增长速度约高4个百分点……”

统计资料以较为专业的形式出现的有:联合国统计年鉴、统计月刊,各国出版发行的经济年鉴等。

(3)统计学则是一门指导统计工作的方法论科学,是统计活动的经验总结。

三者关系为:统计学与统计工作之间的关系是理论与实践的关系,统计资料则是统计工作的成果。

第二节 统计学的产生与发展

人类的统计实践随着早期的记数活动而产生,其历史可追溯到远古的原始社会。

一、统计活动的产生与发展^①

统计活动是适应人类社会实践活动的需要而产生和发展起来的。

原始社会末期,随着国家的产生逐渐产生了一种为全社会所需要的、共同一致的、具有某种总体特性的记数活动——统计。人类社会初期的统计活动是简单和松散的,这种活动在各国的史书中都有记载。

例如,古埃及第一王朝时有石碑刻了“〔清查〕西、北、东各州的所有人民”的铭文;罗马帝国早期的历史学家阿毗安(Appianus)所著《罗马史》下卷“内战史”部分,通篇记载了有关军队人数、船队数和俘虏人数的数字;我国《后汉书》中有“禹平水土,还为九州,今禹贡是也。是以其时九州之地凡二千四百三十万八千二十四顷,定耕者九百三十万六千二十四顷,不耕者千五百万二千顷,民口千三百五十五万三千九百二十三人”的记载。

封建社会,统计活动与政治有了密切联系,各国历代封建统治者为了利用统计来加强封建统治、稳定社会秩序,广泛进行了有关人口、军队、世袭领地、财产等方面的统计。

我国春秋战国时期的“上计制度”和“上计律”大约是最原始的统计制度,“计”指包括统计与会计在内的经济核算。按上计制度的规定,郡县长官在年终将下一年度的民户、垦地和税收等预算写在木卷上呈送国君;郡县木卷分左右两部分,右卷留国君、左卷发还。到下一年终检查预算执行情况,以决定赏罚;到我国明代,建立了世界上最早的人口普查制度——“黄册”和“户帖”。

到了资本主义社会,统计活动有了迅速的发展,除了人口、土地和财产等统计,还建立了工业、农业、商业、对外经贸、银行、保险、交通、邮电和海关等专业统计。与此同时,各国普遍建立起专业的统计制度和统计机构,统计活动也演变为一种专业。

虽然我国的统计活动历史悠久,而且建立了较为完善的统计调查制度,但尚无独立而统一的统计机构。直到清末立宪后,才按西方国家统计组织的模式建立统一的统计机构——宪政编查馆统计局;中华民国后期,国民政府实行“超然统计制度”,行政上建立所谓的“联综组织”,即财政制度分为联立综合的四大系统——行政、主计、出纳、审计,各部门相互辅助、相互监督,主计处下设统计局,主持全国的统计工作。中华人民共和国成立后提倡全面学习苏联,建立了计划经济体制的统计机构——国家统计局。现在我国的国家统计局仍是全国政府统计首脑机关,但其机构设计、职能随着我国市场经济的深化以及与国际统计体系接轨的需要而有所改革。

需要说明的是:以上所介绍的主要为有文字记载的政府统计活动,而统计活动并不限于这些。人们的日常生活中经常会涉及统计,只不过没有意识到它的存在,也没有严格地按统计活动的程序行事而已。

^① 王德发:《统计学》,上海财经大学出版社1997年版。

二、统计学的形成与发展现状^①

统计学作为一门系统的科学,距今只有 300 多年的历史,产生于欧洲资本主义社会初期,经历了古典统计学、近代统计学和现代统计学三个时期,先后产生过政治算术学、国势学、社会统计学和数理统计学等重要学派。

(一)古典统计学

17 世纪开始步入成熟的市民社会的欧洲,由于货币经济的发展,先后产生了三门学科,即德国的国势学、英国的政治算术和以法国为主的概率论。从这三门学科的研究内容和方法来看,对今天的统计学有很大影响,因而被视为现代统计的三大来源。

