经全国中小学教材审定委员会2006年初审通过

普通高中课程标准实验教科书

数学

（选修2-3）

供学习用

SHUXUE

主　　编　严士健　王尚志

副 主 编　张饴慈　李延林　张思明

本册主编　张饴慈　吴江媛

编写人员　（按姓氏笔画排序）

王建波　关　健　吴江媛

张　丹　张饴慈　袁京生

北京师范大学出版社

·北　京·

前　言

你们将进入更加丰富多彩的数学世界.

你们将学到更多重要和有趣的数学知识、技能及应用.

你们将更多地感受到深刻的数学思想和方法.

你们将进一步体会数学对发展自己思维能力的作用，体会数学对推动社会进步和科学发展的意义，体会数学的文化价值.

你们正在长大，需要考虑自己未来的发展. 要学习的东西很多，高中数学的内容都是基础的，时间有限，选择能力是很重要的，你们需要抓紧时间选择发展的方向，选择自己感兴趣的专题，这是一种锻炼.

在高中阶段，学习内容是很有限的. 中国古代有这样的说法："授之以鱼，不如授之以渔"，学会打鱼的方法比得到鱼更重要. 希望同学们不仅关注别人给予你们的知识，更应该关注如何获得知识. 数学是提高"自学能力"最好的载体之一.

在数学中，什么是重要的（What is the key in Mathematics）？20 世纪六七十年代，在很多国家都讨论了这个问题. 大部分人的意见是：问题是关键（The problem is the key in Mathematics）. 问题是思考的结果，是深入思考的开始，"有问题"也是创造的开始. 在高中数学的学习中，同学们不仅应提高解决别人给出问题的能力，提高思考问题的能力，还应保持永不满足的好奇心，大胆地发现问题、提出问题，养成"问题意识"和交流的习惯，这对你们将来的发展是非常重要的.

在学习数学中，有时会遇到一些困难，树立信心是最重要的. 不要着急，要有耐心，把基本的东西想清楚，逐步培养自己对数学的兴趣，你会慢慢地喜欢数学，她会给你带来乐趣.

本套教材由 26 册书组成：必修教材有 5 册；选修系列 1 有 2 册，选修系列 2 有 3 册，它们体现了发展的基本方向；选修系列 3 有 6 册，选修系列 4 有 10 册，同学们可以根据自己的兴趣选修其中部分专题. 习题分为三类：一类是可供课堂教学使用的"练习"；一类是课后的"习题"，分为 A，B 两组；还有一类是复习题，分为 A，B，C 三组.

研究性学习是我们特别提倡的. 在教材中强调了问题提出，抽象概括，分析理

解，思考交流等研究性学习过程．另外，还专门安排了“课题学习”和“探究活动”．

“课题学习”引导同学们递进地思考问题，充分动手实践，是需要完成的部分．

在高中阶段，根据课程标准的要求，学生需要至少完成一次数学探究活动，在必修课程的每一册书中，我们为同学们提供的“探究活动”案例，同学们在教师的引导下选做一个，有兴趣也可以多做几个，我们更希望同学们自己提出问题、解决问题，这是一件很有趣的工作．

同学们一定会感受到，信息技术发展得非常快，日新月异，计算机、数学软件、计算器、图形计算器、网络都是很好的工具和学习资源，在条件允许的情况下，希望同学们多用，“技不压身”．它们能帮助我们更好地理解一些数学的内容和思想．教材中有“信息技术建议”，为同学们使用信息技术帮助学习提出了一些具体的建议；还有“信息技术应用”栏目，我们选取了一些能较好体现信息技术应用的例子，帮助同学们加深对数学的理解．在使用信息技术条件暂时不够成熟的地方，我们建议同学们认真阅读这些材料，对相应的内容能有所了解．教材中信息技术的内容不是必学的，仅供参考．

另外，我们还为同学们编写了一些阅读材料，供同学们在课外学习，希望同学们不仅有坚实的知识基础，而且有开阔的视野，能从数学历史的发展足迹中获取营养和动力，全面地感受数学的科学价值、应用价值和文化价值．

我们祝愿同学们在高中数学的学习中获得成功，请将你们成功的经验告诉我们，以便让更多的朋友分享你们成功的喜悦．

我们的联系方式是：北京师范大学出版社基础教育分社（100875），010-58802811．

目　录

第一章 计数原理

在日常的生产、生活中，我们常常会遇到一些需要计数的问题. 例如：

2004 年中国足球协会超级联赛有 12 支球队参加，每支球队要和其余的 11 支球队进行比赛，而且在主场和客场各赛一次，那么，这次联赛一共要安排多少场比赛呢？

我国许多地区的电话号码，都由 6 位升至 8 位，这样电话号码可以增加多少？

回答这些问题，就会用到本章将要学习的计数知识.

本章主要介绍分类加法计数原理和分步乘法计数原理，我们将利用这两个原理，讨论排列、组合等简单计数问题，并得到重要的二项式定理.

§1　分类加法计数原理和分步乘法计数原理

1.1　分类加法计数原理

实例分析

问题 1　从天津到大连，可以乘飞机，可以乘火车，也可以乘汽车，还可以乘轮船.

每天有 2 个航班的飞机，有 4 个班次的火车，有 2 个班次的轮船，有 1 个班次的汽车. 那么，乘坐以上交通工具从天津到大连，在一天中一共有多少种选择呢？

分析　如图 1-1，从天津到大连，共有乘飞机、火车、轮船、汽车 4 类办法，每类办法中分别又有 2，4，2，1 种方法，共有 2＋4＋2＋1＝9 种方法.

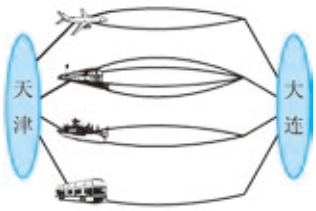


图 1-1

以上问题的特点是：

(1) 完成一件事有若干种方法，这些方法可以分成 n 类；

(2) 用每一类中的每一种方法都可以完成这件事；

(3) 把各类的方法数相加，就可以得到完成这件事的所有方法数.

抽象概括

一般地，有如下原理：

分类加法计数原理　**完成一件事，可以有 n 类办法，在第一类办法中有 m_1 种方法，在第二类办法中有 m_2 种方法，……，在第 n 类办法中有 m_n 种方法. 那么，完成这件事共有**

$$N=m_1+m_2+\cdots+m_n$$

种方法.（也称加法原理）

例 1　在 1，2，3，…，200 中，能够被 5 整除的数共有多少个？

解　能够被 5 整除的数，末位数字是 0 或 5，因此，我们把 1，2，3，…，200 中能够被 5 整除的数分成两类来计数：

第一类：末位数字是 0 的数，一共有 20 个.

第二类：末位数字是 5 的数，一共有 20 个.

根据加法原理，在1，2，3，…，200中，能够被5整除的数共有 20＋20＝40 个.

1.2　分步乘法计数原理

实例分析

图 1-2

问题 2　从 A 村去 B 村的道路有 3 条，从 B 村去 C 村的道路有 2 条，从 C 村去 D 村的道路有 3 条（如图 1-2 所示）. 李明要从 A 村先到 B 村，再经过 C 村，最后到 D 村，一共有多少条线路可以选择？

分析　整个行程必须通过 3 个步骤：先从 A 村到 B 村，再从 B 村到 C 村，然后从 C 村到 D 村.

从 A 村到 B 村有 3 条路，选择这 3 条路中的任意一条都可到达 B 村，再从 B 村到 C 村又有 2 条路. 因此，从 A 村经 B 村到 C 村一共有：$3\times2=6$ 条线路可以选择.

对于这 6 条线路中的每一条，再从 C 村到 D 村又有 3 条路. 因此，整个行程一共有：$3\times2\times3=18$ 条线路可以选择.

以上问题的特点是：

（1）完成一件事需要经过 n 个步骤，缺一不可；

（2）完成每一步有若干方法；

（3）把各个步骤的方法数相乘，就可以得到完成这件事的所有方法数.

抽象概括

一般地，有如下原理：

分步乘法计数原理　**完成一件事需要经过 n 个步骤，缺一不可，做第一步有 m_1 种方法，做第二步有 m_2 种方法，……，做第 n 步有 m_n 种方法. 那么，完成这件事共有**

$$N=m_1\times m_2\times\cdots\times m_n$$

种方法.（也称乘法原理）

例 2　有一项活动，需在 3 名教师、8 名男生和 5 名女生中选人参加.

(1) 若只需 1 人参加，有多少种选法？

(2) 若需教师、男生、女生各 1 人参加，有多少种选法？

解　(1) 只要选出 1 人就可以完成这件事，而选出的 1 人有 3 种不同类型，即教师、男生或女生，因此要分类相加.

第一类：选出的是教师，有 3 种选法.

第二类：选出的是男生，有 8 种选法.

第三类：选出的是女生，有 5 种选法.

根据加法原理，共有 $N=3+8+5=16$ 种选法.

(2) 完成这件事需要分别选出 1 名教师、1 名男生和 1 名女生，可以先选教师，再选男生，最后选女生，因此要分步相乘.

第一步：选 1 名教师，有 3 种选法.

第二步：选 1 名男生，有 8 种选法.

第三步：选 1 名女生，有 5 种选法.

根据乘法原理，共有 $N=3\times8\times5=120$ 种选法.

练　习

1. 完成一项工作，有两种方法，有 5 个人只会用第一种方法，另外有 4 个人只会用第二种方法，从这 9 个人中选 1 人完成这项工作，一共有多少种选法？
2. 有 10 本不同的数学书，9 本不同的语文书，8 本不同的英语书，从中取出数学书、语文书、英语书各一本，共有多少种取法？

习　题　1—1

A　组

1. 在 1，2，3，…，200 中，被 5 除余 1 的数一共有多少个？
2. 在所有的两位数中，个位数字比十位数字大的两位数有多少个？
3. 高二(1)班有学生 56 人，其中男生 38 人，从中选取 1 名男生和 1 名女生做代表，参加学校组织的调查团，则选取代表的方法有多少种？
4. 一个口袋内装有 5 个小球，另一个口袋内装有 4 个小球，所有这些小球的颜色互不相同，从两个口袋内分别取 1 个小球，有多少种取法？
5. 在平面直角坐标系中，确定若干点，点的横坐标取自集合 $P=\{1,2,3\}$，点的纵坐标取自集合 $Q=\{1,4,5,6\}$，这样的点有多少个？
6. 商店里有 15 种上衣，18 种裤子，某人要买 1 件上衣或 1 条裤子，共有多少种选法？若要买上衣、裤子各 1 件，共有多少种选法？

B 组

“渐升数”是指每一位数字比其左边的数字大的正整数(如 236),那么三位渐升数有多少个?其中比 516 大的三位渐升数有多少个?

§2　排　列

问题提出

在日常生活中我们经常遇到下面一些问题，这些问题有什么共同特征呢？

问题 1　3 名同学排成一行照相，有多少种排法？

方法 1　（枚举法）

把 3 名同学用 A,B,C 作为代号，于是有以下 6 种排法：

$A\,B\,C$　　$B\,C\,A$　　$C\,A\,B$

$A\,C\,B$　　$C\,B\,A$　　$B\,A\,C$

方法 2　（分步计数）

A,B,C 三人排成一行，可以看作将字母 A,B,C 顺次排入图1-3的方格中.

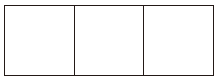


图 1-3

首先排第一个位置：从 A,B,C 中任选 1 人，有 3 种方法.

其次排第二个位置：从剩下的 2 个人中任选 1 人，有 2 种方法.

最后排第三个位置：只有 1 种方法.

根据乘法原理，3 名同学排成一行照相，共有 $3\times2\times1=6$ 种排法.

问题 2　北京、广州、南京、天津 4 个城市相互通航，应该有多少种机票？

方法 1　（枚举法）

列出每一个起点和终点情况，如图 1-4 所示：

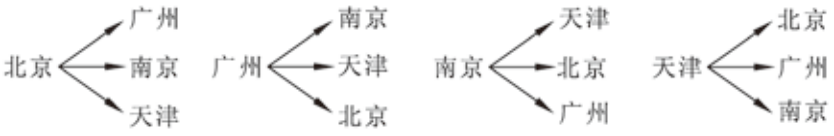


图 1-4

所以一共有 12 种机票.

方法 2　（分步计数）

我们按起始站、终点站的顺序进行排列：

第一步：先确定起始站，起始站有 4 种选择方法.

第二步：再确定终点站，对应于起始站的每一种选择，终点站都有 3 种选择方法.

根据乘法原理，共有 $4\times3=12$ 种机票.

问题 3 从 4 面不同颜色的旗子中，选出 3 面排成一排作为一种信号，能组成多少种信号？

分析 解决这个问题可以分三步进行.

第一步：先选第 1 面旗子，有 4 种选择方法.

第二步：在剩下的 3 种颜色中，再选第 2 面旗子，有 3 种选法.

第三步：在剩下的 2 种颜色中，选最后一面旗子，有 2 种选法.

根据乘法原理，共有 $4\times3\times2=24$ 种选法. 而每种选法对应一种信号，故共能组成 24 种信号.

抽象概括

一般地，从 n 个不同的元素中取出 $m(m\leqslant n)$ 个元素，按照一定顺序排成一列，叫作从 n 个不同的元素中任意取出 m 个元素的一个**排列**. 我们把有关求排列的个数的问题叫作**排列问题**.

在上面讨论的问题中，问题 1 是从 3 个不同元素中取出 3 个元素的排列问题，问题 2 是从 4 个不同元素中取出 2 个元素的排列问题，问题 3 是从 4 个不同元素中取出 3 个元素的排列问题.

练 习 1

1. 写出：
   (1) 从 4 个元素 a,b,c,d 中任取 3 个元素的所有排列；
   (2) 从 5 个元素 a,b,c,d,e 中任取 2 个元素的所有排列.
2. 从 6 名班委中选出 2 人分别担任正、副班长，一共有多少种选法？
3. 9 个人站成一排照相，其中甲必须站在左侧第一个位置，一共有多少种排法？

在前面的问题中，我们计算了几个排列问题，那么，对于一般的排列问题如何计算所有排列的个数呢？

我们把从 n 个不同的元素中任意取出 $m(m\leqslant n)$ 个元素的排列，看成从 n 个不同的球中选出 m 个球，放入排好的 m 个盒子中，每个

盒子里放一个球，我们用乘法原理排列这些球（见表 1-1）：

表 1-1

盒子	1	2	3	…	m
方法数	n	$n-1$	$n-2$	…	$n-(m-1)$

第 1 步：从全体 n 个球中选出一个放入第 1 个盒子，有 n 种选法.

第 2 步：从剩下的 $n-1$ 个球中选出一个放入第 2 个盒子，有 $n-1$ 种选法.

第 3 步：从剩下的 $n-2$ 个球中选出一个放入第 3 个盒子，有 $n-2$ 种选法.

……

第 m 步：从剩下的 $n-(m-1)$ 个球中选出一个放入第 m 个盒子，有 $n-(m-1)$ 种选法.

根据乘法原理，一共有 $n(n-1)(n-2)\cdots[n-(m-1)]$ 种放法.

这样，我们得到：从 n 个不同的元素中任意取出 $m(m\leqslant n)$ 个元素的排列一共有 $n(n-1)(n-2)\cdots(n-m+1)$ 种.

