普通高等教育“十二五”经济管理类规划教材

經華管信 创优系列·管理科学与工程

管理统计学（第2版）

——基于SPSS软件应用

MANAGEMENT STATISTICS, 2EDITION

——BASED ON SPSS

王雪华　　主　编

张　鹏　张承伟　副主编

電子工業出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

http://www.phei.com.cn

普通高等教育“十二五”经济管理类规划教材

經華管信 创优系列·管理科学与工程

管理统计学（第2版）

——基于SPSS软件应用

MANAGEMENT STATISTICS, 2EDITION

——BASED ON SPSS

王雪华　　主　编

张　鹏　张承伟　副主编

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书是大连理工大学精品课程建设成果，是一本适合管理类学生学习的应用统计教材，以其问题导向型写作风格、丰富的案例、结合软件的应用，以及选材的系统性与完整性，而与众不同。全书内容覆盖常用的统计分析方法及SPSS软件应用，共11章，包括概率论基础、数据的搜集与整理、参数估计、假设检验、方差分析、正交试验、相关分析、线性回归、聚类分析与判别分析、生存分析、主成分分析与因子分析。本书配有习题解答和电子课件，供任课教师免费使用。

本书可作为普通高等院校信息管理与信息系统、人力资源管理、国际经济与贸易、金融学等专业管理统计学教材，也是学习统计分析方法及其SPSS软件应用的本科生、研究生的自学用书，还可供经济、管理统计工作者参考。

图书在版编目(CIP)数据

管理统计学：基于SPSS软件应用/王雪华主编. —2版. —北京：电子工业出版社，2014.1
(华信经管创优系列)
ISBN 978-7-121-21831-6

I. ①管… II. ①王… III. ①经济统计学-高等学校-教材 IV. ①F222

中国版本图书馆CIP数据核字(2013)第265477号

策划编辑：秦淑灵
责任编辑：秦淑灵
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　　　　北京市海淀区万寿路173信箱　邮编　100036
开　　本：787×1092　1/16　印张：20　字数：551千字
印　　次：2014年1月第1次印刷
定　　价：43.00元

前　　言

统计学课程是国家教育部确定的高等院校财经类专业 11 门核心课程之一，是一门收集、整理和分析统计数据的方法科学，其目的是探索数据内在的数量规律性，以达到对客观事物的科学认识。管理统计学是经济管理类学科的核心课程，是管理科学和社会科学领域中应用最广泛的数量研究方法。随着管理现代化和科学化进程的不断加快，管理统计学正在发挥着更大的作用并得到更广泛的关注。

SPSS 软件是目前社会科学领域使用的功能强大的统计软件之一，它简化了统计计算和研究工作，降低了运用统计研究方法的数学门槛，大大提高了工作效率，为统计方法在管理科学和社会科学领域的普及应用提供了便利条件。

本书结合编者多年讲授管理统计学课程的教学经验编写而成。全书共 11 章，详细介绍了描述性统计学与推断性统计学两大分支。描述性统计学重点培养统计人员进行统计资料收集、整理等综合能力；推断性统计学重点培养统计研究人员进行统计抽样，运用样本信息对总体进行参数估计、假设检验、方差分析及统计回归等能力。另外，每章均配有精心安排的案例和习题。教学资源中有习题解答和电子课件，供任课教师免费使用。任课老师可登录华信教育资源网 http://www.hxedu.com.cn 免费注册下载。

本书内容丰富，实用性强，侧重采用案例和引入现代统计分析工具 SPSS 等进行管理统计基本知识的介绍，强调统计原理和技术在管理科学中的应用及 SPSS 软件计算原理的透彻理解和正确应用，努力构建一个贯通统计学原理、SPSS 软件使用技巧且应用于统计学与管理科学领域的应用型统计学知识体系。

通过对本书的学习，读者既能根据实际工作需要设计统计调查问卷、将实际问题转化成统计问题，又能利用 SPSS 的强大功能整理和分析数据，并通过科学的数据解释，将统计结果应用于管理实践当中。

本书具有如下特点：

（1）内容全面。教材内容体系、前后内容的衔接、知识点的安排由浅入深，循序渐进。全书围绕管理实际中可能遇到的统计问题，详细介绍管理统计学的研究思路和技术方法，使读者掌握具有重要应用价值的常用统计学原理、方法和工具。

