全国医药职业教育药学类规划教材
QUANGUO YIYAO ZHIYE JIAOYU YAOXUELEI GUIHUA JIAOCAI

（供高职高专使用）

医药数理统计

YIYAO SHULITONGJI

主编 高祖新

中国医药科技出版社

全国医药职业教育药学类规划教材

医药数理统计

（供高职高专使用）

主　编　高祖新
副主编　王小平　徐　伟
编　者　（以姓氏笔画为序）
王小平（中国药科大学高等职业技术学院）
周凤莲（山西生物应用职业技术学院）
徐　宁（山东食品药品职业学院）
徐　伟（沈阳药科大学高等职业技术学院）
高祖新（中国药科大学）

中国医药科技出版社

内 容 提 要

本书是全国医药职业教育药学类规划教材之一，依照教育部［2006］16号文件要求，结合我国高职教育的发展特点，根据《医药数理统计》教学大纲的基本要求和课程特点编写而成。

全书共分为8章，分别介绍了概率论基础；医药应用领域的数据处理与图表展示；数理统计的基本原理、基本概念和基本知识；常用统计推断和统计分析方法、统计软件的实际操作应用等内容。相对其他数理统计教科书，内容切合高职教学的实际，结构体系也更合理完善。

本书适合医药高职教育及专科、函授和自学考试等相同层次不同办学形式教学使用，也可作为医药行业职工培训和自学用书。

图书在版编目（CIP）数据

医药数理统计/高祖新主编．—北京：中国医药科技出版社，2008.6

全国医药职业教育药学类规划教材

ISBN 978-7-5067-3911-5

Ⅰ．医… Ⅱ．高… Ⅲ．数理统计—应用—医药学—高等学校：技术学校—教材 Ⅳ．R311

中国版本图书馆CIP数据核字（2008）第074092号

美术编辑 陈君杞

版式设计 郭小平

出版 中国医药科技出版社

地址 北京市海淀区文慧园北路甲22号

邮编 100082

电话 责编：010-62216635 发行：010-62227427

网址 www.cspyp.cn

规格 787×1092mm 1/16

印张 13¼

字数 298千字

印数 1—3000

版次 2008年6月第1版

印次 2008年6月第1次印刷

印刷 北京通州皇家印刷厂

经销 全国各地新华书店

书号 ISBN 978-7-5067-3911-5

定价 23.00元

前　言

医药数理统计是应用概率论与数理统计原理，对医药、生物等相关领域的数据资料进行搜集整理、分析和解释，以显示其统计规律性的应用科学，其知识和方法的掌握，对于医药高职学生提高其基本科学素质，进一步学习其他专业基础课及专业课程具有非常重要的作用，也为其今后从事相关的工作和科研提供了重要的统计应用工具。

《医药数理统计》作为全国医药职业教育药学类规划教材之一，其编写本着"基础理论适度够用、统计主体重点保证、医药应用背景突出、能力培养综合全面"的指导原则，广泛吸收了国内外教学研究的优秀成果，总结了编者多年的统计教学实践经验，在保持数理统计学科系统性前提下，强调统计理论与医药实际的结合，反映与时俱进的时代特征，强化医药高职学生的针对性和统计软件应用的实用性，突出高职学生的统计应用能力的培养，体现"学以致用"的教学目的性。

总体而言，本书具有以下主要特点。

1. 按照医药高职学生的培养目标和要求，适当选取教材内容的深度和广度。所选概率基础理论精练严谨，必需够用；统计知识方法充实全面，重点讲透；医药专业应用结合实例，贯通全书。教材内容涵盖统计应用所必需的概率论基础；医药应用领域的数据处理与图表展示；数理统计的基本原理、基本概念和基本知识；常用统计推断和统计分析方法；统计软件的实际操作应用等。教学内容更加切合高职教学的实际，结构体系也更为合理完善。

2. 针对医药高职学生的专业特点和学习基础，注重统计方法思想和实际医药应用的阐述，并结合医药应用实例和背景，说明数理统计方法的特点、应用条件和场合等，同时增加非常实用的统计软件应用操作的技能训练。编写内容系统全面，重点突出，深入浅出；理论阐述严谨清晰，简明扼要；举例典型，生动实用，解析透彻；叙述解释简明流畅，通俗易懂；课后思考与练习题，题型多样，富于启发性，并配有参考答案，便于教学和提高学生对知识的掌握水平。

3. 突出医药高职教学的应用性、可操作性。以往的医药统计课程的教学一般重视理论和方法，而忽视了统计实际应用技能的培养。本教材重点增加了计算机应用的统计技能的训练，根据医药高职学生计算机学习的基础，本教材选用了计算机应用中最为普及的Excel软件的统计模块来进行统计软件应