我们把从 n 个不同元素中取出 $m(m\leqslant n)$ 个元素的所有排列的个数，叫作从 n 个不同元素中取出 m 个元素的**排列数**，记作 A_n^m.

$$A_n^m=n(n-1)(n-2)\cdots(n-m+1).$$

规定 $A_n^0=1$. 当 $m=n$ 时，$A_n^n=n\cdot(n-1)\cdot(n-2)\cdot\cdots\cdot2\cdot1$.

我们把 $n\cdot(n-1)\cdot(n-2)\cdot\cdots\cdot2\cdot1$ 记作 $n!$，读作：n 的阶乘. 我们规定 $0!=1$.

我们可以对上面的公式进行变形：

$$\begin{aligned}A_n^m&=n(n-1)(n-2)\cdots(n-m+1)\\&=\frac{n\cdot(n-1)\cdot(n-2)\cdot\cdots\cdot(n-m+1)\cdot(n-m)\cdot(n-m-1)\cdot\cdots\cdot2\cdot1}{(n-m)\cdot(n-m-1)\cdot\cdots\cdot2\cdot1}\\&=\frac{n!}{(n-m)!}.\end{aligned}$$

$$A_n^m=\frac{n!}{(n-m)!}.$$

例 1　计算下列排列数：

（1）A_{50}^3；　（2）A_{15}^3；　（3）A_5^5；　（4）A_6^6.

解　（1）$A_{50}^3=50\times49\times48=117\ 600$；

（2）$A_{15}^3=15\times14\times13=2\ 730$；

（3）$A_5^5=5!=120$；

(4) $A_6^6 = 6! = 720$.

例 2 利用 1,2,3,4 这四个数字,可以组成多少个没有重复数字的三位数?

解 这是从 1,2,3,4 四个数字中,任意选出三个数字排成一排,有多少种排法的排列问题. 故有 $A_4^3 = 4\times3\times2 = 24$ 种排法.

所以,用 1,2,3,4 这四个数字,可以组成 24 个没有重复数字的三位数.

例 3 从某班 50 名学生中选出 6 名学生分别担任 6 个小组的组长,有多少种可能?

解 从 50 名学生中选出 6 名学生担任 6 个小组的组长,相当于从 50 个不同元素中任选 6 个元素进行排列的问题.

所以,一共有

$$A_{50}^6 = 50\times49\times48\times47\times46\times45 = 11\ 441\ 304\ 000$$

种可能.

例 4 有红、黄、蓝 3 种颜色的旗子各一面,如果用它们其中的若干面挂在一个旗杆上发出信号,那么一共可以组成多少种信号?

分析 旗杆上可以挂 1 面旗子,也可以挂 2 面、3 面旗子,因此,需要分类计数. 由于挂出的旗子顺序不同表示的信号也不同,因此,对每一类来说是一个排列问题.

解 第一类:旗杆上挂 1 面旗子,可以组成 A_3^1 种信号.

第二类:旗杆上挂 2 面旗子,可以组成 A_3^2 种信号.

第三类:旗杆上挂 3 面旗子,可以组成 A_3^3 种信号.

根据加法原理,一共可以组成

$$A_3^1 + A_3^2 + A_3^3 = 3 + 3\times2 + 3\times2\times1 = 15$$

种信号.

练 习 2

1. 计算:

(1) A_{15}^4; (2) A_6^6; (3) $A_8^4 - 2A_8^2$; (4) $\dfrac{A_{12}^8}{A_{12}^7}$.

2. 计算下表中的阶乘数,并填入表中:

n	1	2	3	4	5	6	7	8
$n!$								

3. 选择题：

(1) $18\times17\times16\times\cdots\times9\times8$ 等于(　　).

A. A_{18}^{8}　　B. A_{18}^{9}　　C. A_{18}^{10}　　D. A_{18}^{11}

(2) 下列各式中不等于 $n!$ 的是(　　).

A. A_n^n　　B. $\dfrac{1}{n+1}A_{n+1}^{n+1}$　　C. A_{n+1}^{n}　　D. nA_{n-1}^{n-1}

习 题 1—2

A 组

1. 计算：

(1) $5A_5^3+4A_4^2$；　　(2) $A_4^1+A_4^2+A_4^3+A_4^4$.

2. 从 6 名志愿者中选出 4 人分别从事翻译、导游、导购、保洁 4 项工作，选派方案共有多少种？

3. A,B,C,D,E 5 人站成一排，如果 A,B 必须相邻且 B 在 A 的右边，那么排法种数有(　　).

A. 60 种　　B. 48 种　　C. 36 种　　D. 24 种

4. 从黄瓜、白菜、油菜、扁豆 4 种蔬菜品种中选出 3 种，分别种在不同土质的 3 块土地上，其中黄瓜必须种植，种植方法共有多少种？

5. 已知 5 个不同的元素 a,b,c,d,e 排成一排.

(1) a,e 相邻有多少种排法？

(2) a,e 不相邻有多少种排法？

B 组

1. 一块并排 10 垄的田地中，选择 2 垄分别种植 A,B 两种作物，每种作物种植一垄. 为有利于作物生长，要求 A,B 两种作物的间隔不小于 6 垄，选垄方法共有多少种？

2. 甲、乙、丙、丁、戊 5 名学生进行某种劳动技术比赛，决出了第 1 名到第 5 名的名次. 甲、乙两名参赛者去询问成绩. 回答者对甲说“很遗憾，你和乙都未拿到冠军”；对乙说“你当然不是最差的”. 试从这个回答中分析 5 人的名次排列共有多少种情况？

§3 组 合

问题提出

在日常生活中我们经常遇到下面一些问题，这些问题有什么共同特征？它们与排列问题有什么不同吗？

问题 1 某城市有 3 个大型体育场 A,B,C，需要选择 2 个体育场承办一次运动会，有多少种选择方案？

分析 利用枚举法

我们把所有可能都列出来，一共有 AB,AC,BC 3 种，因此有 3 种选择方案.

问题 2 从 a,b,c,d 4 个元素中取出 2 个元素，共有多少种可能？

分析 设取法的总数为 C，其中每一种取法是 a,b,c,d 中的 2 个元素，例如：a,b.

这 2 个元素，可以组成 2 种不同的排列.

这样，就可以分两步来计算“从 4 个不同元素中，任取出 2 个元素”的排列问题.

第一步：先从 4 个元素中取出 2 个元素，总数为 C.

第二步：将取出的 2 个元素进行排列，排列数为 2.

根据乘法原理，$A_4^2=C\times 2$，从而 $C=\dfrac{A_4^2}{2}=\dfrac{4\times 3}{2}=6$.

问题 3 某次团代会，要从 5 名候选人 a,b,c,d,e 中选出 3 人担任代表，共有多少种方案？

分析 我们先考虑从 5 名候选人中选出 3 人的排列数 A_5^3.

其中每一个排列都由 3 人按一定顺序组成，例如：abc 就是 a,b,c 3 人的一种排列. 这 3 人共有 $3!=6$ 种排列，但是，这些不同排列都是这 3 人组成的.

我们可以分两步计算“从 5 人中选取 3 人”的排列问题.

第一步：从 5 名候选人中选取 3 人，设其总数为 C.

第二步：将 3 人进行排列，总数为 A_3^3.

这样，$A_5^3 = C \times A_3^3$. 所以，$C = \dfrac{A_5^3}{A_3^3} = \dfrac{5\times4\times3}{3\times2\times1} = 10$.

即，从 5 人中选出 3 人的方案总数为 10 种.

抽象概括

一般地，从 n 个不同的元素中，任取 $m(m \leqslant n)$ 个元素为一组，叫作从 n 个不同元素中取出 m 个元素的一个**组合**. 我们把有关求组合的个数的问题叫作**组合问题**.

例如，在上面的三个问题中，求“从 3 个体育场中选出 2 个承办运动会”的方案个数问题，就是从 3 个不同的元素中选出 2 个元素的组合问题；求“从 a,b,c,d 4 个元素中取出 2 个元素”的取法问题，就是从 4 个不同的元素中选出 2 个元素的组合问题；求“从 5 名候选人中选出 3 人担任代表”的方案个数问题，就是从 5 个不同的元素中选出 3 个元素的组合问题.

练　习　1

1. 写出：
   (1) 从 4 个元素 a,b,c,d 中任取 3 个元素的所有组合；
   (2) 从 5 个元素 a,b,c,d,e 中任取 2 个元素的所有组合.
2. (1) 从 6 个候选人中选出 3 个人作为职工大会代表，一共有多少种选择方法？
   (2) 从 6 个候选人中选出 3 个人分别担任一、二、三车间的车间主任，一共有多少种选择方法？

对于一个组合问题如何计算组合的个数呢？

我们把从 n 个不同元素中取出 $m(m \leqslant n)$ 个元素的所有组合的个数，叫作从 n 个不同元素中取出 m 个元素的**组合数**，用符号 C_n^m 表示.

组合与排列都是从 n 个不同元素中取出 $m(m \leqslant n)$ 个元素的计数问题，它们的差别是：排列考虑元素顺序，组合不考虑元素顺序. 前面我们已经学习了如何计算排列数，下面，我们看一看能否通过排列数来计算组合数.

先看一个简单情况：从 3 个元素 a,b,c 中任取 2 个元素的组合有 ab,ac,bc 3 种情况，再对每一种组合的 2 个元素进行排列，这样，就可以得到从 3 个元素中取 2 个元素的所有排列(见表 1-2).

表 1-2

组　合	排　列	
ab	ab	ba
ac	ac	ca
bc	bc	cb

从上面的分析可以看出，“从 3 个不同的元素中选出 2 个元素进行排列”这件事，可以分两步进行：

第一步：从 3 个不同元素中取出 2 个元素，一共有 C_3^2 种取法.

第二步：把取出的 2 个元素进行排列，一共有 A_2^2 种排法.

根据乘法原理，我们得到“从 3 个不同的元素中选出 2 个元素进行排列”一共有 $C_3^2 \cdot A_2^2$ 种排法，即 $A_3^2 = C_3^2 \cdot A_2^2$.

由此我们可以得出：

$$C_3^2 = \frac{A_3^2}{A_2^2} = \frac{3\times 2}{2!}.$$

一般地，考虑 C_n^m 与 A_n^m 的关系：把“从 n 个不同的元素中选出 m $(m \leqslant n)$ 个元素进行排列”这件事，分两步进行：

第一步：从 n 个不同元素中取出 m 个元素，一共有 C_n^m 种取法.

第二步：把取出的 m 个元素进行排列，一共有 A_m^m 种排法.

根据乘法原理，我们得到“从 n 个不同的元素中选出 $m(m \leqslant n)$ 个元素进行排列”一共有 $C_n^m \cdot A_m^m$ 种排法，即 $A_n^m = C_n^m \cdot A_m^m$.

由此我们可以得出：

$$C_n^m = \frac{A_n^m}{A_m^m} = \frac{n(n-1)(n-2)\cdots(n-m+1)}{m!}.$$

规定 $C_n^0 = 1$，上述这个公式叫作**组合数公式**.

因为 $A_n^m = \dfrac{n!}{(n-m)!}$，所以上面的组合数公式还可以写成：

$$C_n^m = \frac{n!}{m!\,(n-m)!}.$$

例 1　计算：

(1) C_{10}^4；　　(2) C_7^3.

解　(1) $C_{10}^4 = \dfrac{10\times 9\times 8\times 7}{4\times 3\times 2\times 1} = 210$.

(2) $C_7^3 = \dfrac{7\times 6\times 5}{3\times 2\times 1} = 35$.

例 2　平面内有 12 个点，任何 3 点不在同一直线上，以每 3 点为顶点画一个三角形，一共可以画多少个三角形？

分析　已知“任何 3 点不在同一直线上”，所以在 12 个点中任取 3 个点都可以构成一个三角形，且这 3 个点不必考虑顺序，例如：$\triangle ABC$，$\triangle ACB$，$\triangle BCA$，$\triangle BAC$，$\triangle CAB$，$\triangle CBA$ 都表示同一个三角形. 因此，这是一个组合问题.

解　以平面内 12 个点中的每 3 个点为顶点画三角形，可画的三角形的个数，就是从 12 个不同元素中取出 3 个元素的组合数，即 $C_{12}^3=\dfrac{12\times11\times10}{3\times2\times1}=220$.

因此，一共可以画 220 个三角形.

练　习　2

1. 计算：
   (1) C_6^2；　(2) C_8^3；　(3) $C_7^3-C_6^2$；　(4) $3C_8^3-2C_5^2$.
2. 6 个人聚会，每两人握一次手，一共握多少次手？
3. 学校开设了 6 门选修课，要求每个学生从中选学 3 门，共有多少种选法？

我们来研究下面的两个问题：

问题 4　计算“从 10 人中选出 6 人参加比赛”与“从 10 人中选出 4 人不参加比赛”的方法数.

分析　每次选出 6 人都相当于剩下 4 人，所以，选出 6 人参加比赛和选出 4 人不参加比赛的方法数是一样的. 即 $C_{10}^6=C_{10}^4$.

一般地，组合数有如下性质.

性质 1　$C_n^m=C_n^{n-m}$.

证明　从 n 个元素中选出 m 个元素，相当于从 n 个元素中留下 $n-m$ 个元素，因此有 $C_n^m=C_n^{n-m}$.

这个性质我们也可以利用组合数公式来进行验证：

因为　$$C_n^m=\frac{n!}{m!(n-m)!},$$

$$C_n^{n-m}=\frac{n!}{(n-m)![n-(n-m)]!}=\frac{n!}{m!(n-m)!},$$

所以　$$C_n^m=C_n^{n-m}.$$

问题 5 从 10 名战士和 1 名班长这 11 人中选出 5 人参加比武，一共有多少种方案？

分析 一方面，从 11 人中选出 5 人参加比武，一共有 C_{11}^{5} 种方案.

另一方面，选出的 5 人可以分为两类：

第一类：含有班长，一共有 C_{10}^{4} 种方案.

第二类：不含班长，一共有 C_{10}^{5} 种方案.

根据加法原理，一共有 $C_{10}^{4}+C_{10}^{5}$ 种方案.

由此，我们得到 $C_{11}^{5}=C_{10}^{4}+C_{10}^{5}$.

一般地，组合数还有如下性质.

性质 2 $C_{n+1}^{m}=C_{n}^{m}+C_{n}^{m-1}$.

证明 一方面，C_{n+1}^{m} 表示从 $n+1$ 个不同的元素中取出 m 个元素的组合数.

另一方面，假设这互不相同的 $n+1$ 个元素中有一个元素是 a，我们按照取出的 m 个元素中含不含元素 a，把这 m 个元素的组合分成两类：

第一类：取出的 m 个元素中不含元素 a，这相当于从不含 a 的 n 个不同的元素中取出 m 个元素，一共有 C_{n}^{m} 种取法.

第二类：取出的 m 个元素中含有元素 a，这相当于从不含 a 的 n 个不同的元素中取出 $m-1$ 个元素，一共有 C_{n}^{m-1} 种取法.

根据加法原理，一共有 $C_{n}^{m}+C_{n}^{m-1}$ 种取法.

由此，我们得到 $C_{n+1}^{m}=C_{n}^{m}+C_{n}^{m-1}$.

同样，我们也可以利用组合数公式，来验证这个性质，请大家试试看.