1. 以文字表述来比较各国显著事项的学派——国势学

17 世纪继承了亚里士多德比较叙述的思想、最初孕育于意大利、最终破土发芽于德国的“国势学”,在意大利、荷兰等国有关国家形势研究的基础上逐渐系统化,形成了一门反映国情国力的学科。

1660 年 11 月 20 日,国势学的奠基人海尔曼·康令(Hermann Conving,1606~1681)在德国的赫尔姆斯达德大学讲授“欧洲最近国势学”,分析的对象为西班牙及其殖民地、葡萄牙、英国、法国等,内容涉及社会秩序、立宪、行政、人口、土地、财政、国家组织与机构等,表现形式主要为文字叙述,偶尔也有数据。

有“统计学之父”之誉的阿痕瓦尔曾在哥丁根大学讲授政治统计(国势学)。他继承了康令的观点,代表作是 1749 年出版的《欧洲最主要各国新国势学概要》。

18 世纪后期丹麦的约翰·彼特·安杰逊(Johann Peter Ancherson,1700~1765)于 1741 年编著了《文明国一览表》。1782 年,德国的克罗梅(A. F. W. Crome,1753~1833)受到安杰逊的启发,在其《关于全部欧洲国家的领土和人口》一书中,第一次用图来描述欧洲各国的人口、土地、人口密度、军队、岁入和岁出等项目。由于他们仍是以比较分析各国的政治经济事项为主旨,只是表现手法不同,所以被视为国势学派的支派——图表学派。

2. 以数量分析来研究各国的国情国力的学派——政治算术

政治算术的创始人威廉·配第的代表作为《政治算术》^②与《爱尔兰的政治解剖》。在著作中,配第以一系列分析手段和大量数字描述了英格兰、荷兰、法兰西和爱尔兰等国的经济、军事、政治等方面的情况。



威廉·配第(William Petty,1623~1687)生于英国汉普郡一个毛纺织手工业者家庭。14 岁开始独立谋生,当过水手、家庭教师、海军士兵、医生。1644~1645 年在荷兰莱顿大学攻读医学,后又到法国、英国行医和研究,1649 年获牛津大学医学博士学位,成为医生并兼任皇家医学院教授。1651 年任英国驻爱尔兰军总司令的随从医生。1652 年任爱尔兰总督的私人秘书,后又任爱尔兰土地分配总监。1658 年选为英国议会议员。斯图亚特王朝复辟时期投靠国王查理二世(1630~1685),被封为男爵,并被任命为爱尔兰土地测量总监。1662 年当选为英国皇家学会会员。晚年拥有 10.9 万公顷土地,并经营铁厂、渔场和木材场等企业。马克思对配第的经济思想给予了极高的评价,称他为“现代政治经济学的创始者”、“最有天才和最有创见的经济研究家”,是“政治经济学之父,在某种程度上也可以说是统计学的创始人”。

英国的约翰·格朗特(John Graunt,1620~1674)是政治算术学派的代表人物之一。1662 年出版了《关于死亡率的自然观察和政治观察》,分析了 60 年来伦敦居民死亡的原因及人口变动的关系,首次提出通过大量观察,可以发现新生儿性别比例具有稳定性和不同死因的比例等人口规律;并且第一

① 徐国祥、王德发:《新中国统计思想史》,上海财经大学出版社 1999 年版。

② 杜瑞芝:《数学史辞典》,山东教育出版社 2000 年版。

次编制了“生命表”，对死亡率与人口寿命作了分析，从而引起了普遍的关注。埃德蒙德·哈雷(Edmond Helley, 1656~1742)利用了布雷斯劳(Breslau)的记有死亡年龄的数据，改进了格朗特的生命表并引进了死亡率的定义。瑞士数学家莱昂哈德·欧拉(Leonhard Euler, 1707~1783)提出了平稳生命表的概念。约翰·德威特(John DeWitt, 1625~1672)等人最早讨论退休金和人寿保险的方案。托马斯·罗伯特·马尔萨斯(Thomas Robert Malthus, 1766~1834)、阿尔弗雷德·詹姆斯·洛特克(Alfred James Lotke, 1880~1949)、R. A. 费歇(Ronald Aylmer Fisher, 1890~1962)、威廉·费勒(William Feller, 1906~1970)等人用渐趋复杂的数学来研究生命表的理论，对人类及其总体的统计描述具有显著意义。