用的教学，在最后一章给出Excel软件对应概率统计功能操作与应用的具体指导，同时辅之以上机训练题，以培养学生运用Excel软件进行医药统计分析、解决医药统计实际问题的能力，达到“学以致用”的目的。

本书由高祖新主编，王小平、徐伟任副主编，并由高祖新负责全书的整体设计、编写大纲的拟订和全书统稿、审定工作，王小平协助进行全书的修订统稿工作，各章节的负责编委见各章结尾处。本书编著时注意博采众长，参考了国内外多种教材和参考文献，并得到中国医药科技出版社及编委所在单位的大力支持，在此一并表示衷心的感谢。

尽管我们为提高本书的质量倾注了不少努力，但由于编写时间和学识水平所限，书中疏漏和不妥之处在所难免，恳请各位专家、读者批评指正，以便今后修正完善。

编　者

2008年4月

目　录

绪论 ……………………………………………………………………………（1）

一、统计学及其发展史 ………………………………………………………（1）

二、常用统计软件的应用 ……………………………………………………（2）

第一章　概率与分布 …………………………………………………………（4）

第一节　随机事件和概率 ……………………………………………………（4）

一、随机事件 …………………………………………………………………（4）

二、事件间的关系和运算 ……………………………………………………（5）

三、概率的定义 ………………………………………………………………（7）

四、概率的加法公式 …………………………………………………………（10）

五、条件概率和事件的独立性 ………………………………………………（11）

第二节　随机变量及其分布 …………………………………………………（13）

一、随机变量 …………………………………………………………………（13）

二、离散型随机变量及其分布 ………………………………………………（13）

三、连续型随机变量及其分布 ………………………………………………（14）

四、随机变量的数字特征 ……………………………………………………（15）

五、常见的概率分布 …………………………………………………………（19）

自测思考题一 …………………………………………………………………（25）

习题一 …………………………………………………………………………（26）

第二章　数据的整理与统计描述 ……………………………………………（28）

第一节　数据的分类和整理 …………………………………………………（28）

一、数据的分类 ………………………………………………………………（28）

二、数据资料的统计整理 ……………………………………………………（30）

第二节　数据分布的统计特征描述 …………………………………………（34）

一、数据分布集中趋势的统计描述 …………………………………………（34）

二、数据分布离散程度的统计描述 …………………………………………（37）

第三节　统计图表 ……………………………………………………………（40）

一、统计图 ……………………………………………………………………（40）

二、统计表 ……………………………………………………………………（44）

自测思考题二 …………………………………………………………………（45）

习题二 …………………………………………………………………………（46）

第三章 参数估计 ……（48）

第一节 统计量 ……（48）

一、总体与样本 ……（48）

二、统计量 ……（49）

第二节 统计推断的常用分布 ……（50）

一、样本均值的分布 ……（50）

二、χ^2 分布 ……（50）

三、t 分布 ……（52）

四、F 分布 ……（53）

第三节 参数的点估计 ……（54）

一、参数的点估计 ……（54）

二、估计量的优良性 ……（55）

第四节 参数的区间估计 ……（56）

一、区间估计的概念 ……（56）

二、正态总体均值的区间估计 ……（57）

三、正态总体方差的区间估计 ……（59）

四、总体率的区间估计 ……（60）

自测思考题三 ……（63）

习题三 ……（63）

第四章 假设检验 ……（65）

第一节 假设检验概述 ……（65）

一、假设检验的基本思路 ……（65）

二、假设检验的一般步骤 ……（67）

三、假设检验的两类错误 ……（67）

四、假设检验中的 P 值法 ……（68）

第二节 单个总体参数的假设检验 ……（68）

一、正态总体的均值检验 ……（69）

二、正态总体的方差检验 ……（71）

三、总体率的检验 ……（72）

四、假设检验中的单侧检验 ……（73）

第三节 两个总体参数的假设检验 ……（75）

一、成组设计中的均值检验 ……（75）

二、配对设计中的均值检验 ……（78）

三、成组设计中的方差检验 ……（80）

自测思考题四 ……（82）

习题四 ……（82）

第五章 方差分析 ······ (84)
第一节 单因素方差分析 ······ (84)
一、方差分析的基本概念 ······ (84)