例 3 计算：

(1) C_{100}^{98}；　(2) $C_{99}^{96}+C_{99}^{97}$.

解 (1) $C_{100}^{98}=C_{100}^{2}=\dfrac{100\times 99}{2}=4\ 950$.

(2) $C_{99}^{96}+C_{99}^{97}=C_{99}^{3}+C_{99}^{2}=C_{100}^{3}=\dfrac{100\times 99\times 98}{3\times 2\times 1}=161\ 700$.

练 习 3

1. 计算：

（1）C_{20}^{17}；　　（2）C_{100}^{98}.

2. $C_{12}^{5}+C_{12}^{6}$等于(　　).

A. C_{13}^{5}　　B. C_{13}^{6}　　C. A_{13}^{11}　　D. A_{12}^{7}

习 题 1—3

A 组

1. 计算：

（1）C_{5}^{2}；　　（2）C_{7}^{4}；　　（3）C_{10}^{5}.

2. 计算：

（1）C_{15}^{3}；　　（2）C_{200}^{197}；　　（3）$\dfrac{C_{8}^{3}}{C_{8}^{4}}$；　　（4）$C_{n+1}^{n}\cdot C_{n}^{n-2}$.

3. 求证：$C_{7}^{3}+C_{7}^{4}+C_{8}^{5}=C_{9}^{5}$.

4. 圆上有 10 个点：

（1）过每 2 个点画一条弦，一共可画多少条弦？

（2）过每 3 个点画一个圆内接三角形，一共可画多少个圆内接三角形？

5. 从 1，3，5，7，9 中任取 3 个数字，从 2，4，6，8 中任取 2 个数字，一共可以组成多少个没有重复数字的五位数？

B 组

1. 甲、乙、丙、丁 4 个公司承包 8 项工程，甲公司承包 3 项，乙公司承包 1 项，丙、丁公司各承包 2 项，则共有多少种承包方式？

2. 某校乒乓球队有男运动员 10 名和女运动员 9 名，若要选出男、女运动员各 3 名参加三场混合双打比赛（每名运动员只限参加一场比赛），共有多少种参赛方法？

3. 计算：$\dfrac{C_{100}^{2}+C_{100}^{97}}{A_{101}^{3}}$.

4. 计算：$C_{3}^{3}+C_{4}^{3}+\cdots+C_{10}^{3}$.

§4　简单计数问题

在这一节中，我们利用前面学习的方法解决一些简单的计数问题.

例 1　(1) 5 个相同的球，放入 8 个不同的盒子中，每盒至多放一个球，共有多少种放法？

(2) 5 个不同的球，放入 8 个不同的盒子中，每盒至多放一个球，共有多少种放法？

解　(1) 由于球都相同，盒子不同，每盒至多放一个球，所以，只要选出 5 个不同的盒子，就可以解决问题. 这是一个组合问题.

因此，5 个相同的球，放入 8 个不同的盒子中，每盒至多放一个球，共有 $C_8^5=56$ 种放法.

(2) 方法一：由于球与盒子均不同，每盒至多放一个球，所以这是一个排列问题. 可直接从 8 个不同的盒子中取出 5 个盒子进行排列(即放球)，所以，共有 $A_8^5=8\times7\times6\times5\times4=6\ 720$ 种放法.

方法二：由于每盒至多放一个球，所以，第 1 个球有 8 种放法，第 2 个球有 7 种放法，……，第 5 个球有 4 种放法. 所以，共有 $A_8^5=8\times7\times6\times5\times4=6\ 720$ 种放法.

例 2　在 100 个零件中有 80 个正品、20 个次品，从中任意选 2 个进行检测，其中至少有一个次品的选法有多少种？

分析　2 个零件中至少有一个次品的情况有两种：只有一个次品或两个都是次品. 由于次品不加区别，这是一个组合问题.

解　分类计数.

第一类：只有一个次品，另一个是正品，有 $C_{80}^1\cdot C_{20}^1$ 种选法.

第二类：两个都是次品，有 $C_{80}^0\cdot C_{20}^2$ 种选法.

根据加法原理，其中至少有一个次品的选法共有

$$C_{80}^1\cdot C_{20}^1+C_{80}^0\cdot C_{20}^2=80\times20+1\times\frac{20\times19}{2\times1}=1\ 790$$

种选法.

例 3　某项化学实验，要把 2 种甲类物质和 3 种乙类物质按照先放甲类物质后放乙类物质的顺序，依次放入某种液体中，观察反应结果. 现有符合条件的 3 种甲类物质和 5 种乙类物质可供使用. 问：这

个实验一共要进行多少次，才能得到所有的实验结果？

分析　由于要把 2 种甲类物质和 3 种乙类物质按照先甲后乙的顺序依次放入某种液体中，因此，需要分步计数. 由于同一类物质不同的放入顺序，反应结果可能会不同，因此，这是一个排列问题.

解　第一步：放入甲类物质，共有 A_3^2 种方案.

第二步：放入乙类物质，共有 A_5^3 种方案.

根据乘法原理，共有 $A_3^2 \cdot A_5^3 = 3\times2\times5\times4\times3 = 360$ 种方案.

因此，共要进行 360 次实验，才能得到所有的实验结果.

例 4　将 5 个不同的元素 a,b,c,d,e 排成一排.

(1) a,e 必须排在首位或末位，有多少种排法？

(2) a,e 既不在首位也不在末位，有多少种排法？

(3) a 不排在首位，e 不排在末位，有多少种排法？

解　(1) 按首位是 a 还是 e 分类计数.

第一类：a 排在首位，那么 e 必须排在末位，中间三位是把 b,c,d 进行排列，一共有 $A_3^3 = 3\times2\times1 = 6$ 种排法.

第二类：e 排在首位，那么 a 必须排在末位，中间三位是把 b,c,d 进行排列，一共有 $A_3^3 = 3\times2\times1 = 6$ 种排法.

根据加法原理，a,e 必须排在首位或末位，一共有 $6+6=12$ 种排法.

(2) 按照先排首位和末位，再排中间三位分步计数.

第一步：排出首位和末位.

由于 a,e 既不在首位也不在末位，那么首位和末位是在 b,c,d 中选出两个进行排列，一共有 $A_3^2 = 3\times2 = 6$ 种排法.

第二步：排出中间三位.

由于在 a,b,c,d,e 5 个元素中，已经有 2 个元素排在了首位和末位，因此，中间三位是把剩下的 3 个元素进行排列，一共有 $A_3^3 = 3\times2\times1 = 6$ 种排法.

根据乘法原理，a,e 既不在首位也不在末位，一共有 $6\times6 = 36$ 种排法.

(3) 按照 a 是否排在末位分类计数.

第一类：a 排在末位，此时 e 不排在末位，故一共有 $A_4^4 = 4\times3\times2\times1 = 24$ 种排法.

第二类：a 不排在末位，此时可按照先排 a，再排 e，最后排 b,c,d 分步计数：

第一步：a 排在中间，有 $A_3^1 = 3$ 种排法.

第二步：e 排在除末位及 a 所占位置外的其余位置，有 $A_3^1 = 3$ 种

排法.

第三步:b,c,d 排在其余位置,有 $A_3^3=3\times2\times1=6$ 种排法.

根据乘法原理,第二类有 $3\times3\times6=54$ 种排法.

最后,根据加法原理,a 不排在首位,e 不排在末位,一共有 $24+54=78$ 种排法.

练 习 1

1. 12 名新战士,每人有一个储物箱,每个箱子有一把钥匙,但是钥匙上没有标记箱子号码,班长要想把所有的箱子打开最多要试多少次?
2. 某列火车往返于甲地和乙地之间,中途共停站 5 次,一共要设计几种车票?
3. 某校有 A,B 两个科技活动小组,每组有 12 名学生,其中有 4 名学生两个小组都参加. 现要从这两个小组中选出 3 人作为代表,参加少年宫科技活动大赛,则共有多少种选法?

例 5 用 0,1,2,…,9 这 10 个数字,

(1) 可以组成多少个 5 位数?

(2) 可以组成多少个没有重复数字的 5 位数?

(3) 可以组成多少个没有重复数字且能够被 5 整除的 5 位数?

解 (1)第一步:首位数字可以在 1～9 这 9 个数字中选取,有 9 种可能.

第二步:其他 4 个数位可以在 0～9 这 10 个数字中选取,由乘法原理有 $10\times10\times10\times10=10^4$ 种可能.

最后,根据乘法原理,用 0,1,2,…,9 这 10 个数字,一共可以组成 $9\times10^4=90\ 000$ 个 5 位数.

(2) 由于组成 5 位数中不能有重复数字,所以除了要考虑到首位不是 0 以外,还要考虑到各个数位上的数字互不相同,因此,采用分步计数的方法,先确定首位数字再确定其他数位.

第一步:首位数字,可以在 1～9 这 9 个数字中选择,有 9 种可能.

第二步:其他 4 个数位,可以在剩下的 9 个数字中选择,有 A_9^4 种可能.

根据乘法原理,用 0,1,2,…,9 这 10 个数字,一共可以组成 $9\cdot A_9^4=27\ 216$ 个没有重复数字的 5 位数.

(3) 能够被 5 整除的数,末位有且仅有 0 或 5 两种可能,分两类进行计数.

第一类:末位是 0,由于没有重复数字,所以其他 4 个数位共有 A_9^4 种可能.

第二类:末位是 5,对其他 4 个数位进行分步计数:

第一步：由于首位不能为 0，首位有 C_8^1 种选择.

第二步：其他 3 个数位有 A_8^3 种可能.

所以，共有 $A_9^4+C_8^1\cdot A_8^3=5\ 712$ 种.

因此，用 0，1，2，…，9 这 10 个数字，一共可以组成 5 712 个没有重复数字且能够被 5 整除的 5 位数.

例 6　某批产品中有一等品 100 个，二等品 80 个，三等品 30 个. 从其中任取 10 个进行检验，那么：

(1) 一共有多少种抽取结果？

(2) 全部抽到一等品的结果有多少种？

(3) 抽不到一等品的结果有多少种？

(4) 恰抽到 5 个一等品的结果有多少种？

(5) 恰抽到 1 个一等品、2 个二等品的结果有多少种？

(6) 至少抽到 1 个一等品的结果有多少种？

解　(1) 这批产品一共有 100＋80＋30＝210 个，从其中任取 10 个进行检验，共有 C_{210}^{10} 种抽取结果.

(2) 这批产品中有一等品 100 个，取出 10 个一等品，共有 C_{100}^{10} 种抽取结果.

(3) 抽不到一等品，相当于从二等品和三等品中抽取 10 个进行检验. 二等品和三等品共有 80＋30＝110 个，所以，抽不到一等品的结果共有 C_{110}^{10} 种.

(4) 恰抽取 5 个一等品，剩下的 5 个产品从二等品和三等品中抽取.

分步计数：先抽取 5 个一等品，再抽取 5 个非一等品.

根据乘法原理，一共有 $C_{100}^5\cdot C_{110}^5$ 种抽取结果.

(5) 恰抽到 1 个一等品、2 个二等品，剩下的 7 个产品从三等品中抽取.

分步计数：先抽取 1 个一等品，再抽取 2 个二等品，最后抽取 7 个三等品.

根据乘法原理，一共有 $C_{100}^1\cdot C_{80}^2\cdot C_{30}^7$ 种抽取结果.

(6) “抽取的 10 个产品中，至少有 1 个一等品”，是“没有抽到一等品”的反面，因此，用所有的抽取结果数，减去没有抽到一等品的结果数即可.

所以，至少抽到 1 个一等品的结果共有 $C_{210}^{10}-C_{110}^{10}$ 种.

练 习 2

1. 某班新年联欢会原定的 5 个节目已排成节目单，开演前又增加了 2 个新节目，如果将这 2 个新节目插入原节目单中且不相邻，那么有多少种方法？
2. 某学校有 6 个电脑教室，每周一、周三、周五的晚上向学生开放 2 个教室，那么一共有多少种开放方案？
3. 5 件不同的礼品分送给 4 个人，每个人至少得到 1 件礼品，而且礼品必须全部送出，那么送礼品的方法一共有多少种？
4. 校学生会成立了物理、生物两个课外活动组，由于两个活动组的活动时间相同，所以每名同学只能参加一个活动组．甲、乙、丙 3 名同学刚刚完成了报名，他们可能有多少种报名结果呢？

习 题 1—4

A 组

1. 3 个不同的球放入 5 个不同的盒子，每个盒子至多放 1 个球，共有多少种放法？
2. 3 个不同的球放入 5 个不同的盒子，每个盒子放球数量不限，共有多少种放法？
3. 6 名同学排成一排，其中甲、乙两人相邻的排法有多少种？
4. 从 1，2，3，4，7，9 中任取不相同的两个数，分别作为对数的底数和真数，能得到多少个对数值？
5. 已知集合 $M=\{1,-1,3\}$，$N=\{-4,5,6,-7\}$，从两个集合中各取一个元素作为点的坐标，则这样的坐标在直角坐标系中可表示第一、第二象限内多少个点？
6. 从数字 0，1，2，3，4，5，6 中，任取 3 个不同的数字作为系数 a，b，c 的值，组成二次函数 $y=ax^2+bx+c$，则一共可以组成多少个解析式？
7. 某一天的课程表要排入政治、语文、数学、物理、体育、美术共 6 节课，如果第一节不排体育，最后一节不排数学，那么共有多少种排法？
8. 将 5 封信投入 3 个邮筒，有多少种投法？
9. 乘积 $(a_1+a_2+a_3)(b_1+b_2+b_3+b_4)(c_1+c_2+c_3+c_4+c_5)$ 展开后共有多少项？

B 组

1. 某校刊设有 9 门文化课专栏，由甲、乙、丙 3 名同学每人负责 3 个专栏，其中数学专栏必须由甲负责，则共有多少种分工方法？
2. 将 3 种作物全部种植在如图的 5 块试验田里，每块试验田种植一种作物且相邻的试验田不能种植同一种作物，则共有多少种种植方法？

（第 2 题）

3. 某城市的电话号码是由 8 个数字组成的，那么：
   （1）理论上该城市最多可以有多少个电话号码？（提示：电话号码各数位可重复）
   （2）理论上该城市首位为 8 的电话号码最多可以有多少个？

§5　二项式定理

5.1　二项式定理

问题提出

问题　计算乘积 $(a_1+b_1)(a_2+b_2)(a_3+b_3)$，并思考乘积中的每一项是怎样得到的.

分析理解

我们不难得到

$$\begin{aligned}&(a_1+b_1)(a_2+b_2)(a_3+b_3)\\=\ &a_1a_2a_3+a_1a_2b_3+a_1b_2a_3+a_1b_2b_3+\\&b_1a_2a_3+b_1a_2b_3+b_1b_2a_3+b_1b_2b_3.\end{aligned}$$

把每一项的产生过程填入表 1-3，并观察乘积的特点.

表 1-3

	从第一个因式 a_1+b_1 中取	从第二个因式 a_2+b_2 中取	从第三个因式 a_3+b_3 中取	乘　积
1	a_1	a_2	a_3	$a_1a_2a_3$
2	a_1	a_2	b_3	$a_1a_2b_3$
3	a_1	b_2	a_3	$a_1b_2a_3$
4	a_1	b_2	b_3	$a_1b_2b_3$
5	b_1	a_2	a_3	$b_1a_2a_3$
6	b_1	a_2	b_3	$b_1a_2b_3$
7	b_1	b_2	a_3	$b_1b_2a_3$
8	b_1	b_2	b_3	$b_1b_2b_3$

我们发现：

1. 3 个因式相乘的结果一共有 8 项，每一项都是 3 个字母的乘积，这 3 个字母分别来自 3 个不同的因式.