随着欧洲资本主义的日趋发展，各国对经济研究的要求日渐迫切，以至政治算术很快传遍欧洲大陆，并形成了不同的学派，其中英国的葛瑞贡·金(Gregong King, 1648~1712)是经济学派的代表者，1696年编著了《关于英国国势自然的政治的观察及其结论》，书中加总计算了1688年英国国民收入总额。查尔斯·戴韦南特(Charles Davenant, 1656~1714)是经济学派的另一名代表人物，他根据“不同国家社会情况相似”的原则，推算和比较了法国与荷兰的贸易往来、土地利用、生产经营、生活方式等方面。

值得说明的是，当时的中国虽然没有直接受到政治算术的影响，也没有进行政治算术研究的课题，但也出现了一些提倡封建变革的思想家，他们的著作与政治算术阐述的内容有异曲同工之妙。其中较有代表性的是清代的魏源(1794~1857)，他曾受林则徐之嘱，根据林则徐主持翻译的《四洲志》，参考了西方的著作，编成了《海国图志》。在60卷著作中，他系统地叙述了世界各国的政治、生产、贸易、金融、军事和历史等情况，其表达方式不仅有文字描述，还有数字说明。

3. 从赌博数学发展起来的概率论

16世纪早期，意大利的卡尔达诺等数学家开始研究骰子中的博弈机会，一位法国侯爵提出在博弈的得点中探求赌金的划分问题。

骰子与赌博游戏

骰子，亦称色子，为正多面体，通常作为桌上游戏的小道具，是古老的赌具之一。骰子也是容易制作和取得的随机数发生器。最常见的骰子是六面骰，它是一颗正立方体，上面分别有一到六个孔(或数字)，其相对两面数字之和必为七。

相传，骰子的发明人是三国时代的文学家曹植，最初用做占卜的工具，后来才演变成后宫嫔妃的游戏，掷骰子点数赌酒或赌丝绸、香袋等物。当时骰子的点穴上涂的是黑色。据清代赵翼考证，红四点最早为唐玄宗所使用。

考古学家曾在出土的两千多年前古埃及坟墓的壁上，发现绘有以羊的后足跟制成的称为astragal之赌具的赌戏。这种骨头有四个面，并不对称，每次投掷会落在四个面之一方。

大约从公元前1200年起，人们摩擦纯天然的骨骼(如羊角脚上的距骨)，使其成为一个粗糙的立方体，再在上面刻数字。

公元前1400年的埃及出现了现在相对面数字之和是7的骰子。

著名的希腊历史学家希罗多德(Herodotus)在其巨著《历史》中写道：早在公元前1500年，埃及人为了忘却饥饿的困扰，经常聚集在一起掷骰子和紫云英，这是一种叫做“猎犬与胡狼”的游戏，按照一定规则，根据掷出各种不同的紫云英而移动筹码。

赌金分配与概率雏形

“赌金分配”(division problem)或称“得点问题”(problem of points)。

公元1494年,意大利数学家帕西奥尼(1445~1509)出版了一本有关算术技术的书,提出在一场赌博中,某一方先胜6局便算赢家,那么,在甲方胜了4局、乙方胜了3局的情况下,因出现意外,赌局被中断,无法继续,此时,赌金应该如何分配?帕西奥尼的答案是:应当按照4:3的比例把赌金分给双方。当时,许多人都认为帕西奥尼的分法不是那么公平合理。因为,已胜了4局的一方只要再胜2局就可以拿走全部的赌金,而另一方则需要胜3局,并且至少有2局必须连胜,这样要困难得多。但是,人们又找不到更好的解决方法。在这以后100多年中,先后有多位数学家研究过这个问题,但均未得到过正确的答案。