二、方差分析的原理与方差分析表 ······ (85)
三、方差分析的解题步骤 ······ (89)
四、单因素方差分析应用举例 ······ (89)
第二节 双因素方差分析 ······ (90)
一、无交互作用的双因素方差分析 ······ (90)
二、双因素方差分析问题举例 ······ (93)
自测思考题五 ······ (94)
习题五 ······ (95)
第六章 相关分析与回归分析 ······ (97)
第一节 相关分析 ······ (97)
一、相关关系 ······ (97)
二、相关分析 ······ (98)
三、相关系数的显著性检验 ······ (99)
第二节 回归分析 ······ (101)
一、一元线性回归 ······ (102)
二、回归方程的显著性检验 ······ (104)
三、利用回归方程进行预测 ······ (107)
自测思考题六 ······ (109)
习题六 ······ (109)
第七章 正交试验设计与分析 ······ (112)
第一节 正交表与正交设计 ······ (112)
一、正交表与正交设计 ······ (112)
二、正交设计的基本步骤 ······ (113)
第二节 正交试验的直观分析 ······ (114)
一、表头设计 ······ (114)
二、直观分析法的步骤 ······ (116)
第三节 正交试验的方差分析 ······ (117)
自测思考题七 ······ (121)
习题七 ······ (122)
第八章 数据处理与统计分析的 Excel 应用 ······ (123)
第一节 数据整理与统计作图的 Excel 应用 ······ (124)

一、用 Excel 进行统计作图 ……(124)
二、用 Excel 生成频数分布表与直方图 ……(127)
三、用 Excel 计算常用统计量 ……(130)
第二节 常用分布概率计算的 Excel 应用 ……(132)
一、用 Excel 计算二项分布 ……(132)
二、用 Excel 计算正态分布 ……(134)
三、用 Excel 计算 χ^2 分布 ……(136)
四、用 Excel 计算 t 分布 ……(138)
第三节 总体参数置信区间的 Excel 应用 ……(140)
一、用 Excel 求 σ^2 已知时总体均值的置信区间 ……(140)
二、用 Excel 求 σ^2 未知时总体均值的置信区间 ……(141)
第四节 假设检验的 Excel 应用 ……(143)
一、用 Excel 进行单个正态总体的参数检验 ……(143)
二、用 Excel 进行两个正态总体的参数检验 ……(147)
第五节 方差分析的 Excel 应用 ……(150)
一、用 Excel 进行单因素方差分析 ……(150)
二、用 Excel 进行双因素方差分析 ……(152)
第六节 相关分析与回归分析的 Excel 应用 ……(155)
一、用 Excel 制作散点图 ……(156)
二、用 Excel 计算相关系数 ……(159)
三、用 Excel 进行一元线性回归分析 ……(160)
上机训练题 ……(162)

附录：常用统计表 ……(165)

附表 1 二项分布表 ……(165)
附表 2 泊松分布表 ……(168)
附表 3 标准正态分布表 ……(170)
附表 4 标准正态分布的双侧分位数表 ……(172)
附表 5 χ^2 分布表 ……(173)
附表 6 t 分布表 ……(175)
附表 7 F 分布表 ……(177)
附表 8 二项分布参数 p 的置信区间表 ……(181)
附表 9 检验相关显著性的临界值表 ……(185)
附表 10 正交表 ……(186)

习题参考答案 ……(192)

中英文索引 ……(199)

参考文献 ……(202)

绪　论

随着社会的发展，21 世纪的今天，我们已进入了知识经济时代。知识经济时代也是信息经济时代，数据资料作为信息的主要载体，在我们社会生产和科学研究的各个领域中正起着越来越重要的作用。而在我们所从事的医药研究和生产中，无论是疾病防治、药物研发、临床试验、公共卫生等各领域，还是新药研制、药物鉴定、药理分析、试验设计、药政管理、处方筛选、医药信息等医药领域的各个方面，都需要进行大量的数据资料的整理和分析。**医药数理统计**正是应用概率论与数理统计的原理和方法，对医药、生物等相关领域研究对象的数据资料信息进行搜集、整理、分析和解释，以显示其总体特征和统计规律性的应用科学。其中**概率论**（probability）是从数量侧面来研究随机现象统计规律性的数学学科，而**数理统计**（mathematical statistics）则是以概率论为基础，通过对随机现象观察数据的收集整理和分析推断来研究其统计规律的学科。

一、统计学及其发展史

在日常生活中，统计既可以指统计数据的搜集活动，即统计工作；也可以指统计活动的结果，即统计数据；还可指分析统计数据的方法和技术，即统计学。**统计学**（statistics）是对研究对象的数据资料进行搜集、整理、分析和解释，以显示其总体特征和统计规律性的科学。