2. 所有的项包含了 3 个因式中各选 1 个字母相乘的所有可能情况.

当 $a_1=a_2=a_3=a$，$b_1=b_2=b_3=b$ 时，以上 3 个因式相乘结果的 8 项中有同类项，合并同类项后，得到 4 个不同的项：

a^3 ——只有 1 项，相当于从 3 个因式中都不取 b 只取 a，即 $C_3^0=1$；

a^2b ——有 3 项，相当于从 3 个因式中的 1 个因式中取 b，其余 2 个因式中取 a，即 $C_3^1=3$；

ab^2 ——有 3 项，相当于从 3 个因式中的 2 个因式中取 b，其余 1 个因式中取 a，即 $C_3^2=3$；

b^3 ——只有 1 项，相当于从 3 个因式中都取 b，即 $C_3^3=1$.

因此，得到 $(a+b)^3=a^3+3a^2b+3ab^2+b^3$

$$=C_3^0a^3+C_3^1a^2b+C_3^2ab^2+C_3^3b^3.$$

考虑 $(a+b)^n$ 的展开式情况：

1. $(a+b)^n$ 是 n 个因式 $a+b$ 的乘积，展开时在每个因式中可取 a 或 b，有两种取法，由乘法原理，共有 2^n 种取法，所以其展开后的式子共有 2^n 项. 其中，每一项都可以表示为 $a^{n-r}b^r(r=0,1,\cdots,n)$ 的形式.

2. 在已展开的式子中，考虑 $a^{n-r}b^r$ 的同类项，由于 $a^{n-r}b^r$ 是由 $n-r$ 个因式 $a+b$ 中选 a，r 个因式 $a+b$ 中选 b 得到的，且 b 选定后，a 的选法也随之确定，因此，$a^{n-r}b^r$ 出现的次数相当于从 n 个因式 $a+b$ 中取 r 个 b 的组合数 C_n^r. 所以，合并同类项之后，$a^{n-r}b^r$ 的系数是 C_n^r. 于是，

$$(a+b)^n=C_n^0a^n+C_n^1a^{n-1}b+\cdots+C_n^ra^{n-r}b^r+\cdots+C_n^nb^n.$$

这个公式称为**二项式定理**，等号右边的式子称为 $(a+b)^n$ 的**二项展开式**，$(a+b)^n$ 的二项展开式共有 $n+1$ 项，其中各项的系数 $C_n^r(r=0,1,2,\cdots,n)$ 称为**二项式系数**，$C_n^ra^{n-r}b^r$ 称为二项展开式的第 $r+1$ 项，又称为**二项式通项**.

在二项式定理中，如果设 $a=1$，$b=x$，则得到公式：

$$(1+x)^n=1+C_n^1x+C_n^2x^2+\cdots+C_n^rx^r+\cdots+x^n.$$

例 1 展开 $(x+2)^5$.

解 $(x+2)^5=C_5^0x^52^0+C_5^1x^42^1+C_5^2x^32^2+$

$$C_5^3x^22^3+C_5^4x^12^4+C_5^5x^02^5$$

$=x^5+10x^4+40x^3+80x^2+80x+32.$

例 2　展开$\left(2+\frac{1}{x}\right)^4$.

解　$\left(2+\frac{1}{x}\right)^4$

$=C_4^0 2^4\left(\frac{1}{x}\right)^0+C_4^1 2^3\left(\frac{1}{x}\right)^1+C_4^2 2^2\left(\frac{1}{x}\right)^2+C_4^3 2\left(\frac{1}{x}\right)^3+C_4^4\left(\frac{1}{x}\right)^4$

$=16+\frac{32}{x}+\frac{24}{x^2}+\frac{8}{x^3}+\frac{1}{x^4}.$

例 3　展开$\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)^6$.

解　$\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)^6=\left(\frac{x-1}{\sqrt{x}}\right)^6=\frac{(x-1)^6}{x^3}$

$=\frac{1}{x^3}[C_6^0 x^6(-1)^0+C_6^1 x^5(-1)^1+C_6^2 x^4(-1)^2+$
$C_6^3 x^3(-1)^3+C_6^4 x^2(-1)^4+C_6^5 x^1(-1)^5+$
$C_6^6 x^0(-1)^6]$

$=\frac{1}{x^3}(x^6-6x^5+15x^4-20x^3+15x^2-6x+1)$

$=x^3-6x^2+15x-20+\frac{15}{x}-\frac{6}{x^2}+\frac{1}{x^3}.$

例 4　求$(x-2y)^6$展开式中的第 4 项，并指出该项的系数.

解　由二项式通项得，$(x-2y)^6$展开式中的第 4 项为

$$C_6^3 x^{(6-3)}\cdot(-2y)^3=-160\,x^3y^3,$$

这一项的系数是−160.

注　意

二项展开式中项的系数与二项式系数的区别.

练　习

1. 展开$(2x+y)^5$.
2. 展开$\left(\frac{1}{x}-1\right)^4$.
3. 求$\left(\frac{1}{x}-x\right)^6$展开式中第 2 项的系数.
4. 求$(a-2b)^{10}$展开式中的第 8 项.

5.2 二项式系数的性质

当 n 依次取 1,2,3,…时，$(a+b)^n$ 展开式的二项式系数如图 1-5 所示：

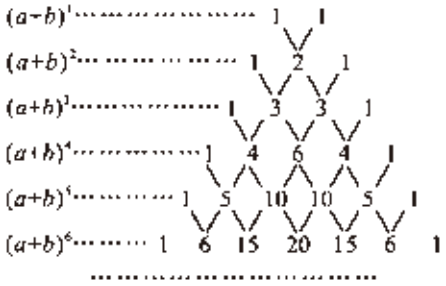


图 1-5

图 1-5 中所示的表叫作**二项式系数表**，它有这样的规律：表中每行两端都是 1，而且除 1 以外的每一个数都等于它“肩上”的两个数的和. 事实上，设表中任一不为 1 的数为 C_{n+1}^r，那么它“肩上”的两个数分别为 C_n^{r-1} 及 C_n^r，由组合数的性质 2 知道：

$$C_{n+1}^r = C_n^{r-1} + C_n^r.$$

早在 1261 年，我国南宋数学家杨辉的著作《详解九章算法》中就有类似的表（见图 1-6）：

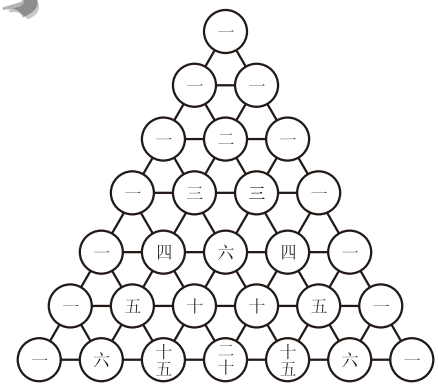


图 1-6

这个表称为**杨辉三角**.

结合“杨辉三角”可以得出**二项式系数的对称性**：与首末两端“等距离”的两个二项式系数相等.

这个性质可以由公式 $C_n^m = C_n^{n-m}$ 得到.

例 5 根据“杨辉三角”写出 $(a+b)^7$ 的二项式系数.

解 从图 1-6 所示的“杨辉三角”中知道，$(a+b)^6$ 的各个二项式系数为 1,6,15,20,15,6,1. 根据其规律，有

1　6　15　20　15　6　1

1　7　21　35　35　21　7　1

所以，$(a+b)^7$ 的二项式系数为 1，7，21，35，35，21，7，1.

这样，可以将二项式系数表延伸下去，从而可根据“杨辉三角”来求二项式系数.

例 6　求证：$2^n=C_n^0+C_n^1+\cdots+C_n^r+\cdots+C_n^n$.

证明　由二项式定理，有

$$(a+b)^n=C_n^0a^n+C_n^1a^{n-1}b+\cdots+C_n^ra^{n-r}b^r+\cdots+C_n^nb^n.$$

令 $a=b=1$，则

$$2^n=C_n^0+C_n^1+\cdots+C_n^r+\cdots+C_n^n.$$

练　习

1. 根据“杨辉三角”写出 $(a+b)^8$ 的二项式系数.
2. 根据“杨辉三角”求 $(5a+b)^{10}$ 展开式中的第 3 项的系数.

阅读材料

杨　辉

杨辉，字谦光，中国南宋时期杰出的数学家和数学教育家. 他著名的数学书共五种二十一卷. 著有《详解九章算法》十二卷、《日用算法》二卷、《乘除通变本末》三卷、《田亩比类乘除捷法》二卷、《续古摘奇算法》二卷.

杨辉的数学研究与教育工作的重点是在计算技术方面，他在《续古摘奇算法》中介绍了各种形式的“纵横图”及有关的构造方法，同时，“垛积术”是杨辉继沈括“隙积术”后，关于高阶等差级数的研究. 杨辉在“纂类”中，将《九章算术》246 个题目按解题方法由浅入深的顺序，重新分为乘除、分率、合率、互换、衰分、叠积、盈不足、方程、勾股九类.

早在 1261 年，杨辉在《详解九章算法》一书中提出了二项式系数的三角形排法，即著名的“杨辉三角”. 法国数学家帕斯卡在 17 世纪也建立了二项式系数的三角形表示法.

杨辉十分重视教育普及工作. 他编写的这几种书内容都较浅显，适合于教学之用或做普及读本.

习 题 1—5

A 组

1. 展开$(1+3x)^4$.

2. 展开$\left(x-\frac{1}{2}\right)^5$.

3. 展开$(\sqrt{x}+y)^5$.

4. 展开$\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)^5$.

5. 求$(1-x)^3$展开式的各项系数.

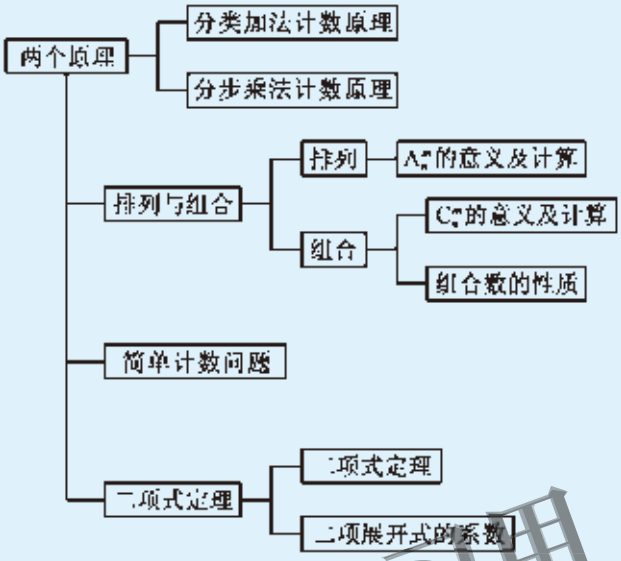
B 组

1. 求证:$C_n^0-C_n^1+C_n^2-C_n^3+\cdots+(-1)^nC_n^n=0$.

2. 用乘法原理求出$(a+b+c)^5$ 的项数.

◆ 本章小结建议

一、本章知识结构图



二、复习指导

1. 根据本章知识结构图回顾本章知识的主要内容.

2. 准确地叙述计数的两个基本原理：分类加法计数原理和分步乘法计数原理. 结合实例，思考“分类”和“分步”在两个计数原理中的作用.

3. 准确地叙述排列问题和组合问题. 结合具体实例，讨论它们的区别和联系.

4. 掌握排列公式和组合公式的计算.

5. 在证明组合数性质：$C_{n+1}^m = C_n^m + C_n^{m-1}$ 时，如何使用加法原理？

6. 结合具体实例，体会加法原理和乘法原理在解决简单计数问题中的作用.

7. 思考二项式定理的证明过程，进一步体会加法原理和乘法原理. 说明二项式系数的意义.

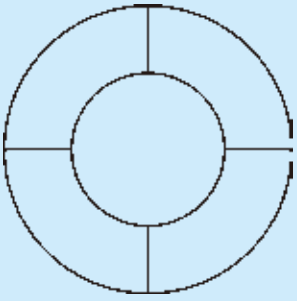
复 习 题 一

A 组

1. 写出 4 个元素 a，b，c，d 的所有排列.
2. 8 个乒乓球队每两个队比赛一场，一共要有多少场比赛？
3. 在某试验田中，分别对一种作物的用肥、用水量和温度进行实验. 用肥有 3 种选择，用水量有 3 种选择，温度控制有 2 种选择，则该试验田应该分成多少部分？
4. 要把 6 名农业技术员分到 3 个乡支援工作，甲乡需要 2 名，乙乡需要 3 名，丙乡需要 1 名，一共有多少种分配方案？
5. 2 名医生和 4 名护士被分配到 2 所学校为学生体检，每校分配 1 名医生和 2 名护士. 分配方法有多少种？
6. 从 5 名男生、3 名女生中选 4 名代表，至少有 1 名女生的选法有多少种？
7. 银行储蓄卡的密码是一个 4 位数，至少多少人中就会有两个人的密码相同？
8. 展开 $(x-2)^4$.

B 组

1. 6 名同学排成一排，其中甲、乙两人不相邻的排法有多少种？
2. 正六边形的中心和顶点共 7 个点，以其中 3 个点为顶点的三角形共有多少个？
3. 求以下问题的排列数：
   (1) 4 男 3 女排成一排，3 女相邻；
   (2) 4 男 3 女排成一排，3 女不能相邻；
   (3) 4 男 3 女排成一排，女不能排在两端；
   (4) 4 男 3 女，男女相间排成一排.
4. 如图，节日花坛中有 5 个区域，要把 4 种不同颜色的花分别种植到这 5 个区域中，要求相同颜色的花不能相邻栽种，一共有多少种种植方案？
5. 展开 $(2x+\sqrt{x})^4$.

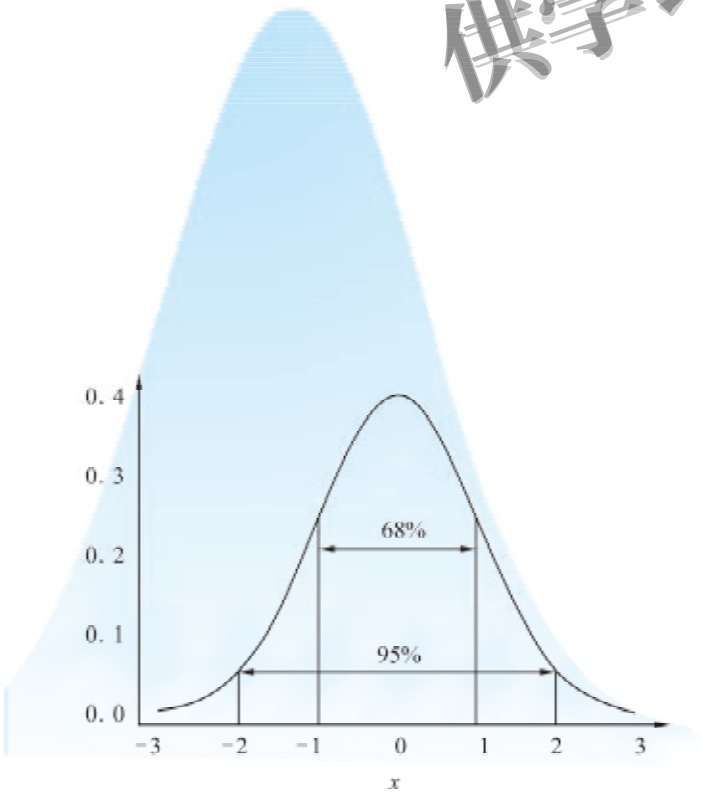


(第 4 题)

第二章 概 率

随机现象在日常生活中随处可见，寻找随机现象的规律并从数学上加以刻画是概率学习的主要目标.