半个世纪以后,曾以发表三次方程的求解公式而闻名于世的意大利著名数学家卡当诺(Cardano, 1501~1576)讨论了一个类似的问题:需要分析的不是已经赌过的次数,而是剩下的次数。在帕西奥尼的问题里,胜了5局的赌徒只要再赢一局,便可以结束整场赌博。所以,假如比赛不中断的话,有五种可能,即他第一局胜,第二局胜,第三局胜,第四局胜或者所有4局都输掉。卡当认为,总赌金应按照 $(1+2+3+4):1=10:1$ 的比例来分配(其实,该结果是错的,正确的答案应该是15:1)。

到了17世纪,为了解答这个问题,法国贵族梅莱(Mere)不厌其烦地投掷骰子,并把结果记在本子上,试图找出答案。公元1651年夏天,梅莱在旅途中偶遇法国数学家、物理学家、哲学家帕斯卡,他向帕斯卡请教一个亲身经历的“赌金分配”问题,以消磨旅途时光。梅莱的问题是:一次梅莱和赌友掷骰子,各押赌注32个金币,梅莱若先掷出三次“6点”,或赌友先掷出三次“4点”,就算赢了对方。赌博进行了一段时间,梅莱已掷出了两次“6点”,赌友也掷出了一次“4点”。这时,梅莱奉命要立即去晋见国王,赌博只好中断。那么,两人应该怎样分这64个金币的赌金呢?

赌友说,梅莱要再掷一次“6点”才算赢,而他自己若能掷出两次“4点”也就赢了。这样,自己所得应该是梅莱的一半,即得64个金币的 $1/3$,而梅莱得 $2/3$ 。梅莱争辩说,即使下一次赌友掷出了“4点”,两人也是秋色平分,各自收回32个金币,何况那一次自己还有一半的可能得16个金币呢?所以他主张自己应得全部赌金的 $3/4$,赌友只能得 $1/4$ 。

梅莱的问题居然把帕斯卡给难住了。他为此苦苦想了三年,终于在1654年悟出了一点道理。于是他把自己的想法写信告诉他的好友,当时号称数坛“怪杰”的费马,两人对此展开热烈的讨论。后来,荷兰数学家惠更斯(C. Huygens, 1629~1695)也加入了他们的探讨行列。最后,他们一致认为,梅莱的分法是对的!惠更斯还把他们讨论的结果载入1657年出版的《论赌博中的计算》。这本书至今被公认为概率论的第一部著述。

梅莱的分法为什么是对的?帕斯卡和费马又是怎么想的?这些疑问随着本书下面章节逐步展开,便能迎刃而解。

事实上,按卡当诺的想法,在中断赌博之后所设想的4局比赛中,每局都有胜负两种可能,总共有 $2 \times 2 \times 2 \times 2 = 16$ 种可能。其中只有最后一种,即第一个赌徒4局全负时,第二个赌徒才可能得到赢,而其余15种情况都是输。因此,他们的赌金分配比例应当是15:1。

赌金风波终于以古典概率论雏形的诞生而宣告平息。

17世纪,法国的数学家巴里瑟·帕斯卡(Balise Pascal, 1623~1662)和法国律师、业余数学家皮埃尔·德·费马(Pierre de Fermat, 1601~1665)研究了意大利帕乔里的著作《摘要》,建立了通信联系。在他们的相互通信以及著作中,费马引入了随机变量和数学期望的概念,建立了概率论的基本原则,为概率的数学模型——有限概率空间的抽象奠定了博弈基础,尽管这种总结直至1933年才由柯尔莫戈罗夫给出。1665年,帕斯卡在其遗著《数学三角形》中提出了上述得点问题的一般解法,并在此基础上发展了组合理论和“帕斯卡三角形”,为计算二项式系数奠定了基础。

1657年,荷兰物理学家、天文学家、数学家惠更斯发表的《论赌博中的计算》是概率论最早的论著,显示了他在数学上的造诣,也被公认为概率论的创始人之一。

帕斯卡、费马和惠更斯三人在通信讨论中,都涉及了数学期望(mathematical expectation)的概念,奠定了古典概率论的基础。

瑞士巴塞尔大学教授雅各布·伯努利(Jakob Bernoulli, 1654~1705)在他去世后的1713年发表的遗著《猜度术》中给出了概率论发展中极为重要的显示二项试验结果的定理,构成概率论中的第一个极限定理——“伯努利大数定理”,即“在多次重复试验中,频率有越趋稳定的趋势”。