统计实践作为一种社会实践活动已有四五千年的历史，早在人类社会的初期——还没有文字的原始社会，就有了“结绳记事”等统计计数活动；在我国公元前二千多年的夏朝就有了人口和土地的统计数字记载。此后，随着社会生产力的发展，统计实践的内容、规模和范围越来越大。但是，将统计实践上升到理论，使之成为一门系统科学——统计学，距今只有 300 多年的历史。

最初的统计方法是随着社会政治和经济的需要而逐步得到发展的，直到 18 世纪概率论被引进之后，统计才逐渐形成一门成熟的科学。17 世纪中叶，法国数学家帕斯卡（B. Pascal，1623 - 1662）和费马（P. Fermat，1601 - 1665）等对赌徒 Méré 提出的赌局问题的解决，开创了概率论研究的新纪元。1662 年格朗特（J. Graunt，1620 - 1674）基于伦敦死亡人数资料的研究所进行的死亡率推算，是历史上最早出现的统计推断。他在其代表作《关于死亡表的自然的和政治的观察》（1662 年）一书中，还通过大量观察的方法，研究并发现了一系列人口统计规律，如男性的死亡率高于女性，男婴和女婴的出生性别比大约为 14:13 等，并运用各种方法对统计数据进行间接的推算和印证。而最早将古典概率论引进统计学领域的是法国天文学家、数学家拉普拉斯（P. S. Laplace，1749 - 1827），他提出了研究随机现象的分析方法，完善了古典概率论的结构，并阐明了统计学大数法则，进行了大样本推断的尝试。德国数学家高斯（F. Gauss，1777 - 1855）发现了正态分布方程，他还成功地将正态分布理论用于描述观察误差的分布，并用于行星轨迹的预测。比利

时统计学家凯特勒（A. Quetelet，1796－1874）发现了大量随机现象的统计规律性，并开创性地应用了许多统计方法，完成了统计学和概率论的结合，出版了《概率论书简》、《统计学的研究》、《社会物理学》等一系列统计学重要著作，被认为是数理统计学的创始人。此后，以概率论为基础的统计理论和方法被称数理统计。

从19世纪中叶到20世纪中叶，数理统计和应用统计得到蓬勃发展并达到成熟。德国的大地测量学者赫尔梅特（F. Helmert，1843－1917）在1876年研究正态总体的样本方差时，发现了χ^2分布（卡方分布）。英国生物学家、人类学家高尔顿（F. Galton，1822－1911）将正态分布理论用于社会学方面的研究，并在生物遗传学中提出了著名的回归、相关等概念，创立了回归分析法。法国医生路易斯（P. C. A. Louis，1787－1872）研究了当时流行的用“放血”疗法治疗伤寒和肺炎效果，1835年提出了医学观察中的抽样误差和混杂概念、临床疗效对比的前瞻性原则和疗效比较的“数量化”方法，被誉为“临床统计之父”。他的学生盖瓦勒特（J. Gavarret，1808－1890）1840年在巴黎出版了世界上第一部医药统计教科书——《医学统计学》。数理统计学的奠基人之一、英国数学物理学家、统计学家皮尔逊（K. Pearson，1857－1936）进一步发展了回归与相关的理论，提出了总体、标准差、正态曲线等重要术语和矩估计法、χ^2拟合优度检验法，并创建了生物统计学，为20世纪数理统计和生物统计学的发展奠定了基础。英国统计学家戈塞特（W. S. Gosset，1876－1937）在1908年以“Student”为笔名在《生物计量学》杂志上发表了论文“平均数的规律误差”，首先提出了t统计量的精确分布——t分布，开创了小样本统计理论的先河，使统计学进入了以推断统计学为主流的现代统计学时期。而英国统计学派的代表人物费歇尔（R. Fisher，1890－1962）系统地发展了抽样分布理论，建立了以最大似然估计法为中心的点估计理论，首创了试验设计法并提出方差分析法，所发表的论文《理论统计学的数学基础》（1921年）和《点估计理论》（1925年），奠定了统计学沿用至今的数学框架，被誉为现代数理统计学的奠基人之一。其后美国统计学家奈曼（J. Neyman，1894－1981）和小皮尔逊（E. Pearson，1895－1980，K. Pearson之子）合作，20世纪30年代提出了似然比检验，并建立了置信区间理论，在数学上完善了假设检验和区间估计的理论体系。美国统计学家沃尔德（A. Wald，1902－1950）所建立的序贯分析和统计决策理论，美国统计学家威尔克斯（S. Wilks，1906－1964）所创立的多元方差分析、多项式分布、多变量容许区间等一系列多元分析方法，开创了数理统计学的新局面。