在本章中，我们将在必修课程的基础上，学习利用随机变量描述和分析某些随机现象的方法；学习二项分布、超几何分布以及正态分布等重要的概率模型；学习离散型随机变量的均值、方差，并用所学知识解决一些简单的实际问题.



§1　离散型随机变量及其分布列

问题提出

在必修课程中，我们已经认识了大量的随机现象，了解了一些随机现象的规律. 那么，如何用数学语言来清楚地刻画每个随机现象的规律呢？

分析理解

了解一个随机现象的规律，是指了解这个随机现象中所有可能出现的结果及每个结果的概率. 例如，任意掷一枚均匀的骰子，所得到的点数是一个随机现象，了解这一随机现象的规律，既要知道所有可能出现的结果是“掷出 1 点”“掷出 2 点”“掷出 3 点”“掷出 4 点”“掷出 5 点”“掷出 6 点”，还要知道每个结果出现的概率都是$\frac{1}{6}$.

在数学中，一个常用的做法是把每个可能出现的结果对应于一个数，即用一个数来表示一个结果. 在上面的例子中，通常用 1 表示“掷出 1 点”，2 表示“掷出 2 点”……又如，在观测天气时，我们可以规定：1 表示晴天，2 表示阴天，3 表示下雨等，这种表示方法会给我们带来很大的方便.

抽象概括

我们将随机现象中试验（或观测）的每一个可能的结果都对应于一个数，这种对应称为一个**随机变量**，通常用大写的英文字母如 X，Y 来表示. 实际上，随机变量是从随机试验每一个可能的结果所组成的集合到实数集的映射. 例如，掷一枚均匀的骰子时，令 X 表示骰子掷出的点数，则 X 的所有可能取值 1，2，3，4，5，6，对应着掷骰子所有可能出现的结果. 即“$X=1$”表示“掷出 1 点”，“$X=2$”表示“掷出 2 点”……

例 1　已知在 10 件产品中有 2 件不合格品. 现从这 10 件产品中任取 3 件，这是一个随机现象.

(1)写出该随机现象所有可能出现的结果；

(2)试用随机变量来描述上述结果.

解 (1)这 10 件产品中有 2 件不合格品，有 8 件合格品.

因此，从 10 件产品中任取 3 件，所有可能出现的结果是：“不含不合格品”“恰有 1 件不合格品”“恰有 2 件不合格品”.

(2)令 X 表示取出的 3 件产品中的不合格品数.

则 X 所有可能的取值为 0，1，2，对应着任取 3 件产品所有可能出现的结果. 即

“$X=0$”表示“不含不合格品”；

“$X=1$”表示“恰有 1 件不合格品”；

“$X=2$”表示“恰有 2 件不合格品”.

例 2 连续投掷一枚均匀的硬币两次，用 X 表示这两次投掷中正面朝上的次数，则 X 是一个随机变量. 分别说明下列集合所代表的随机事件：

(1) $\{X=0\}$；　　(2) $\{X=1\}$；

(3) $\{X\leqslant 1\}$；　　(4) $\{X>0\}$.

解 (1)$\{X=0\}$ 表示使得随机变量对应于 0 的那些结果组成的事件，即两次都掷得反面朝上. 所以$\{X=0\}=\{$两次都是反面朝上$\}$.

(2)$\{X=1\}=\{$第一次正面朝上，第二次反面朝上$\}\cup\{$第一次反面朝上，第二次正面朝上$\}=\{$恰有一次正面朝上$\}$.

(3)$\{X\leqslant 1\}$ 表示使得随机变量 X 对应的数值不超过 1 的那些结果组成的事件，即$\{$至多一次正面朝上$\}$，也可以表示为：$\{X\leqslant 1\}=\{$两次都是反面朝上$\}\cup\{$第一次正面朝上，第二次反面朝上$\}\cup\{$第一次反面朝上，第二次正面朝上$\}$.

(4)$\{X>0\}=\{$第一次正面朝上，第二次反面朝上$\}\cup\{$第一次反面朝上，第二次正面朝上$\}\cup\{$两次都是正面朝上$\}=\{$至少一次正面朝上$\}$.

练　习

1. 用随机变量来描述随机现象可能的结果：

(1) 连续掷一枚均匀的硬币 3 次，正面朝上的次数；

(2) 一个口袋中装有除颜色外其他均相同的 8 个红球、3 个黄球，任意摸出两球，摸到黄球的个数.

2. 用 X 表示 10 次射击中命中目标的次数，分别说明下列集合所代表的随机事件：

(1) $\{X=8\}$；　　(2) $\{1<X\leqslant 9\}$；

(3) $\{X\geqslant 1\}$；　　(4) $\{X<1\}$.

3. 找出日常生活中的随机事件，并尝试用随机变量描述随机现象可能的结果.

对于随机试验我们引入了随机变量的概念. 这样，了解随机试验的规律就转化为了解随机变量的所有可能取值，以及随机变量取各个值的概率. 了解了上述两点，我们就可以说了解了这个随机试验的规律. 例如：用 X 表示投掷一枚均匀的骰子所得的点数，则 X 是一个随机变量，它的可能取值为 1，2，…，6. 由于掷得各点的概率相等，因而事件 $\{X=i\}$ 的概率为 $\frac{1}{6}$ $(i=1,2,\cdots,6)$，记作：

$$P(X=i)=\frac{1}{6} \qquad (i=1,2,\cdots,6).$$

知道了以上两点，投掷一枚均匀的骰子所得点数的规律也就弄明白了. 我们也常把上式列成表 2-1：

表 2-1

$X=i$	1	2	3	4	5	6
$P(X=i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

抽象概括

上面所举的随机变量的取值能够一一列举出来，这样的随机变量称为**离散型随机变量**.

我们设离散型随机变量 X 的取值为 a_1，a_2，…随机变量 X 取 a_i 的概率为 $p_i(i=1,2,\cdots)$，记作：

$$P(X=a_i)=p_i \quad (i=1,2,\cdots), \tag{1}$$

或把上式列成表 2-2：

表 2-2

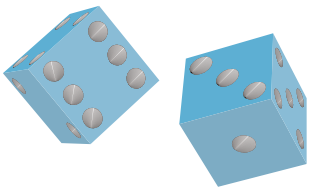
$X=a_i$	a_1	a_2	…
$P(X=a_i)$	p_1	p_2	…

表2-2或(1)式称为**离散型随机变量 X 的分布列**. 显然 $p_i>0$，$p_1+p_2+\cdots=1$.

如果随机变量 X 的分布列为表 2-2 或(1)式，我们称随机变量 X 服从这一分布(列)，并记为

$$X\sim\begin{pmatrix} a_1 & a_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}.$$

需要强调的是，随机变量 X 的分布列完全描述了随机现象的规律. 即当我们了解了随机变量 X 的分布(指(1)式中的 a_i，$p_i(i=1,2,\cdots)$都是已知数)时，就了解了这个随机变量的所有可能取值及取各个值的概率.



例 3 连续投掷一枚均匀的骰子两次，用 X 表示所得点数之和，试列出 X 的分布列.

解 连续投掷一枚均匀的骰子两次，共有如下 $6\times6=36$ 种可能的结果：

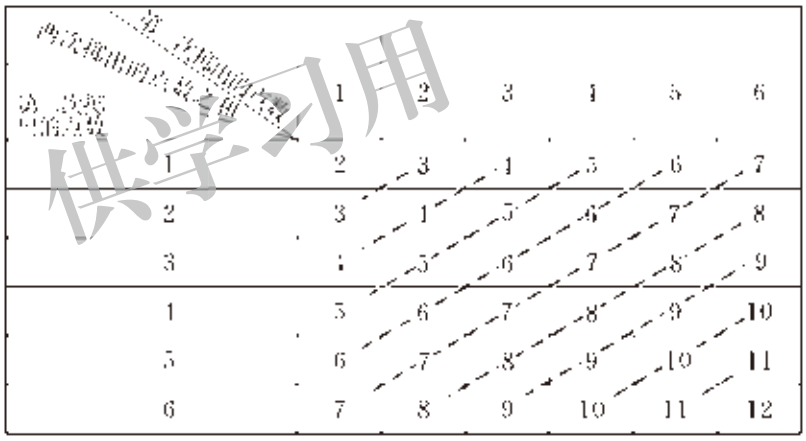
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

(我们用(i,j)表示事件“第一次掷得 i 点，第二次掷得 j 点”. 例如，(3,4) 表示第一次掷得 3 点，第二次掷得 4 点.)

显然这 36 种结果发生的概率是相同的，都是$\frac{1}{36}$.

由此，可以求出两次投掷所得结果的点数之和 X. 为了清楚地表示，可以列出表 2-3：

表 2-3



根据表 2-3，X 的可能取值为 2,3,…,12；其中使得 X 取值为 2 的可能结果只有 1 种(1,1)，因此 X 取值为 2 的概率为 $P(X=2)=\frac{1}{36}$. 使得 $X=3$ 的可能结果有 2 种：(1,2)或(2,1)，因此 $X=3$ 的概率为 $P(X=3)=\frac{2}{36}$. 同理可求得随机变量 X 取其他值的概率，最后可得 X 的分布列如下(见表 2-4)：

表 2-4

$X=x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(X=x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

例 4　用 X 表示投掷一枚均匀的骰子所得的点数，利用 X 的分布列求出下列事件发生的概率：

（1）掷出的点数是偶数；

（2）掷出的点数大于 3 而不大于 5；

（3）掷出的点数超过 1.

解　容易得到 X 的分布列为

$$P(X=i)=\frac{1}{6}\quad (i=1,2,\cdots,6).$$

根据上式，可得：

（1）掷出的点数是偶数是指$\{X=2\}$ 或$\{X=4\}$ 或$\{X=6\}$，因此掷出的点数是偶数的概率为

$$\begin{aligned}&P(\{X=2\}\cup\{X=4\}\cup\{X=6\})\\=&P(X=2)+P(X=4)+P(X=6)\\=&\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}.\end{aligned}$$

（2）掷出的点数大于 3 而不大于 5 是指掷得 4 点或 5 点，它发生的概率为

$$P(3<X\leqslant 5)=P(X=4)+P(X=5)=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}.$$

（3）掷出的点数超过 1 的对立事件是掷得 1 点，因此掷出的点数超过 1 的概率为

$$P(X>1)=1-P(X=1)=1-\frac{1}{6}=\frac{5}{6}.$$

从上面的问题可以看到，如果知道了随机变量 X 的分布列，X 在各个范围取值的概率通过计算就可以得到，因此，这个随机变量的规律就掌握了.

习　题　2—1

1. 下列随机变量的可能取值各是什么？

（1）某人做了 10 次试验，每次试验都可能成功或失败，10 次试验中成功的次数；

（2）一个箱子中装有 12 件产品，其中有 3 件次品，从这批产品中任意挑选出 4 件，取出的产品中次品的件数.

2. 连续掷两次均匀的硬币，用 X 表示正面朝上的次数，求：

（1）$P(X=1)$；　　（2）$P(X\leqslant 2)$；　　（3）$P(0\leqslant X<2)$.

3. 在一个箱子中装有编号分别为 1，2，3，4，5 的完全一样的 5 个球，现从中同时取出两个球，设 X 为取出的两球的最大编号，求 X 的分布列.

（注：可以使用计算器完成习题）

§2　超几何分布

已知在 10 件产品中有 4 件次品，现从这 10 件产品中任取 3 件，用 X 表示取得的次品数，试写出 X 的分布列.

首先，从这 10 件产品中任取 3 件，共有 C_{10}^{3} 种取法，每一种取法都是等可能的.

已知在 10 件产品中有 4 件次品，故 X 的可能取值为 0，1，2，3.

其中，“$X=0$”表示“任取的 3 件产品中不含次品”，这意味着，从 4 件次品中取出 0 件，再从 10－4 件正品中取出 3－0 件，由分步乘法计数原理可知，共有 $C_4^0C_{10-4}^{3-0}$ 种取法，故事件“$X=0$”的概率为

$$P(X=0)=\frac{C_4^0C_{10-4}^{3-0}}{C_{10}^{3}}=\frac{20}{120}\approx 0.166\ 7.$$

类似地，“$X=1$”表示“任取的 3 件产品中恰有 1 件次品”，这意味着，取出 1 件次品和 3－1 件正品，共有 $C_4^1C_{10-4}^{3-1}$ 种取法，故

$$P(X=1)=\frac{C_4^1C_{10-4}^{3-1}}{C_{10}^{3}}=\frac{60}{120}=0.5.$$

同理可得，　$P(X=2)=\frac{C_4^2C_{10-4}^{3-2}}{C_{10}^{3}}=\frac{36}{120}=0.3$，

$$P(X=3)=\frac{C_4^3C_{10-4}^{3-3}}{C_{10}^{3}}=\frac{4}{120}\approx 0.033\ 3.$$

事实上，“$X=k$”($k=0,1,2,3$)表示“取出的 3 件产品中恰有 k 件次品”，这意味着，从 4 件次品中取出 k 件，再从 10－4 件正品中取出 3－k 件，共有 $C_4^kC_{10-4}^{3-k}$ 种取法，故 X 的分布列为

$$P(X=k)=\frac{C_4^kC_{10-4}^{3-k}}{C_{10}^{3}}\quad(k=0,1,2,3).$$

近似计算后，也可以列成表 2-5：

表 2-5

$X=k$	0	1	2	3
$P(X=k)$	0.166 7	0.5	0.3	0.033 3

例 1　一个袋中装有 8 个红球和 4 个白球，这些球除颜色外完全相同. 现从袋中任意摸出 5 个球，用 X 表示摸出红球的个数.

(1)求 $P(X=3)$；

(2)试写出 X 的分布列.

解　(1)袋中共有 $8+4=12$ 个球，从中任意摸出 5 个球，共有 C_{12}^5 种不同的取法，每一种取法都是等可能的.

“$X=3$”表示“摸出的 5 个球中恰有 3 个红球”，即“摸出 3 个红球和 2 个白球”，共有 $C_8^3C_4^2$ 种取法，则

$$P(X=3)=\frac{C_8^3C_4^2}{C_{12}^5}=\frac{14}{33}\approx 0.424\ 2.$$

(2)X 的可能取值为 1,2,3,4,5.