1733年,法国数学家亚伯拉罕·棣莫弗^①(Abraham de Moivre, 1667~1754)在其出版的著作《分析杂论》中,用正态分布估计大量抛掷硬币出现正面的次数,导出了对二项分布的一个近似,使每一个概率都等于正态曲线下的一块面积,进而形成棣莫弗的中心极限定理——概率论中第二个基本极限定理的原始雏形。这个超越时代的成果险些被历史遗忘,所幸著名法国数学家皮埃尔·西蒙·拉普拉斯(Pierre Simon Laplace, 1749~1827)在1812年发表的巨著《概率的分析理论》中拯救了这个默默无闻的理论。在这本书中,拉普拉斯扩展了棣莫弗的理论,明确地给出了概率的古典定义,指出二项分布可用正态分布逼近,形成著名的“棣莫弗—拉普拉斯定理”。但同棣莫弗一样,拉普拉斯的发现在当时并未引起很大反响。直到19世纪末,中心极限定理的重要性才被世人所知。

此外,拉普拉斯还和数位数学家完善了正态分布理论,建立了最小二乘法的理论。

1837年,法国数学家西蒙·丹尼斯·泊松(Simeon Denis Poisson, 1781~1840)在《关于判断的概率之研究》一文中推广了伯努利形式下的大数定律,研究得出了一种新的描述随机现象的一种常用分布——泊松分布。

1901年,俄国数学家李雅普诺夫(Lyapunov, 1857~1918)用更普通的随机变量定义中心极限定理并在数学上进行了精确的证明,第一次科学地解释了为什么实际中遇到的许多随机变量近似服从于正态分布。如今,中心极限定理被非正式地认为是概率论中的首席定理。

苏联数学家安德列·柯尔莫哥洛夫^②(A. N. Kolmogorov, 1903~1987)在概率论发展史上作出了重大贡献,是现代概率论、概率论公理化系统的创立者,在随机过程论和数理统计领域有众多开创性贡献。1928年,他得到了随机变量序列服从大数定律的充要条件;1929年,得到了独立同分布随机变量序列的重对数律;1930年,得到了强大数定律的非常一般的充分条件;1931年,发表了《概率论的解析方法》一文,奠定了马尔可夫过程^③论的基础;1932年,得到了含二阶矩的随机变量具有无穷可分分布律的充要条件;1934年,出版了《概率论基本概念》一书,在世界上首次以测度论和积分论为基础建立了概率论公理化系统,这是一部具有划时代意义的巨著,在科学史上写下了苏联数学最光辉的一页;1935年,提出了可逆对称马尔可夫过程的概念及其特征所服从的充要条件,这种过程成为统计物理、排队网络、模拟退火、人工神经网络、蛋白质结构的重要模型;1936~1937年,给出了可数状态马尔可夫链状态分布;1939年,定义并得到了经验分布与理论分布最大偏差的统计量及其分布函数;20世纪30~40年代,他和辛钦一起发展了马尔可夫过程和平稳随机过程论,并应用于大炮自动控制和工农业生产中,在卫国战争中立了功;1941年,他得到了平稳随机过程的预测和内插公式;1955~1956年,他和他的学生苏联数学家普洛考浩洛夫(Y. V. Prokhorov)开创了取值于函数空间上概率测度的弱极限理论,这个理论和苏联数学家司考若郝德(A. B. Skorohod)引入的D空间理论是弱极限理论的划时代成果。

到了近代,出现了理论概率及应用概率的分支,即将概率论应用到不同范畴,从而开展了不同学

① 陈希孺:《数理统计学简史》,湖南教育出版社2002年版。

② 百度百科 baike. baidu. com/view/65843. htm, 2012-4-30。

③ 马尔可夫过程在物理、化学、生物、工程技术和经济管理等领域中有十分广泛的应用,仍然是当今世界数学研究的热点和重点之一。

科。因此,现代概率论已经成为一个非常庞大的数学分支。

(二)近代统计学

18 世纪上半叶,比利时统计学家、数学家和天文学家凯特勒(Lambert Adolphe Jacques Quetelet, 1796~1874),把德国的国势学、英国的政治算术以及意大利、法国的古典概率论结合起来,开创了近代统计学的新纪元。