随着自然科学和社会经济的进步和发展，数理统计在理论上不断成熟与完善，应用上日益广泛和深入。数理统计也成为研究自然现象和社会经济现象数量方面的极为用力的工具，并逐步渗透到各个学科领域，形成了许多边缘学科，如：信息论、决策论、排队论、可靠性理论、自动控制、统计质量管理、生物统计、医药统计、社会统计、水文统计、统计物理学、计量经济学、计量心理学等，成为现代科学发展的一个重要标志。

二、常用统计软件的应用

随着电子计算机的应用和普及，特别是计算机统计软件的深入发展，人们的数据处理能力大为增强，以往受计算能力限制的数理统计有关理论和方法，其处理实际问题的能力也得到了空前提高。统计软件是利用计算机软件技术呈现统计数据，进行数据分析，模拟

和实现统计过程的一类专业应用软件，是统计方法应用的重要载体，在医药统计数据处理和统计分析中具有日益重要的地位。

在实际处理时，尤其是对于数据量较大的实际问题，一般通过计算机利用有关统计软件进行有关数据整理、统计图表显示和统计分析等工作。目前常用的统计软件主要有 SAS（统计分析系统）、SPSS（社会科学统计软件）、Excel（电子表格）等。

（一）SAS（统计分析系统）

SAS 系统，全称 Statistical Analysis System（统计分析系统），是模块化、集成化的应用软件系统，具有完备的数据管理、数据分析、数据存取、数据显示等功能，除统计分析外还有制图、矩阵运算、运筹规划、质量控制和医药临床研究等功能，为医药研究、经济管理、社会科学、自然科学等各领域的众多用户所采用，是当前最流行的国际标准通用的统计分析软件，但其操作略为繁琐。

（二）SPSS（社会科学统计软件）

SPSS，全称 Statistical Package for Social Science（社会科学统计软件），是集数据整理、分析功能于一身的组合式软件包，以其强大的统计分析功能、方便易用的用户操作方式、灵活的表格分析报告和精美的图形展现形式，与 SAS 同为当前世界上最为流行的应用最广泛的专业统计分析软件，不仅应用于社会科学领域，而且广泛应用于商务经济、医药卫生、政府部门、教学科研和自然科学研究等各个领域。

（三）Excel（电子表格软件）

Excel 作为 Microsoft Office 办公软件包的最重要的组件之一，是一个功能强大且使用简便的电子表格软件。它不仅具有强大的制表和绘图功能，而且内置了数学、统计、财务等十类 300 多种函数，同时还提供数据分析、规划求解、方案管理器等多种分析方法和工具，可进行各种数据处理、基本统计分析、数学计算和辅助决策操作等。

由于 Excel 软件普及程度高，操作运算也较为简便，本书第八章主要介绍 Excel 软件的统计分析与运算处理操作，以提高和拓展数据处理和统计分析的应用能力。

目前，医药数理统计的理论方法及应用已广泛渗透到医药研究与实践的各个领域，成为进行医药科学研究的重要前提和手段。有关医药数理统计的知识、方法和必要的统计软件应用技能训练，也已成为每个医药科技工作者必不可少的专门知识和技能，其学习和掌握对于有效而正确地利用数据资料进行医药领域的研究和实践具有极为重要的意义。

（高祖新）

第一章 概率与分布

当我们兴冲冲地外出旅行，天公不作美，却下起了大雨；当我们购买彩票时，总希望自己中大奖，但长时间与大奖无缘。在现实生活中，确实有很多这类现象，其结果具有不确定性，我们称之为**随机现象**（random phenomena）。

概率论作为从数量侧面来研究随机现象统计规律性的数学学科，为解决这种不确定性问题提供了有效的方法。在本章中，我们将学习如何用概率来度量不确定性，并介绍概率、随机变量及其分布等的概率论基本知识，从而为以后几章学习统计推断等数理统计基本理论和统计分析方法奠定基础。

第一节 随机事件和概率

一、随机事件

为了研究随机现象的统计规律性，我们把各种科学实验或观测等统称为**试验**（experiment）。如果试验具有下列特点：

（1）试验可以在相同的条件下重复进行；

（2）每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果；

（3）进行一次试验之前不能确定哪一个结果会出现。

则称这种试验为**随机试验**（random experiment），简称试验。

随机试验的所有可能结果组成的集合称为**样本空间**（sample space），记为 $\boldsymbol{\Omega}$。样本空间的元素，即随机试验的每个可能结果，称为**基本事件**（elemental event）或**样本点**（sample point），记为 $\boldsymbol{\omega}$。