“$X=k$”($k=1,2,3,4,5$)表示“摸出的 5 个球中恰有 k 个红球”，即“摸出 k 个红球和 $5-k$ 个白球”，共有 $C_8^kC_4^{5-k}$ 种取法，故 X 的分布列为

$$P(X=k)=\frac{C_8^kC_4^{5-k}}{C_{12}^5}\quad(k=1,2,3,4,5).$$

近似计算后，列成表 2-6：

表 2-6

$X=k$	1	2	3	4	5
$P(X=k)$	0.010 1	0.141 4	0.424 2	0.353 6	0.070 7

对照前面的实例分析，如果我们将袋中的 12 个球看作 12 件产品，8 个红球看作 8 件次品，4 个白球看作 12−8 件正品. 任取 5 个球中红球的个数 X 可以看作是任取 5 件产品中所含的次品数. 从而可以看出例 1 与前面“取次品问题”是同一类问题.

抽象概括

一般地，设有 N 件产品，其中有 $M(M\leqslant N)$ 件次品. 从中任取 $n(n\leqslant N)$ 件产品，用 X 表示取出的 n 件产品中次品的件数，那么

$$P(X=k)=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}\quad(\text{其中 } k \text{ 为非负整数}).$$

如果一个随机变量的分布列由上式确定，则称 X 服从参数为 N，M，n 的**超几何分布**.

超几何分布是随机现象中十分常见的一类分布.

思考交流

下列随机变量 X 是否服从超几何分布？如果服从，那么各分布的参数(即定义中的 N,M,n)分别是多少？

(1) 一个班共有 45 名同学，其中女生 20 人，现从中任选 7 人，用 X 表示其中女生的人数；

(2) 从一副扑克牌(去掉大、小王，共 52 张)中取出 a 张牌，用 X 表示取出的黑桃的张数.

尝试列举一些超几何分布的例子，并与同学进行交流.

例 2 高三(1)班的联欢会上设计了一项游戏：准备了 10 张相同的卡片，其中只在 5 张卡片上印有“奖”字. 游戏者从 10 张卡片中任意抽取 5 张，如果抽到 2 张或 2 张以上印有“奖”字的卡片，就可获得一件精美小礼品；如果抽到的 5 张卡片上都印有“奖”字，除精美小礼品外，还可获赠一套丛书. 一名同学准备试一试，那么他能获得精美小礼品的概率是多少？能获赠一套丛书的概率又是多少？

分析 可以将 10 张卡片看作是 10 件“产品”，5 张印有“奖”字的卡片看作 5 件“次品”，任意抽取的 5 张卡片中印有“奖”字的卡片数，可以看作是任取 5 件“产品”中所含的“次品”数.

解 设 X 表示抽取 5 张卡片中印有“奖”字的卡片数，则 X 服从参数为 $N=10,M=5,n=5$ 的超几何分布.

X 的可能取值为 0,1,2,3,4,5，则 X 的分布列为

$$P(X=k)=\frac{C_5^k C_5^{5-k}}{C_{10}^5}\quad (k=0,1,2,3,4,5).$$

若要获得精美小礼品，只需 $X\geqslant 2$，故获得精美小礼品的概率为

$$\begin{aligned}P(X\geqslant 2)&=1-P(X<2)\\&=1-P(X=0)-P(X=1)\\&=1-\frac{C_5^0 C_5^5}{C_{10}^5}-\frac{C_5^1 C_5^4}{C_{10}^5}\\&=\frac{113}{126}\approx 0.896\ 8.\end{aligned}$$

若要获赠一套丛书，必须 $X=5$，故获赠一套丛书的概率为

$$P(X=5)=\frac{C_5^5 C_5^0}{C_{10}^5}=\frac{1}{252}\approx 0.004\ 0.$$

由上面的计算可以看出，该同学参加游戏，获得精美小礼品的概率大约为 0.896 8，希望很大. 但获赠一套丛书的概率大约只有 0.004 0，希望不大.

阅读材料

彩票中的概率

有一种“36 选 7＋1”的彩票是从 01,02,…,36 这 36 个数中任选择 7 个,开奖时公布的是 7 个正选数与 1 个特选数.若你选择的 7 个数恰是公布的 7 个正选数,将获特等奖;若你选择的 7 个数中有 6 个是公布的正选数,另一个是公布的特选数,将获一等奖……下表是该彩票某一期的中奖公布:

中奖号码									
正选号码	11	24	25	26	27	33	34	特选号码	36
	中奖注数					单注奖金			
特等奖	1 注					500 万元			
一等奖	2 注					591 005 元			
二等奖	132 注					8 954 元			
三等奖	397 注					500 元			
四等奖	18 149 注					50 元			
五等奖	239 059 注					5 元			

特等奖和一等奖的奖金十分具有诱惑力.但我们知道,从 01,02,…,36 中任取 7 个数字,共有 C_{36}^{7} 种取法,每种取法都是等可能的,却只有一种取法和公布的正选数都一致,因此,获特等奖的概率为 $\frac{1}{C_{36}^{7}}\approx 0.000\ 000\ 12$,大约只有八百三十五万分之一,这一概率非常小.

而要想获得一等奖,就必须取到 6 个正选数和 1 个特选数,共有 $C_{7}^{6}C_{1}^{1}$ 种取法,这些取法都是等可能的.因此,获一等奖的概率为 $\frac{C_{7}^{6}C_{1}^{1}}{C_{36}^{7}}\approx 0.000\ 000\ 84$,也不足百万分之一.

事实上,要想获得五等奖(奖金 5 元),需要取到 3 个正选数和 1 个特选数,或者取到 4 个正选数,共有 $C_{7}^{3}C_{1}^{1}C_{28}^{3}+C_{7}^{4}C_{28}^{3}$ 种取法,概率也只有 $\frac{C_{7}^{3}C_{1}^{1}C_{28}^{3}+C_{7}^{4}C_{28}^{3}}{C_{36}^{7}}\approx 0.027\ 47$,还不足十分之一.因此,即便是获得 5 元奖金,也是小概率事件.

练　习

已知某社区的 10 位选民代表中有 5 位支持候选人 A,现随机采访他们中间的 4 位,求其中至少有 2 名支持候选人 A 的概率.

习 题 2—2

1. 某班有 30 名男生和 10 名女生，现从中随机选出 5 名学生，计算所选学生中女生数的分布列.
2. 一批 100 个计算机芯片中含 2 个不合格的芯片，现随机地从中取出 5 个芯片作为样本.
   (1) 计算样本中含不合格芯片数的分布列；
   (2) 求样本中至少含有一个不合格芯片的概率.
3. 一般地，将扑克牌中的 J，Q，K 叫花牌. 某人从一副已洗均匀的扑克牌(去掉大、小王，共 52 张)中依次摸 5 张，所摸扑克牌中恰好有 3 张花牌的概率是多少？若 X 表示摸 5 张扑克牌中的花牌数，求 X 的分布列.
4. 一个 10 人的办公室里有 5 名男性和 5 名女性. 现在需要形成一个由 4 人组成的委员会，研究办公环境中是否允许吸烟的问题. 管理方声明人员是随机选择的. 但是最终选择的结果为“4 人都是男性”.
   (1) 选择 4 人都是男性的概率是多少？
   (2) 管理方的声明可信吗？
   (注：可以使用计算器完成习题)

§3　条件概率与独立事件

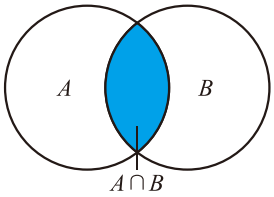
问题提出

100 件产品中有 93 件产品的长度合格，90 件产品的质量合格，85 件产品的长度、质量都合格. 现在，任取一件产品，若已知它的质量合格，那么它的长度合格的概率是多少?

分析理解

如果我们令 $A=\{$产品的长度合格$\}$，$B=\{$产品的质量合格$\}$，那么 $A\cap B=\{$产品的长度、质量都合格$\}$.

现在，任取一件产品，已知它的质量合格(即 B 发生)，则它的长度合格(即 A 发生)的概率为$\frac{85}{90}$.



那么，此概率$\left(\frac{85}{90}\right)$与事件 A 及 B 发生的概率有什么关系呢?

由题目可知：

$$P(A)=\frac{93}{100},\ P(B)=\frac{90}{100},\ P(A\cap B)=\frac{85}{100},$$

因此，在事件 B 发生的前提下，事件 A 发生的概率为

$$\frac{85}{90}=\frac{\frac{85}{100}}{\frac{90}{100}}=\frac{P(A\cap B)}{P(B)}.$$

抽象概括

实际上，上面的例子是求已知 B 发生的条件下(即质量合格)，A 发生(即长度合格)的概率，称为 ***B* 发生时 *A* 发生的条件概率**，记为 $P(A|B)$. 从上面的例子中可以看出，当 $P(B)>0$ 时，我们有

$$P(A|B)=\frac{P(A\cap B)}{P(B)}\quad(\text{其中，}A\cap B\text{ 也可以记成 }AB).$$

类似地，当 $P(A)>0$ 时，***A* 发生时 *B* 发生的条件概率**为

$$P(B|A)=\frac{P(AB)}{P(A)}.$$

问题提出

从一副扑克牌(去掉大、小王,共 52 张)中随机取出 1 张,用 A 表示取出的牌是“Q”,用 B 表示取出的牌是红桃. 试利用 $P(B)$ 及 $P(AB)$ 计算 $P(A|B)$.

分析理解

由于 52 张牌中有 13 张红桃,则 B 发生(即取出的牌是红桃)的概率为

$$P(B)=\frac{13}{52}=\frac{1}{4}.$$

而在 52 张牌中既是红桃又是“Q”的牌只有 1 张,故

$$P(AB)=\frac{1}{52}.$$

根据条件概率的计算公式,有

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{\frac{1}{52}}{\frac{1}{4}}=\frac{1}{13}.$$

另外,由于 52 张牌中共有 4 张“Q”,因而 $P(A)=\frac{4}{52}=\frac{1}{13}$.

不难发现

$$P(A|B)=P(A).$$

即,已知取出的牌是红桃时它为“Q”的概率等于取出的牌是“Q”的概率.

容易理解,取出的牌是红桃不影响取出的牌是“Q”的概率.

抽象概括

对两个事件 A, B, 如果 $P(A|B)=P(A)$,则意味着事件 B 发生不影响事件 A 的概率. 设 $P(B)>0$, 根据条件概率的计算公式, $P(A)=P(A|B)=\frac{P(AB)}{P(B)}$, 我们能得到 $P(AB)=P(A)P(B)$.

一般地, 对两个事件 A, B, 如果 $P(AB)=P(A)P(B)$, 则称 **A, B 相互独立**. 可以证明,如果 A, B 相互独立,则 A 与 $\overline{B}$,$\overline{A}$ 与 B,$\overline{A}$ 与 $\overline{B}$ 也相互独立.

思考交流

在实际应用中，我们常常根据实际问题的条件，利用直觉来判断事件间的“相互独立性”. 例如，掷硬币时，{第一次掷出正面}与{第二次掷出正面}是相互独立的；{甲厂的产品是次品}与{乙厂的产品是次品}是相互独立的.

请列举出日常生活中的一些独立事件的例子.

例　通过调查发现，某班学生患近视的概率为 0.4. 现随机抽取该班的 2 名同学进行体检，求他们都近视的概率.

解　如果用 $A_i(i=1,2)$ 表示抽取的第 i 名学生患近视，则 $P(A_1)=P(A_2)=0.4$. 可以认为 2 名同学是否近视是相互独立的，因此

$$P(\text{两名同学都近视})=P(A_1A_2)=P(A_1)P(A_2)$$
$$=0.4\times0.4=0.16.$$

前面我们讨论了两个事件的相互独立性. 事实上，对多个事件我们也可以讨论它们的相互独立性. 如果 $A_1,A_2,\cdots,A_n$ 相互独立，则有

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n).$$

思考交流

有人以为，把一枚均匀硬币掷 4 次，事件“第 1 次出现正面，第 2 次出现反面，第 3 次出现正面，第 4 次出现反面”的发生是正常的. 而事件“4 次都出现正面”的发生就不太正常，好像前者发生的概率大. 你同意这种观点吗？

练　习

4 个射手独立地进行射击，设每人中靶的概率都是 0.9. 试求下列各事件的概率：

(1) 4 人都中靶；

(2) 4 人都没中靶；

(3) 两人中靶，另两人没中靶.

阅读材料

概率与法庭

下面是美国历史上一个很有名的案例，控辩双方都使用了概率论的知识.

1964 年 6 月 18 日上午，年老的 Brooks 太太在购物后沿着洛杉矶市郊区的一条小巷回家. 一名袭击者从背后将她推倒，抢了她的钱包. Brooks 太太瞥到逃跑的袭击者是一名穿着黑色的衣服、梳着马尾辫的金发年轻女子.

正在小巷一端自家屋前浇灌草坪的 Bass 被叫喊声所吸引，她看到一名女子跑出小巷，进入一辆部分是黄色的小汽车，车由一名留着络腮胡子的黑人驾驶逃离了现场.

按照这些描述，警察很快便逮捕了一对年轻夫妇：Malcolm Collins 和他的妻子 Janet Collins. Malcolm 是黑人，虽然脸刮得很干净，但有证据表明他最近留过络腮胡子. Janet 是金发，并且常常将她的头发梳成马尾辫. 他们还驾驶一辆部分是黄色的林肯车. 原告请了一位数学教授作为专家证人，根据目击者的描述，这位教授提供了一份犯罪嫌疑人的特征，而且保守地估计了各个特征的概率（如下表所示）.

特征	概率
金发女子	$\frac{1}{3}$
梳马尾辫的女子	$\frac{1}{10}$
留小胡子的男子	$\frac{1}{4}$
留络腮胡子的男子	$\frac{1}{10}$
部分是黄色的小汽车	$\frac{1}{10}$
种族混合同车	$\frac{1}{1\,000}$

原告据此提出随机地挑出具有这些特征的一对夫妇的概率是上述概率的乘积，也就是$\frac{1}{12\,000\,000}$，这是一个很小的概率，不足千万分之一，因此，被告是有罪的. 陪审团同意了这一说法并且宣布对 Collins 夫妇犯有二级抢劫罪的裁决.

在本案随后的上诉中，辩护律师辩护如下：首先，承认以下事实，假设在洛杉矶地区有 5 000 000 对夫妇，随机地挑选出来适合目击者描述的一对夫妇的概率是$\frac{1}{12\,000\,000}$，这个概率是非常小的. 但是，如果该地区有一对这样的夫妇就可能有许多对这样的夫妇. 计算表明，“在有一对这样夫妇的条件下，有多对这样的夫妇”这个随机事件的概率不低. 因此，无法断言 Collins 夫妇是罪犯.

加利福尼亚州最高法院最终推翻了本案最初的有罪裁决.

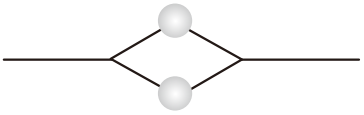
由此可以看出，概率论知识对于法庭裁决起着重要的参考作用.