（2）通俗易懂。本书在编写过程中力求做到概念叙述准确、严谨，语言简练通俗。为了面向实际需要，强调应用，书中略去了复杂的数学推导过程，通过丰富完整的统计实例引出章节的内容，从数据的获取、统计学原理、方法与研究实际问题和 SPSS 软件的使用及案例的分析等方面，以问题导向型的写作风格引导读者进行学习。

（3）实践性强。书中案例丰富，覆盖面广，既包括统计学理论的内容，又有增加兴趣和提供管理实践分析能力的案例。在介绍实例的同时也全面系统地介绍了如何选择 SPSS 的选项，以及这些选项的功能与应用，最后对实例统计分析的结果进行解释，得到正确的结论。

本书适合作为高等院校工商管理类专业“统计学”课程教材和工程技术人员学习 SPSS 的参考书。另外，本书还可作为自学用书。

本书适合于学时较少，并与多媒体教学和案例讨论方法相匹配的教学方式。对于本书的学时数，建议课堂教学32～48学时，上机时间12～16学时。为了解决学时少、内容多的矛盾，建议在教学中注重培养学生分析和解决实际问题的能力，采用任务驱动、案例教学相结合的教学模式，以提高学生学习的兴趣和主动性。

本书第1，2章由张承伟编写，第3，4，5，6，9章由张鹏编写，第7，8，10，11章由王雪华编写。全书由张承伟、王雪华统稿。

在本书的编写过程中，研究生刘淼、刘莹、葛月、刘畅、肖君德等在素材的收集、整理、录入及案例编写、习题解答、SPSS软件操作等方面做了大量工作，编者在此表示诚挚的谢意！

由于编者水平和能力所限，书中难免有疏漏和不当之处，敬请读者和专家指正。反馈建议请发至作者 E-mail：zcw@dlut.edu.cn。

编　者

于大连理工大学

目　　录

第 1 章　概率论基础

【引例】　现实中的统计。

X 商店位于 Y 市西郊，是一家以经营生鲜食品、日杂用品为主的中型百货商店。在 X 商店正式营业的第一年末，商店经理决定购入一批挂历进行销售，但购入挂历的数量成为困扰经理的一个难题——一方面如果购入的挂历数量不够，那么待挂历售尽便会出现缺货损失，从而只能眼睁睁地看着大笔生意被竞争对手抢走；另一方面如果购入的挂历数量过多，多余的存货积压必然会造成流动资金的短缺及存货费用的增加，因此只能做削价处理，这必将给商店带来经济上的损失。

为了使收益值的期望最大，经理请教了在高校任教的王老师来为商店确定一个合适的挂历进货量。已知商店每出售一件挂历可获得纯利润 7 元；但如果在春节以前不能售尽，则需要做削价处理，每件将亏损 3 元。

王老师调查了 Y 市与鼎文商店各种情况类似的十家商店，统计了每家商店在最近四年春节前挂历的销售情况。根据调查结果，得到了表 1–1 中的数据。

表 1–1　十家商店在最近四年的挂历销售量分布表

销售量	100	200	300	400	500	600	合计
次数	2	5	13	11	6	3	40

根据表 1–1 中的数据，王老师通过计算得到了另外两个表——挂历销售量的概率分布表和收益与收益期望分布状况表，并得出结论：当挂历的进货量为 400 件时，商店的期望收益最大，为 2075 元。按照经理的“收益值的期望最大”的要求，王老师向经理建议商店的进货量为 400 件。

上述引例中所涉及的概率、概率分布、期望等概念均属于概率论的范畴。概率论是研究随机现象规律性的数学分支，在科学研究和社会生产实践中有着十分广泛的应用，是统计研究的基础。本章将介绍一些概率论的基础理论，包括事件与概率、概率的基本性质、条件概率与事件独立性，以及随机变量及其分布。

1.1　事件与概率

事件与概率是概率论研究中的两个最基本的概念。围绕着这两个概念，本节将介绍三部分内容，包括随机试验与随机事件、事件的关系及运算、事件的概率等内容。

1.1.1　随机试验与随机事件

1. 随机试验与样本空间

在自然界和人类社会生产实践中，存在着两类现象。一类现象在一定条件下必然发生（或必然不发生）。例如，在标准大气压下，水的沸点是 100℃；又如向上抛掷一枚石子，由于受地心引力的作用，石子在上升到一定高度之后必然下落。由于这类现象具有确定的结果，故称为确定性现象。