样本空间 Ω 的子集称为**随机事件**（random event），简称**事件**（event），通常用大写字母 A、B、C 等表示。设 A 是一个随机事件，一般 A 由一个或多个基本事件组成，当 A 中的一个基本事件出现时，就称**事件 A 发生**。

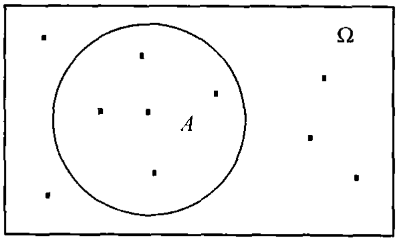

图 1-1 样本空间和事件 A

样本空间 Ω 包含所有的样本点，它是 Ω 自身的子集，在每次试验中它总是发生的，称为**必然事件**（certain event）。空集 Φ 不包含任何样本点，也是 Ω 的子集，它在每次试验中都不发生，称为**不可能事件**（impossible event）。

例 1-1 考察随机事件："掷一枚硬币，观察向上的面"。令

$$\omega_1 = \{出现正面\}, \omega_2 = \{出现反面\},$$

则其样本空间 $\Omega = \{\omega_1, \omega_2\}$。

例 1-2 我们考察随机试验："掷一枚骰子，观察其出现的点数"。

如果用 $\{i\}$ 表示 {出现 i 点}，则该试验共有六个基本事件：

$$\{1\},\ \{2\},\ \{3\},\ \{4\},\ \{5\},\ \{6\}$$

其样本空间 $\Omega=\{1,2,3,4,5,6\}$。"出现偶数点"这一随机事件是由 2、4、6 这三个基本事件组成，可表示为 $\{2, 4, 6\}$。在该试验中"点数不超过 6"就是必然事件，"出现 7 点"就是不可能事件。

二、事件间的关系和运算

（一）事件的包含与相等

若事件 A 发生必然导致事件 B 发生，则称**事件 B 包含事件 A** 或称**事件 A 包含于事件 B**，记为 $\boldsymbol{B\supset A}$ 或 $\boldsymbol{A\subset B}$。

例如，若记 $A=$ {考试优秀}，$B=$ {考试及格}，则 $A\subset B$。

若 $A\subset B$ 且 $B\subset A$，则称**事件 A 与事件 B 相等**，记为 $\boldsymbol{A=B}$。

对任一事件 A，有 $\Phi\subset A\subset\Omega$。在概率论中常用一个长方形表示样本空间 Ω，用其中的圆（或其它几何图形）表示事件，这类图形称为 **Venn 图**（Venn graph）。如图 1-2 表示 $A\subset B$ 的 Venn 图。

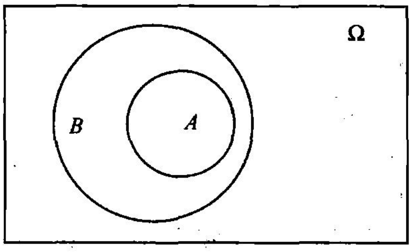

图 1-2 $A\subset B$

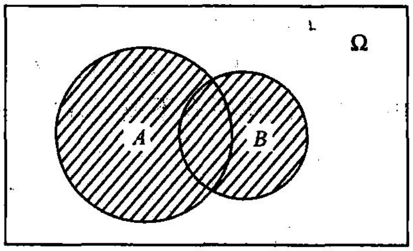

图 1-3 $A+B$（或 $A\cup B$）

（二）事件的和（或并）

事件 A 与事件 B 中至少有一个发生的事件称为**事件 A 与事件 B 的和**（或并），记为 $\boldsymbol{A+B}$（或 $\boldsymbol{A\cup B}$）（如图 1-3 中阴影部分所示）。

事件的和的定义可推广到多个事件，即：

$A_1+A_2+\cdots+A_n$ 表示 A_1、A_2、…、A_n 中至少有一个发生的事件。

（三）事件的积（或交）

事件 A 与事件 B 同时发生的事件称为**事件 A 与事件 B 的积**（或交），记为 $\boldsymbol{AB}$（或 $\boldsymbol{A\cap B}$）（如图 1-4 中阴影部分所示）。