习 题 2—3

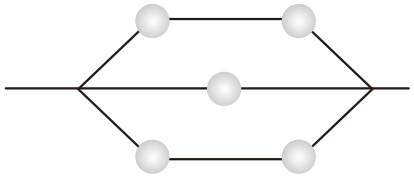
1. 在一个坛子中装有 16 个除颜色外完全相同的玻璃球，其中有 2 个红的，3 个蓝的，5 个绿的，6 个黄的. 从中任取一球，放回后，再取一球，求第一次取出红球且第二次取出黄球的概率.
2. 妇女绝育手术每 4 000 例中会有 1 例失败，男子绝育手术的成功率为99.9%. 在双方都做了绝育手术的夫妇中，随机抽取一对夫妇，求他们仍怀上了孩子的概率.
3. n 台机床独立工作，每台机床正常工作的概率都是 0.99，求 n 台机床都正常工作的概率. 借助计算器对 $n=1\,000$ 给出结果，用此说明一个大纺织厂为什么要有一支常备的机修队伍.
4. 下列图中表示由若干个某种电子元件组成的电路. 已知每个元件的可靠性是 0.9，而且各个元件的可靠性是彼此独立的. 分别求出下列电路畅通的概率：

(1)

(2)



(3)



（第 4 题）

（注：可以使用计算器完成习题）

§4 二项分布

我们知道，随机变量的分布可以描述随机现象的规律. 前面我们已经研究了一类十分常见而重要的分布——超几何分布，本节我们研究另一类常见而重要的分布——二项分布.

某射击运动员进行了 4 次射击，假设每次射击击中目标的概率都为$\frac{3}{4}$，且各次击中目标与否是相互独立的. 用 X 表示这 4 次射击中击中目标的次数，求 X 的分布列.

每次射击都有两种可能的结果：击中目标或没击中目标，并且每次击中目标的概率都是 $p=\frac{3}{4}$，没击中目标的概率均为 $1-p=\frac{1}{4}$. 在对目标进行的 4 次射击中，击中目标次数 X 的取值为 0，1，2，3，4.

如果 $X=0$，即 4 次都没击中目标，这只包含 $C_4^0=1$ 种情况. 由于每次射击都是独立的，从而

$$P(X=0)=C_4^0\left(\frac{1}{4}\right)^4.$$

同理可得

$$P(X=4)=C_4^4\left(\frac{3}{4}\right)^4.$$

如果 $X=2$，即 4 次中有 2 次击中目标，这包含 C_4^2 种情况，每种情况发生的可能性都为$\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2$，因此

$$P(X=2)=C_4^2\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2.$$

类似地，可以得到

$$P(X=1)=C_4^1\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^3,$$

$$P(X=3)=C_4^3\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1.$$

这样，X 的分布列可以写成如表 2-7 所示的形式：

表 2-7

$X=k$	0	1	2
$P(X=k)$	$C_4^0\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^4$	$C_4^1\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^3$	$C_4^2\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2$
$X=k$	3	4	
$P(X=k)$	$C_4^3\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1$	$C_4^4\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^0$	

在上面的问题中，如果将一次射击看成做了一次试验，思考如下问题：

1. 一共进行了多少次试验？每次试验有几个可能的结果？

2. 如果将每次试验的两个可能的结果分别称为“成功”（击中目标）和“失败”（没击中目标），那么，每次试验成功的概率是多少？它们相同吗？

3. 各次试验是否相互独立？独立性在随机变量 X 的分布列的计算中，具体应用在哪里？

进行 n 次试验，如果满足以下条件：

（1）每次试验只有两个相互对立的结果，可以分别称为“成功”和“失败”；

（2）每次试验“成功”的概率均为 p，“失败”的概率均为 $1-p$；

（3）各次试验是相互独立的.

用 X 表示这 n 次试验中成功的次数，则

$$P(X=k)=C_n^k p^k(1-p)^{n-k}\quad(k=0,1,2,\cdots,n).$$

若一个随机变量 X 的分布列如上所述，称 X 服从参数为 n，p 的**二项分布**，简记为 $X\sim B(n,p)$.

下列随机变量 X 服从二项分布吗？如果服从二项分布，其参数各是什么？

（1）掷 n 枚相同的骰子，X 为出现“1”点的骰子数；

(2) n 个新生婴儿，X 为男婴的个数；

(3) 某产品的次品率为 p，X 为 n 个产品中的次品数；

(4) 女性患色盲的概率为 0.25%，X 为任取 n 个女人中患色盲的人数.

二项分布有着十分广泛的应用. 尝试列举一些二项分布的例子，并与其他同学进行交流.

例 1 某公司安装了 3 台报警器，它们彼此独立工作，且发生险情时每台报警器报警的概率均为 0.9. 求发生险情时，下列事件的概率：

(1) 3 台都没报警；　　(2) 恰有 1 台报警；

(3) 恰有 2 台报警；　　(4) 3 台都报警；

(5) 至少有 2 台报警；　　(6) 至少有 1 台报警.

分析 在发生险情时，我们将每台报警器是否报警看成做了一次试验，那么一共做了 3 次试验，并且它们彼此是独立的；在每次试验中，把“报警”看作成功，“未报警”看作失败，那么每次试验成功的概率都是 0.9. 如果令 X 为在发生险情时 3 台报警器中报警的台数，那么 X 服从参数为 $n=3$，$p=0.9$ 的二项分布.

解 令 X 为在发生险情时 3 台报警器中报警的台数，那么 X 服从参数为 $n=3$，$p=0.9$ 的二项分布，则它的分布列为

$$P(X=k)=C_3^k 0.9^k(1-0.9)^{3-k} \quad (k=0,1,2,3),$$

即

$X=k$	0	1	2	3
$P(X=k)$	0.001	0.027	0.243	0.729

由此可知 3 台都未报警、恰有 1 台报警、恰有 2 台报警和 3 台都报警的概率分别为 0.001，0.027，0.243，0.729.

另外，至少有 2 台报警的概率为

$$\begin{aligned} P(X\geqslant 2)&=P(X=2)+P(X=3)\\ &=0.243+0.729=0.972; \end{aligned}$$

至少有 1 台报警的概率为

$$\begin{aligned} P(X\geqslant 1)&=1-P(X=0)\\ &=1-0.001=0.999. \end{aligned}$$

由上面的计算可见，安装多台报警器能使发生险情时得到预警的概率比只用 1 台报警器的概率大.

练　习

1. 已知一批产品的次品率为 $p=0.12$，从中任取 5 件，求取得各次品数的概率.
2. 已知一种疾病的发病率为 0.002，并且每人是否患此病是彼此独立的. 若某单位共有 800 人，求该单位至少有 2 人患此病的概率.

问题提出

有人这样认为，投掷一枚均匀的硬币 10 次，恰好 5 次正面朝上的概率很大. 你同意他的想法吗？

在必修课程中，我们曾经做过上面的试验，通过试验数据我们发现 10 次投掷中恰有一半正面朝上的可能性不到 25%，并不是通常想象得那么大. 那么，“投掷一枚均匀的硬币 10 次，恰好有 5 次正面朝上”的概率究竟是多少呢？试着计算一下.

思考交流

1. 如果用 X 表示 10 次投掷中正面朝上的次数.

（1）X 服从二项分布吗？与其他同学交流你的理由.

（2）这个分布的参数是多少？

（3）求出 X 的分布列，并计算出“投掷一枚均匀的硬币 10 次，恰有 5 次正面朝上”的概率.

2. 有的同学可能会继续思考，10 次投掷中恰有一半正面朝上的可能性不大，那么增加投掷次数，比如 100 次，恰好出现一半“正面朝上”（即 50 次“正面朝上”）的可能性会不会大一些呢？

动手实践

1. 每人做一次试验（即投掷 100 次均匀的硬币，或利用科学计算器产生随机数进行模拟），记录正面朝上的次数. 计算恰好得到了 50 次正面朝上的同学的人数占全班同学总人数的比例.

2. 每人再做一次试验，计算恰好得到了 50 次正面朝上的同学的人数占全班同学总人数的比例.

信息技术建议

我们在必修课程中学过了怎样用 Excel 来进行模拟试验，但更简捷的办法是编程模拟. 可供选择的编程语言很多，如 Basic，Fortran，C 语言等. 我们可以用 C 语言编写一个模拟“投掷一枚均匀的硬币 100 次”的程序，并统计这 100 次试验中得到正面朝上的次数（具体程序见附录）. 利用图形计算器也可很容易进行模拟，具体操作步骤参见本节的“信息技术应用”栏目.

分析理解

有人给出了一个掷均匀硬币的模拟试验（参见费勒著《概率论及其应用》），试验相当于 100 个人，每人都掷 100 次均匀硬币，记录下各自掷出正面的次数如下：

54	46	53	55	46	54	41	48	51	53
48	46	40	53	49	49	48	54	53	45
43	52	58	51	51	50	52	50	53	49
58	60	54	55	50	48	47	57	52	55
48	51	51	49	44	52	50	46	53	41
49	50	45	52	52	48	47	47	47	51
45	47	41	51	49	59	50	55	53	50
53	52	46	52	44	51	48	51	46	54
45	47	46	52	47	48	59	57	45	48
47	41	51	48	59	51	52	55	39	41

在这 100 个数字中，50 出现了 7 次，因此，“掷 100 次硬币，恰好出现 50 次正面”的频率是 $\frac{7}{100}=0.07$.

我们也可以从理论上计算一下这个概率. 用 Y 表示 100 次投掷中正面朝上的次数，则易知 Y 服从参数为 $n=100$，$p=\frac{1}{2}$ 的二项分布. 因此，100 次投掷中恰有一半正面朝上的概率为

$$P(Y=50)=C_{100}^{50}\left(\frac{1}{2}\right)^{50}\left(\frac{1}{2}\right)^{100-50}=C_{100}^{50}\left(\frac{1}{2}\right)^{100}\approx 0.08.$$

由此可见，前面算出的频率与理论上的概率值是很接近的.

人们也许存在着这样的误解，认为“掷 100 次出现 50 次正面”是必然的，或者说它的概率应该很大，但通过做试验和计算表明这个概率只有 8%左右. 由此可见，学习概率的知识有时能够帮助我们澄清一些误解.

例 2 某车间有 5 台机床，每台机床正常工作与否彼此独立，且正常工作的概率均为 0.2. 设每台机床工作时需电力 10 kW，但因电力系统发生故障现只能提供 30 kW 的电力，问此时车间不能正常工作的概率有多大.

分析 我们将每台机床是否工作看成一次试验，那么一共有 5 次试验，并且它们彼此是独立的；在每次试验中，把正常工作看作“成功”，不能正常工作看作“失败”，那么每次试验“成功”的概率都是 0.2. 如果令 X 为 5 台机床中正常工作的台数，那么 X 服从参数为

$n=5$，$p=0.2$的二项分布.

而题目中“车间不能正常工作”是指需用电力超过 30 kW，即$X\geqslant 4$.

解　设 X 为 5 台机床中工作的台数，则 X 服从参数为 $n=5$，$p=0.2$的二项分布，即

$$P(X=k)=C_5^k 0.2^k(1-0.2)^{5-k}\quad (k=0,1,\cdots,5).$$

那么

$$\begin{aligned}P(X\geqslant 4)&=P(X=4)+P(X=5)\\&=C_5^4\times 0.2^4\times 0.8+C_5^5\times 0.2^5\times 0.8^0\\&\approx 0.007.\end{aligned}$$

这是一个概率很小的事件，几乎不会发生. 因此，如果车间不能正常工作时不会造成破坏性后果，那么在只能提供 30 kW 的电力的情况下仍可以安排生产.

信息技术应用

“投硬币 100 次，50 次正面朝上”的概率有多大

抛掷一枚均匀硬币 100 次，恰好有 50 次正面朝上的概率为 $C_{100}^{50}\left(\frac{1}{2}\right)^{50}\left(\frac{1}{2}\right)^{100-50}\approx 0.079\ 6$. 下面用图形计算器的数据处理功能进行模拟试验，每个模拟试验都是将硬币投 100 次，一共做 150 个投币模拟试验，统计“50 次正面朝上”的个数，计算出它的频率.

1. 打开图形计算器进入数据编辑窗口，在 c1 列输入 seq(x,x,1,100)，确认后产生 100 次试验的序号(如图 2-1 所示)；

2. 在 c2 列输入 seq(int(rand() * 2),x,1,100)，确认后产生 100 次试验的结果(如图 2-2 所示)，其中 1 表示正面朝上，0 表示反面朝上；

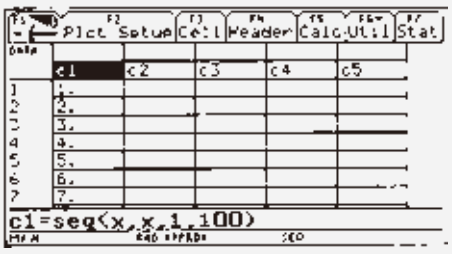


图 2-1

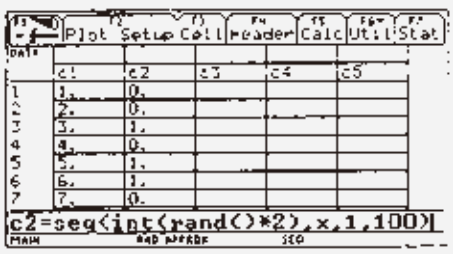


图 2-2

3. 在 c3 的第一行中输入 sum(c2)，得到 100 次试验中正面朝上的次数(如图 2-3 所示)；

4. 再到 c3 的第二行中输入 sum(c2)，确认后 c2 中将重新产生 100 次试验的结果，并且在 c3 第二行得到第二个 100 次试验中正面朝上的次数(如图 2-4 所示)；

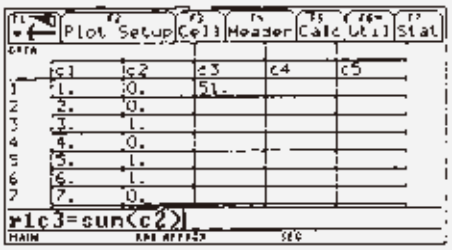


图 2-3

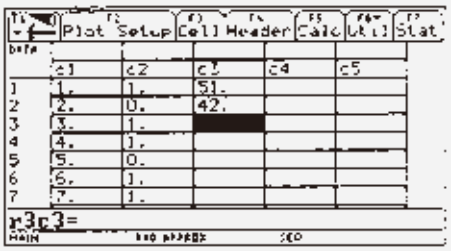


图 2-4

5. 如此继续下去直到在 c3 的第 150 行产生第 150 个 100 次试验正面朝上的次数(如图 2-5 所示)；

6. 作出 c3 列的直方图，并且跟踪图形得到 50 次正面朝上的次数为 11 次(如图 2-6 所示)，从而说明“硬币抛掷 100 次恰好有 50 次正面朝上”的频率是 11/150≈0.073 3，比较接近理论值 0.079 6，如果我们进行更多的试验，我们会发现，随着试验次数的增多，频率会越来越接近 0.079 6.