然而，并不是所有的现象都具有确定性的结果。例如，抛掷一枚硬币，当硬币落地后，可能是正面朝上，也可能是反面朝上，而在硬币落地前不能预知究竟哪一面朝上。同样地，自动机床加工制造同一零件，加工出来的零件可能是合格品也可能是不合格品；同一门炮向同一目标发射多发同种炮弹，弹落点也不一样，等等。以上列举的现象均具有不确定性，即在基本条件不变的情况下，一系列试验或观察会得到不同的结果，并且在每次试验或观察之前不能预知会出现哪种结果，这类现象称为随机现象。概率论研究的对象就是随机现象。

【例 1–1】 生活中随机现象的例子。

① 抛掷一颗骰子，出现的点数；

② 一天内进入某超市的顾客数；

③ 某一生产线生产出的灯泡的寿命；

④ 某批产品的不合格率。

为了探索和研究随机现象的规律性，通过随机试验（简称试验）来对随机现象进行调查、观察或实验。具体来说，随机试验应满足如下条件：

- 试验可以在相同的条件下重复进行；
- 试验有多种可能的结果，并且事先可以明确所有可能出现的结果；
- 试验完成之前不能预知会出现哪一个结果。

一个随机试验的所有可能结果的集合称为样本空间，通常用 Ω 表示。样本空间的元素，即试验的每一个可能结果，称为这个试验的样本点，用 ω 表示。

【例 1–2】 试列出例 1–1 中随机现象的样本空间。

解： ① 掷一颗骰子的样本空间为 $\Omega_1=\{\omega_1,\ \omega_2,\ \cdots,\ \omega_6\}$，其中 ω_i 表示出现的 i 点，$i=1,\ 2,\ \cdots,\ 6$。也即掷一颗骰子的样本空间为 $\Omega_1=\{1,\ 2,\ \cdots,\ 6\}$。

② 一天内进入某超市顾客数的样本空间为 $\Omega_2=\{0,\ 1,\ 2,\ \cdots\}$，其中“0”表示一天内无人光顾该超市。

③ 某生产线生产出的灯泡的寿命的样本空间为 $\Omega_3=\{t\mid t\geqslant 0\}$。

④ 产品的不合格率一定是介于 0 与 1 之间的一个实数，因此其样本空间为 $\Omega_4=\{y\mid 0\leqslant y\leqslant 1\}$。

在例 1–2 中，Ω_1 中的样本点的个数为有限个，是比较简单的样本空间；而 Ω_2、Ω_3 和 Ω_4 中的样本点的个数为无限个，但 Ω_2 中的样本点可以按照某种次序排列出来，即 Ω_2 中有可列个样本点。在概率研究中，将包含有限个或可列个样本点的样本空间称为离散样本空间，如例 1–2 中的 Ω_1 和 Ω_2；而将包含无限个或不可列个样本点的样本空间称为连续样本空间，如例 1–2 中的 Ω_3 和 Ω_4。

2. 随机事件

样本空间 Ω 的某个子集称为随机事件，简称事件，通常用大写字母 A，B，C，…表示。

随机事件表示试验可能出现的结果，这个结果可以是仅由一个样本点组成的基本事件，也可以是由多个样本点组成的复合事件。

对于某一事件 A，当且仅当它所包含的某一样本点出现时，称事件 A 发生。

例如，在掷骰子试验中，投掷一颗均匀的骰子，其样本空间为 $\Omega=\{1,\ 2,\ \cdots,\ 6\}$。现在从不同的角度考察该实验的结果：记事件 A 为“出现 2 点”，事件 B 为“出现偶数点”，事件 C 为“出现的点数小于 7”，事件 D 为“出现的点数大于 6”。

其中，A 为基本事件，当且仅当掷出 2 点时，事件 A 发生，即 $A=\{2\}$；事件 B 发生当且仅当下列样本点之一发生：掷出 2 点、掷出 4 点和掷出 6 点，它由三个基本事件复合而成，即 $B=\{2,\ 4,\ 6\}$。