之后,由于政治、经济等历史等条件的不同,在德国以凯特勒的“统计基本理论”为核心形成了德国社会统计学,在英国则以凯特勒的“论数学与大量观察的关系”为基础、以概率为核心形成了数理统计学。

1. 社会统计学

社会统计学^①于 18 世纪形成于德国,是以国势学的研究目的、政治算术的研究对象和方法、凯特勒的基本统计理论为核心而形成的。后来扩展到美、日等国。由于对某些理论问题的看法不同,这一学派内部产生了分歧,形成“新”、“旧”两大社会统计学派。

“旧派”的观点认为“统计学是一门以大量观察为基础的研究社会现象规律的独立的社会科学”,而“新派”除了同意“统计是采用大量观察法研究社会经济现象的数量方面”之外,还认为“统计学是一门方法论的科学”,并主张“采用抽样法”。

社会统计学由德国弗赖堡大学和海德堡大学教授克尼斯(Karl Gustav Adolf Knies, 1821~1898)首创;德国统计学家、统计教育家恩格尔(Christian Lorenz Ernst Engel, 1821~1896)为形成社会统计学作出了重要贡献,著名的反映居民生活水平高低的恩格尔法则就是他提出并以他的名字命名的;社会统计学的完成人物是德国统计学家梅尔(Georg Von Mayr, 1841~1925),他最终确定了社会统计学的研究范围、性质和研究方法。

2. 数理统计学

数理统计学在内容上分为描述统计(descriptive statistics)和推断统计(inferential statistics)。前者的重点在于研究数据的收集、整理和对数据进行直观的描述,后者则是由部分数据推测总体所进行的统计处理。

(1)描述统计。描述统计是在 19 世纪 20 年代以前发展起来的。“描述”是当时一些研究生物进化的学者在长期的研究实践中提炼出来的,因此他们又被称为生物统计学派。其创始人是弗朗西斯·高尔顿(Francis Galton, 1822~1911),他通过对大量数据的描述和比较,出色地引入中位数、百分位数、四分位数、四分位差和分布、相关、回归等重要概念和方法。

(2)推断统计。贝叶斯(T. Bayes, 1702~1761)的论文《论机会学说问题的求解》可视为最早的一种统计推断程序。拉普拉斯和高斯等利用贝叶斯公式估计参数的研究,促使统计学摆脱观测数据的单纯描述而向强调推断的阶段过渡。

数理统计学从其发展的时间进程看,可分为三个时代:卡尔·皮尔逊统计时代、费歇统计时代和奈曼—皮尔逊统计时代^②。但随着 1946 年瑞典统计学家克拉默尔发表了《统计学的数学方法》一书,用测度论系统总结了数理统计的发展,标志着统计学开始由近代转向现代。因此,上述三个时代的后两个时代可归入现代统计学范畴。

^① “社会统计学”一词出现在那普于 1871 年在《经济学·经济学报》第 17 卷“文献栏”内发表的《关于凯特勒社会统计学和人类学著作的报告》中。19 世纪 80 年代,“社会统计学”一词被广泛应用。1898 年,凯尔(A. N. Kiaer)和汉斯(E. Hansseu)以此为名合著了《社会统计学》一书。而“数理统计学”一词早在 1867 年由德国数学家威蒂斯坦(T. Wittstciu)开始使用。从统计学的形成与发展来看,早期的统计学都是以社会现象为研究对象的,没有必要在“统计学”前面加上“社会”两字。因此,“统计学”已经约定成俗、相沿袭用,成为一个专用名词,而在“数理统计学”一词出现以后,为了与之相区别,才在原有包括政治统计、人口统计、经济统计、犯罪统计、伦理统计或道德统计在内的“统计学”前面加上“社会”两字(高庆丰:《欧美统计学史》,中国统计出版社 1987 年版,第 110 页)。

^② H. O. Lancaster 撰文,吴喜之译:《统计学简史》,摘自 Encyclopedia of Statistical Science, Wiley, 2005。