例如，若记 $A=$ {患糖尿病}，$B=$ {患高血压}，则 $A+B=$ {患糖尿病或高血压}，$AB=$ {同时患糖尿病和高血压}。

事件的交的定义可推广到多个事件，如：

$A_1A_2\cdots A_n$ 表示 A_1、A_2、…、A_n 同时发生的事件。

（四）事件的差

如果事件 A 发生而事件 B 不发生，则称这样的事件为**事件 A 与事件 B 的差**，记为 $A-B$（如图 1－5 中阴影部分所示）。

例如，掷一枚骰子，设 $A=$｛点数大于3｝，$B=$｛点数为奇数点｝，则 $A-B=$｛点数为4或6｝，$B-A=$｛点数为1或3｝。

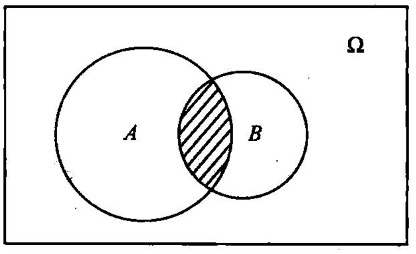

图 1－4　AB（或 $A\cap B$）

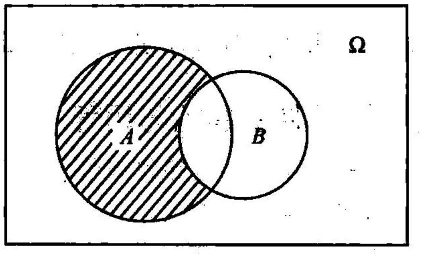

图 1－5　$A-B$

（五）互不相容事件

若 $AB=\Phi$，则称事件 A 与 B 是**互不相容或互斥的**（mutually exclusive），即事件 A 与事件 B 不能同时发生（如图 1－6）。

（六）对立事件

称“事件 A 不发生”的事件为 A 的**对立事件**（complementary event 或逆事件），记为 $\bar{A}$，它由样本空间中所有不属于 A 的基本事件所构成（如图 1－7 中阴影部分所示）。此时有：

$$A\bar{A}=\Phi,\ A+\bar{A}=\Omega。$$

如在药品质量检验中，$A=$｛重量过重｝，$B=$｛重量过轻｝，A 和 B 不可能同时发生，即 A 和 B 是互不相容的。如将重量过轻或过重均视为不合格，则 $C=$｛不合格｝$=A+B$，$D=$｛合格｝$=\bar{C}$

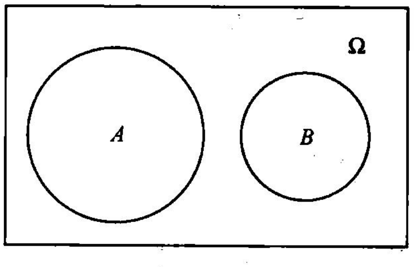

图 1－6　A 与 B 互不相容

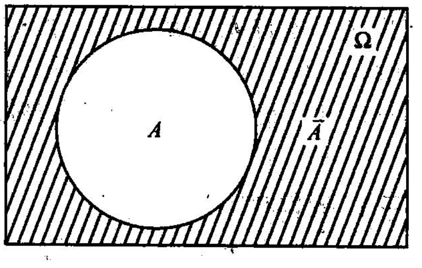

图 1－7　A 的对立事件 $\bar{A}$

（七）事件的运算律

（1）**交换律**：$A+B=B+A$；$AB=BA$。

（2）**结合律**：$(A+B)+C=A+(B+C)$；$(AB)C=A(BC)$。

（3）**分配律**：$(A+B)C=AC+BC$；$A+(BC)=(A+B)(A+C)$。

(4) **差积转换律**：$A-B=A\overline{B}=A-AB$

(5) **德·摩根 (De Morgan) 对偶律**：$\overline{A+B}=\overline{A}\,\overline{B}$；$\overline{AB}=\overline{A}+\overline{B}$。

对更一般的情形，有

$$\overline{A_1+A_2+\cdots+A_n}=\overline{A_1}\,\overline{A_2}\cdots\overline{A_n};$$

$$\overline{A_1A_2\cdots A_n}=\overline{A_1}+\overline{A_2}+\cdots+\overline{A_n}。$$

对于上述运算规律，我们可以利用 Venn 图和事件间的关系来验证其正确性。在事件表示中，我们称以运算符号联结起来的事件表示式为**事件式**（event expression）。在事件式中，事件的运算还应遵循下列运算顺序：先求“对立”，再求“积”，最后求“和”、“差”，遇有括号，先算括号内的。

在熟练掌握事件间的关系与运算的基础上，可用已知的简单事件来表示复杂事件。

例 1-3 现从一批含有次品的药品中连续抽取 3 件，设 A、B、C 分别表示抽取的第一件、第二件、第三件为合格品，试用 A、B、C 分别表示下列事件。

(1)“只有一件合格” $=A\overline{B}\,\overline{C}+\overline{A}B\overline{C}+\overline{A}\,\overline{B}C$；

(2)“至少一件合格”