对于事件 C，在一次试验中，由于每次抛掷骰子出现的点数必然小于7，因此事件 C 必然发生，即 $C=\{1, 2, \cdots, 6\}$。通常，把样本空间 Ω 本身称为必然事件，事件 C 就是一个必然事件。

同样地，对于事件 D，由于每次抛掷骰子出现的点数不可能大于6，因此事件 D 不可能发生，即 $D=\{\}$。通常，把空集 $\varnothing$ 称为不可能事件，事件 D 即不可能事件。

严格来讲，必然事件和不可能事件反映了确定性现象，本质上不是随机事件，然而为了研究方便，还是把必然事件和不可能事件作为随机事件的两个极端情形来处理。

1.1.2　事件的关系及运算

在一个随机试验中，样本空间可以定义的随机事件显然不止一个，同时，事件与事件之间必然存在这样那样的联系。为了能够更好地理解及运用随机试验的结果，下面将借助文氏图分析事件的关系及运算。

1. 事件之间的关系

事件间的包含关系：若事件 A 发生必然导致事件 B 发生，则称 A 包含于 B，或 B 包含 A，记为 $A\subset B$ 或 $B\supset A$，如图1–1所示。

例如，在掷骰子试验中，若记事件 A 为“出现2点”，事件 B 为“出现偶数点”，则 $A\subset B$。显然，对于任一事件 A，必有 $\varnothing\subset A\subset\Omega$。

事件间的相等关系：若事件 A 发生必然导致事件 B 发生，同时事件 B 发生必然导致事件 A 发生，则称事件 A 与 B 相等，记为 $A=B$。相等的两事件在实质上是对同一事件的不同语言描述。

事件间的互不相容关系：若事件 A 与事件 B 不可能同时发生，则称事件 A 与 B 互不相容，如图1–2所示。

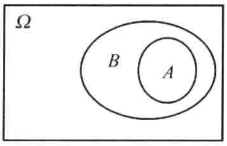

图1–1　$A\subset B$

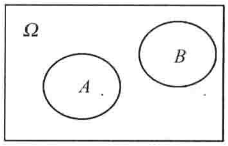

图1–2　A 与 B 互不相容

同样以掷骰子试验为例，在试验中，“出现的点数小于2”与“出现的点数大于4”两个事件不可能同时发生，因而它们是互不相容事件。

2. 事件的运算

对于样本空间中的事件，可以通过以下四种事件的基本运算，得到新的事件。现定义两个事件 A 和 B，事件的运算有如下四种。

事件的并：由属于事件 A 或 B 的所有样本点构成的集合称为事件 A 与 B 的并，记为 $A\cup B$。特别地，对于互不相容的事件 A 和 B，称它们的并为和，记为 $A+B$。

事件 $A\cup B$ 表示事件 A 和 B 至少发生一个。例如，在掷骰子试验中，若记事件 A 为“出现的点数大于1小于3”，事件 B 为“出现的点数大于2小于4”，则事件 $A\cup B$ 表示“出现的点数大于1小于4”。

事件的交：由同时属于事件 A 和 B 的所有样本点构成的集合称为事件 A 与 B 的交，记为 $A\cap B$ 或 AB。

事件 $A \cap B$ 表示事件 A 和 B 同时发生。例如，在掷骰子试验中，若记事件 A 为“出现的点数大于1小于4”，事件 B 为“出现的点数大于2小于5”，则事件 $A \cap B$ 表示“出现的点数大于2小于4”，即出现3点。

事件的差：由属于事件 A，但不属于事件 B 的所有样本点构成的集合称为事件 A 与 B 的差，记为 $A-B$。

事件 $A-B$ 表示事件 A 发生而事件 B 不发生。例如，在掷骰子试验中，若记事件 A 为“出现的点数大于1小于4”，事件 B 为“出现的点数大于1小于3”，则事件 $A-B$ 表示“出现的点数大于等于3小于4”，即出现3点。

事件的逆：由样本空间中不属于事件 A 的所有样本点构成的集合称为事件 A 的逆，记为 $\overline{A}$。

事件 $\overline{A}$ 是事件 A 的对立事件，表示事件 A 不发生。例如，在掷骰子试验中，若记事件 A 为“出现的点数小于3”，则事件 $\overline{A}$ 表示“出现的点数大于等于3”。

以上四种事件运算的文氏图如图1-3所示。

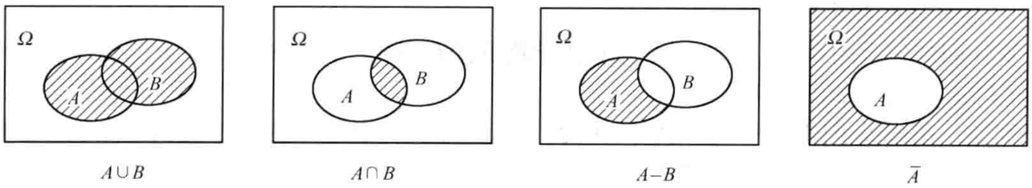

图1-3 四种事件运算的文氏图

【例1-3】 假设一批产品中有3件次品，且产品的外形没有任何差别。现随机地从这批产品中依次抽取3件，若以 A 记“第一次抽到次品”，以 B 记“第二次抽到次品”，以 C 记“第三次抽到次品”，试用 A、B 和 C 的关系表示下列各事件。

① 三次都抽到次品。

② 只有第一次抽到次品。

③ 三次都没有抽到次品。

④ 至少抽到一件次品。

⑤ 最多抽到一件次品。

⑥ 最多抽到两件次品。

解：

① ABC

② $A\overline{B}\,\overline{C}$

③ $\overline{A}\,\overline{B}\,\overline{C}$

④ $A \cup B \cup C$

⑤ 最多抽到一件次品，即 A，B，C 中只有一个发生或 A，B，C 全不发生，即

$$\overline{A}\,\overline{B}C \cup \overline{A}B\overline{C} \cup A\overline{B}\,\overline{C} \cup \overline{A}\,\overline{B}\,\overline{C}$$

⑥ 最多抽到两件次品，即是 A，B，C 全发生的对立事件，即

$$\overline{ABC}$$

（3）事件运算的性质

事件的运算与集合的运算一样，必须满足如下运算法则。

① 交换律：$A \cup B = B \cup A$，$AB = BA$。

② 结合律：$(A\cup B)\cup C=A\cup(B\cup C)$，$(AB)C=A(BC)$。

③ 分配律：$(A\cup B)\cap C=AC\cup BC$，$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$。

④ 对偶律(德摩根公式)：$\overline{A\cup B}=\overline{A}\cap\overline{B}$，$\overline{A\cap B}=\overline{A}\cup\overline{B}$。

特别地，德摩根公式可以推广到对于 n 个事件或可列个事件求对偶的问题。

1.1.3　事件的概率

对于一次随机试验，在试验结束之前并不能确定某个事件是否会发生，这是由随机试验的基本性质所决定的。例如，在一次摸球试验中，假定袋子中有包括 9 只白球和 1 只黑球在内的 10 只小球，在实验结束之前，并不能确定会摸出黑球还是白球。显然摸出白球的可能性比摸出黑球的可能性大得多。

随机事件发生的可能性大小不仅可以比较，而且是可以量化的。对于一个随机事件 A，若可以用一个数 $P(A)$ 来表示其发生的可能性大小，这个数 $P(A)$ 就称为随机事件 A 的概率。

概率度量随机事件发生的可能性的大小，它由随机事件自身所决定，反映了随机现象的内在规律。那么，概率究竟应该如何量化呢?

1. 概率的统计定义

对于随机事件 A，如果它在 N 次试验中发生了 n 次，则称

$$F_N(A)=n/N \tag{1-1}$$

为随机事件 A 在 N 次试验中出现的频率。

显而易见，频率具有如下性质。

① 非负性：对于随机事件 A，必有 $F_N(A)\geqslant 0$。

② 规范性：对于必然事件 Ω，在 N 次试验中出现的次数应为 N，即 $F_N(\Omega)=1$。

③ 可加性：若 A 和 B 为互不相容事件，则 $F_N(A\cup B)=F_N(A)+F_N(B)$。

以上三条性质为频率的基本性质，根据这些性质，还可以得出许多其他的性质，例如，“任何随机事件在 N 次试验中出现的频率都不大于 1”，“不可能事件在 N 次试验中出现的频率为 0”，“对于有限个两两互不相容的事件的频率也具有可加性”等。