$=A\overline{B}\,\overline{C}+\overline{A}B\overline{C}+\overline{A}\,\overline{B}C+AB\overline{C}+\overline{A}BC+A\overline{B}C+ABC=A+B+C$；

(3)“3 件都合格” $=ABC$；

(4)“3 件全不合格” $=\overline{A}\,\overline{B}\,\overline{C}=\overline{A+B+C}$。

三、概率的定义

除了必然事件和不可能事件外，任一随机事件在一次试验中都有发生与不发生的可能性，人们往往通过实际分析来估计某个事件发生的可能性的大小。例如遇到某种天气，人们常会说“今天十有八九会下雨”，这个“十有八九”就是表示“今天下雨”这个事件发生的可能性的大小。这是人们通过大量观察得出的一种统计规律，即已经历了若干次这种天气，下雨的天数所占的比例大约是$\frac{8}{10}$到$\frac{9}{10}$。一般地，人们希望用一个适当的数值来表示一个事件在一次试验中发生的可能性的大小。

定义 1-1 事件 A 发生的**概率**（probability）是事件 A 在试验中出现的可能性大小的数值度量，用 $P(A)$ 表示。

基于对概率的不同情形的应用和不同解释，概率的定义有所不同，主要有统计概率、古典概率和主观概率等定义。

（一）统计概率

定义 1-2 在相同的条件下重复进行 n 次试验，若事件 A 发生了 m 次，则称 m 为事件 A 在这 n 次试验中发生的**频数**（frequence），称比值$\frac{m}{n}$为事件 A 在这 n 次试验中发生的**频率**（frequency 或 relative frequency），记为 $f_n(A)=\frac{m}{n}$。

人们通过大量实践，发现随机事件的一个极其重要的特性：在充分多次的重复试验中，事件 A 的频率 $f_n(A)$ 总在某定值 p 附近波动，且随着试验次数的增加，它逐渐趋向

于p，这就是**频率的稳定性**（stability of relative frequency）。许多统计学家如 K. Pearson 曾做过大量的抛硬币试验，发现正面向上的频率都逐渐接近于0.5。我国历次人口普查中男性所占的比例都逐渐接近于0.516（见表1-1）。

表1-1 我国历次普查中男性占总人口的比例 单位：万人

普查年份	总人口	男性	女性	男性占总人口的比例
1953	59435	30799	28636	0.5182
1964	69458	35652	33806	0.5133
1982	100818	51944	48874	0.5152
1990	113368	58495	54873	0.5160
2000	126583	65355	61228	0.5163

它表明数p是事件A本身客观存在的一种固有属性，因此，数p可以对事件A发生的可能性大小进行度量。

定义1-3 当试验次数n充分大时，事件A的频率$f_n(A)$总在区间$[0, 1]$上的某个数值p附近波动，且随着试验次数n的增加，其波动的幅度越来越小，则称p为事件A的**统计概率**（statistical probability），记为$P(A)$，即$P(A)=p$。

实际应用中，利用概率的统计定义，即可将试验次数充分大时事件A出现的频率值作为事件的概率的近似值，即$P(A)\approx\frac{m}{n}$，这在概率不易求出时很有效。

例1-4 （关于抽烟和肺癌的关系调查）在某城市随机抽取10万个40岁以上从不抽烟的男性和10万个40岁以上抽烟的人，根据随访结果表明，前一组最后因肺癌死亡的是30个，而后一组是600个。因此我们得到：

一般人致癌的概率：$p_1\approx\frac{m}{n}=\frac{30}{100000}=0.03\%$，

抽烟致癌的概率：$p_2\approx\frac{m}{n}=\frac{600}{100000}=0.6\%$，

抽烟致癌的概率是一般人致癌的概率的20倍。

（二）古典概率

根据上述概率的定义，要计算某事件的概率，就得做大量的重复试验，然而在某些情况下根据事件本身所具有的对称性，可以直接计算事件的概率，这就是古典概率。

定义1-4 设随机试验具有如下两个特征

(1) 样本空间所含的基本事件只有有限个，即$\Omega=\{\omega_1, \omega_2, \cdots, \omega_n\}$；

(2) 每一个基本事件发生的可能性相等，即$P(\omega_1)=P(\omega_2)=\cdots=P(\omega_n)$。

则称试验所对应的概率模型为**古典概型**（classical probability model）或**有限等可能概型**。

定义1-5 对于给定的古典概型，若样本空间中基本事件总数为n，而事件A包含其中的m个基本事件，则称比值$\frac{m}{n}$为事件A的**古典概率**（classical probability），记为$P(A)$，即