另外，对于多次重复试验，随机事件 A 的频率还具有另外一项重要的性质——频率稳定性。

人们经过长期的生产实践发现：在相同条件下进行的多次重复试验，随着试验重复次数 N 的增加，随机事件 A 的频率 $F_N(A)$ 会在某一固定的常数 a 附近摆动，频率的这个性质称为频率稳定性，这个固定的常数 a 就是概率。

下面，以抛硬币试验为例来说明频率的稳定性。

在抛一枚硬币时，既可能出现正面，也可能出现反面，预先做出判断是不可能的，但是假如硬币均匀，直观上出现正面与出现反面的机会应该相等，即在大量试验中出现正面的频率应该接近于 50%，为了验证这一点，历史上曾有不少人做过这个试验，结果如表 1-2 所示。

表 1-2　历史上抛硬币试验的若干结果

试验者	抛硬币次数	出现正面次数	出现正面频率
德摩根	2048	1061	0.5181
蒲丰	4040	2048	0.5069
费勒	10000	4979	0.4979
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

从表1-2可以看出，出现正面的频率在0.5附近摆动，根据频率稳定性可知，出现正面的概率为0.5。

由于概率与频率的密切关系，在实际应用中，常常需要根据历史数据，统计某一事件发生的频率，以估计其概率。例如，在北京奥运会前夕，中国气象局分析了北京市观象台1951—2007年历时57年的气象资料，得到北京地区在2008年8月8日奥运会开幕式当天降水的概率为47%。

2. 概率的古典定义

在讨论概率的统计定义时曾提到过，抛一枚均匀的硬币，直观上出现正面与出现反面的概率是相等的，并且历史上大量的试验数据也验证了这一观点。类似于抛硬币的试验，在人类的生产实践中，存在着许多这类随机现象，诸如掷骰子、产品抽样检查等，对于这些随机现象进行深入分析之后，可以发现，它们之间存在以下两个基本的共同点：

① 试验具有有限个可能出现的结果；

② 试验的每个基本事件出现的可能性都是相等的。

具有以上两个基本特点的概率模型称为古典概型。古典概型在概率论的发展初期即被注意，它的内容简单，应用却很广泛，许多最初的概率论结果也是由它得出的。

在古典概型中，假定样本空间 $\Omega=\{\omega_1, \omega_2, \cdots, \omega_n\}$，则对于每个基本事件 ω_i 有

$$P(\omega_1)=P(\omega_2)=\cdots=P(\omega_n)=\frac{1}{n}$$

进一步讲，对于古典概型，如果一个试验有 n 个基本事件，其中随机事件 A 包含的基本事件个数为 m，那么随机事件 A 的概率为

$$P(A)=\frac{\text{事件 } A \text{ 所包含的样本点的个数}}{\text{样本点总数}}=\frac{m}{n} \tag{1-2}$$

式中，由于事件 A 包含样本点的出现必然导致 A 的出现，因此又称这些样本点为 A 的“有利场合”。

在古典概型中，通过以上公式求解随机事件 A 的概率，首先要明确事件 A 中所包含的样本点数和样本空间中的样本点总数，计算时需要熟练地运用排列和组合的相关知识，具有一定的技巧性。

【例1-4】 口袋中有5个白球、3个黑球，从中任取两个，求取到的两个球颜色不同的概率。

解： 从8个球中任取两个，共有 C_8^2 种不同的取法，将每一种取法作为该试验的一个样本点，可以得到取球试验的样本空间。由于是随机取球，任意两个小球被同时取出的概率是相等的，因此这个问题是古典概型。

记“取到的两个球颜色不同”为事件 A，则事件 A 包含的样本点数为 $C_5^1C_3^1$，因此取到两个不同颜色的球的概率为

$$P(A)=\frac{C_5^1C_3^1}{C_8^2}=\frac{15}{28}$$

摸球模型在实际问题中有很重要的应用。例如，如果把例1-4中的“白球”、“黑球”替换为“正品”、“次品”，就可以用来求解产品质量抽样检查问题。另外，还可以向口袋中加入其他颜色的球，使摸球模型更具有代表性，这时就可以描述具有更多等级的产品抽样问题，如将产品分为一等品、二等品、三等品和等外品的产品抽样检查问题。

3. 概率的几何定义

通过古典概型成功地解决了一类问题，这类问题有且只有 n 个基本事件，并且每个基本事