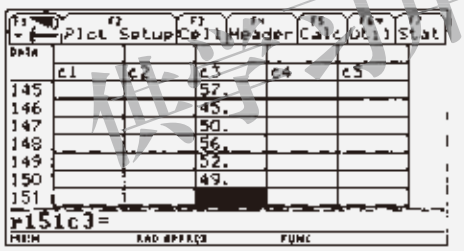


图 2-5

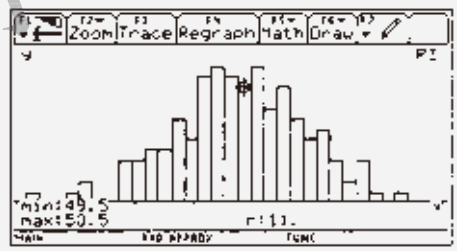


图 2-6

事实上，抛掷硬币 100 次，正面朝上的次数有 101 个可能的结果，正面朝上的次数为 n 的概率为 $P_{100}(n)=C_{100}^{n}0.5^{100}$ ($n=0$, 1, 2, …,100)，在图形计算器中输入 $P_{100}(n)$ 的表达式，作出 $P_{100}(n)$ 的图像，同时显示分布列(如图 2-7～图 2-10 所示). 可以看到当 $n=0$ 时概率约为 7.889×10^{-31}，这说明抛掷硬币 100 次，恰好 100 次反面朝上的概率很小，几乎不发生；当 $n=34$ 时概率约为 0.000 46，说明抛掷硬币 100 次，恰好 34 次正面朝上的概率为 0.000 46，在 2 700个抛掷硬币 100 次的试验中大概会出现 1 次；当 $n=50$ 时概率最大，约为 0.08，这说明抛掷硬币 100 次，恰好 50 次正面朝上的概率约为 0.08，在 150 个抛掷硬币 100 次的试验中大概会出现 12 次，尽管其概率最大，但也没有超过 8%.

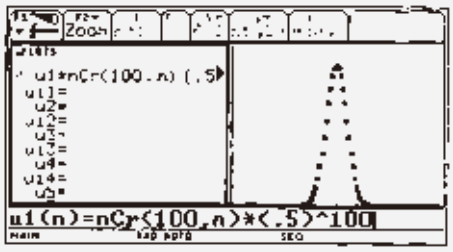


图 2-7

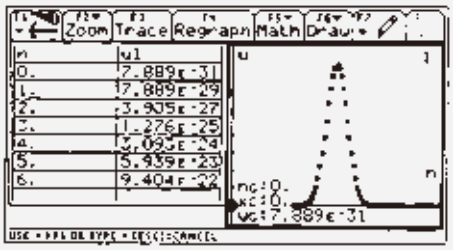


图 2-8

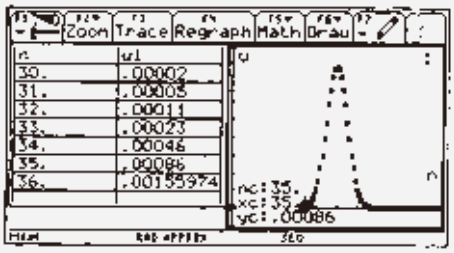


图 2-9

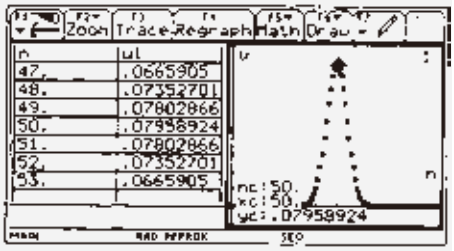


图 2-10

阅读材料

需要多少条外线

利用二项分布的知识能帮助人们解决很多实际问题. 看一个具体的例子：某电话总机有 1 000 个分机，要想用户每次要外线时都不占线需设置 1 000 条外线，但这样的成本较高. 如果设置太少的外线，又会使用户因经常要不通外线而抱怨. 如何处理这个矛盾呢？一个可以接受的想法是：如果用户能以较大的概率，例如 99% 的概率要通外线，那么一方面可以降低成本，另一方面用户也能满意. 如果每个分机平均每小时有 3 分钟需要外线，即需要外线的概率为 $\frac{3}{60}=\frac{1}{20}$，并且每个分机要外线与否是相互独立的，那么此时需要多少条外线呢？

有人会认为需要 1 000×99%＝990 条外线，是这样吗？我们不妨用学过的知识计算一下. 设 X 是这 1 000 个分机用户同时需要外线的个数，则 $X\sim B\left(1\,000,\frac{1}{20}\right)$. 假设 N 条外线能使用户以 99% 的概率要通外线，则 N 应满足 $P(X\leqslant N)\geqslant 99\%$. 通过计算可知只需 $N\geqslant 68$. 这表明，我们只需约 70 条外线即能使用户以 99% 的概率要通外线.

上面的例子表明，了解随机现象发生的规律是非常有意义的. 为了 99% 的能要通外线，我们无须保留 1 000 条外线中的 99%（即 1 000×0.99＝990 条），50%（即 500 条）也用不到，原来线路的 10%（即 100 条）就已足够了. 只要我们在允许的范围内稍作一些“牺牲”（1% 的用户可能要不通外线），就能大大降低成本.

习 题 2—4

A 组

1. 已知平均每 1 000 人中只有一人具有某种特定的稀有血型. 某小城共有 10 000 人，求没有人具有这种血型的概率.
2. 某篮球运动员投篮的命中率为 0.7，现投了 8 次球，求下列事件的概率：
   (1) 恰有 4 次投中；
   (2) 至少有 4 次投中；
   (3) 至多有 4 次投中.
3. n 支步枪独立射击目标，每支步枪的命中率都是 0.001，求至少有一支步枪命中目标的概率，并试讨论 n 充分大时的结果.
4. 5 名工人独立地工作，假定每名工人在 1 h 内平均有 12 min 需要电力(即任一时刻需要电力的概率为$\frac{12}{60}$).
   (1) 求在同一时刻恰有 3 名工人需要电力的概率；
   (2) 如果在同一时刻最多只能供给 3 名工人需要的电力，求超过负荷的概率.
5. 某小吃店根据以往的统计数据，估计出 90%的顾客喜欢吃炸鸡，余下 10%的顾客喜欢吃炸鱼. 现有 50 名顾客准备明天来吃午饭. 小吃店经理事先准备了 5 份炸鱼. 试分别求下列事件的概率：
   (1) 这 50 名顾客中恰有 5 人要炸鱼；
   (2) 这 50 名顾客中少于 5 人要炸鱼；
   (3) 这 50 名顾客中多于 5 人要炸鱼.

B 组

一批机床，每台发生故障的概率均为 0.01，且发生故障后由一个维修工就可排除. 现甲厂订购了 20 台机床，配备了 1 名维修工，乙厂订购了 80 台机床，配备了 3 名维修工. 试问：甲、乙两厂因机床故障又不能及时维修的概率各是多少？比较这两个概率，有什么实际意义？
(注：可以使用计算器完成习题)

§5　离散型随机变量的均值与方差

问题提出

在 10 件某种产品中，有 4 件次品. 从这 10 件产品中任取 3 件，用 X 表示取得产品中的次品数. 在前面第 2 节中我们已经求得 X 的分布列为

$X=k$	0	1	2	3
$P(X=k)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

现在我们关心，取 3 件该产品时，平均会取到几件次品？

那么，怎样的一个数能够“代表”这个随机变量取值的平均水平呢？

分析理解

我们先来看看一个常见的求平均值的问题. 设有 12 个西瓜，其中有 4 个重 5 kg，3 个重 6 kg，5 个重 7 kg，求西瓜的平均质量.

由平均数的意义，西瓜的平均质量应为 12 个瓜的总质量除以西瓜的总个数，即

$$\frac{5\times4+6\times3+7\times5}{12}=\frac{73}{12}\ (\text{kg}).$$

上式也可写成如下形式：

$$5\times\frac{4}{12}+6\times\frac{3}{12}+7\times\frac{5}{12}=\frac{73}{12}\ (\text{kg}).$$

其中$\frac{4}{12}$，$\frac{3}{12}$，$\frac{5}{12}$分别为质量是 5 kg，6 kg 和 7 kg 的西瓜数在总西瓜数中所占的比例. 上式告诉我们，如果知道各个质量所占的比例，则平均质量等于各个质量乘相应的比例再求和.

那么，在前面“取次品问题”中，根据 X 的分布列，有

$$0\times\frac{1}{6}+1\times\frac{1}{2}+2\times\frac{3}{10}+3\times\frac{1}{30}=1.2.$$

它表示，在一次的抽取中，3 件产品中平均有 1.2 件是次品，而 $\frac{1.2}{3}=\frac{4}{10}$，相当于 10 件产品中有 4 件次品. 这样，平均数 1.2 就代表“取次品问题”中随机变量 X 的平均取值.

抽象概括

现在，设随机变量 X 的可能取值为 $a_1, a_2, \cdots, a_r$，取 a_i 的概率为 $p_i (i=1, 2, \cdots, r)$，即 X 的分布列为

$$P(X=a_i)=p_i \quad (i=1, 2, \cdots, r).$$

定义 X 的**均值**为

$$a_1P(X=a_1)+a_2P(X=a_2)+\cdots+a_rP(X=a_r)$$
$$=a_1p_1+a_2p_2+\cdots+a_rp_r,$$

即随机变量 X 的取值 a_i 乘上取值为 a_i 的概率 $P(X=a_i)$ 再求和.

X 的均值也称作 X 的**数学期望**（简称期望），它是一个数，记为 EX，即

$$EX=a_1p_1+a_2p_2+\cdots+a_rp_r.$$

均值 EX 刻画的是 X 取值的“中心位置”，这是随机变量 X 的一个重要特征.

根据均值定义，我们知道，随机变量的分布完全确定了它的均值. 但反过来，两个不同的分布可以有相同的均值. 这表明分布描述了随机现象的规律，从而也决定了随机变量的均值. 而均值只是刻画了随机变量取值的“中心位置”这一重要特征，并不能完全决定随机变量的性质. 不过，有许多情形随机变量的分布并不好求，人们只能借助均值等刻画随机变量的数字特征来描述它. 也有一些问题，人们只需要了解随机变量的均值而无须全面了解其分布. 因此，均值自有它本身的重要意义.

例 1 令 X 为掷一枚均匀骰子出现的点数，求 EX.

解 X 的分布列是

$$P(X=i)=\frac{1}{6} \quad (i=1, 2, \cdots, 6),$$

根据均值的定义，可知

$$EX=1\times\frac{1}{6}+2\times\frac{1}{6}+\cdots+6\times\frac{1}{6}=3.5.$$

思考交流

投掷一枚均匀的骰子，只可能出现 1 点，2 点，……，6 点，怎样解释这个均值 3.5 呢？

例 2　设 X 只取 0，1 两个值，并且

$$P(X=0)=1-p, P(X=1)=p,$$

求 EX.

解
$$\begin{aligned} EX &= 0 \cdot P(X=0)+1 \cdot P(X=1) \\ &= 0 \cdot (1-p)+1 \cdot p=p. \end{aligned}$$

对一些常见的分布，我们可以求出它们的均值.

当随机变量服从参数为 n，p 的二项分布时，其均值为 np. 这一结论在直观上是明显的. 例如独立射击时，每次命中率为 $p=0.9$，则 $n=10$ 次射击命中次数的均值为 $np=10\times 0.9=9$ 次；掷 $n=10$ 次均匀硬币，正面出现的次数的均值为 $np=10\times\frac{1}{2}=5$ 次.

当随机变量 X 服从参数为 N, M, n 的超几何分布时，它的均值 $EX=n\frac{M}{N}$. 这一结论在直观上也是明显的. 设 $N=100$ 个产品中有 $M=10$ 个次品，任取 $n=20$ 个，取到的次品的均值是 $n\frac{M}{N}=20\times\frac{10}{100}=2$ 个.

练　习

一游戏者从标有 2～10 的 9 张卡片中随机取出一张，如果取出的卡片是奇数，则他赢得 1 元；如果取出的卡片是偶数，则他输掉 1 元. 求他每次的平均收益.

问题提出

据气象预报，某地区下月有小洪水的概率为 0.25，有大洪水的概率为 0.01. 该地区某工地上有一台大型设备，为保护设备有以下 3 种方案.

方案 1：运走设备，此时需花费 3 800 元；

方案 2：建一保护围墙，需花费 2 000 元，但围墙无法防止大洪水，当大洪水来临时，设备会受损，损失费为 60 000 元；

方案 3：不采取措施，希望不发生洪水，此时大洪水来临将损失 60 000 元，小洪水来临将损失 10 000 元.

你会采取哪一种方案呢？

分析理解

如果下月没有洪水，那么显然方案 3 最好；但如果有小洪水，方案 3 将损失 10 000 元，此时方案 2 较佳；但如果大洪水来临，则只有方案 1 才能免受巨大损失. 如此看来，各个方案都有被选择的理由. 一个合理的选择标准是比较各个方案的平均损失.

在方案 3 中，用 X 表示损失，则易知 X 的分布列如下：

X	60 000	10 000	0
P	0.01	0.25	0.74

因此，方案 3 的平均损失为：

$EX=60\ 000\times0.01+10\ 000\times0.25+0\times0.74=3\ 100$(元).

方案 2 的平均损失为围墙建设费 2 000 元和设备受损的平均费用：$60\ 000\times0.01+0\times0.99=600$(元)，即 $2\ 000+600=2\ 600$(元).

方案 1 的损失只是运走设备的费用 3 800 元.

由此可见，平均而言方案 2 的损失最小，可供选择.

思考交流

如果采取方案 2，或者损失 60 000 元，或者损失 2 000 元，怎样解释平均损失 2 600 元呢？如果采取方案 3，有可能一分钱不花，而方案 2 至少需要花 2 000 元，如何理解选择方案 2 平均损失最小呢？

问题提出

均值能够反映随机变量取值的“平均水平”，但有时两个随机变量即使均值相同，其取值差异也可能很大. 我们还需要另一个数来反映随机变量取值的集中程度.

例如，有 A,B 两种不同品牌的手表，它们的“日走时误差”分别为 X，Y（单位：s），X,Y 的分布列如下：

$$X\sim\begin{pmatrix}-0.01 & 0.00 & 0.01\\ 1/3 & 1/3 & 1/3\end{pmatrix},$$

$$Y\sim\begin{pmatrix}-0.50 & 0.00 & 0.50\\ 1/3 & 1/3 & 1/3\end{pmatrix}.$$

(1) 分别计算 X,Y 的均值，并进行比较；

(2) 这两个随机变量的分布有什么不同，如何刻画这种不同？

分析理解

根据 X,Y 的分布列计算可以得到 $EX=EY=0$，也就是说这两种表的平均日走时误差都是 0. 因此，仅仅根据平均误差，不能判断出哪一种品牌的表更好. 但进一步观察，我们可以发现 A 品牌的表的误差只有 ± 0.01 s，而 B 品牌的表的误差为 ± 0.5 s，A 品牌的表要好一些.

如何刻画一个随机变量的取值与其均值的偏离程度呢？一般地，设 X 是一个离散型随机变量，我们用 $E(X-EX)^2$ 来衡量 X 与 EX 的平均偏离程度，$E(X-EX)^2$ 是 $(X-EX)^2$ 的期望，并称之为随机变量 X 的**方差**，记为 DX. 设离散型随机变量 X 的分布列为 $P(X=a_i)=p_i(i=1,2,3,\cdots,r)$，则

$$DX=\sum_{i=1}^{r}(a_i-EX)^2p_i.$$

方差越小，则随机变量的取值就越集中在其均值周围；反之，方差越大，则随机变量的取值就越分散. 例如，A 品牌手表日走时误差的方差为 $E(X-EX)^2=E(X-0)^2=(-0.01)^2\times\frac{1}{3}+(0.01)^2\times\frac{1}{3}\approx 0.000\ 067$，$B$ 品牌手表日走时误差的方差为 $E(Y-EY)^2=E(Y-0)^2=(-0.5)^2\times\frac{1}{3}+(0.5)^2\times\frac{1}{3}\approx 0.167$. 显然，$A$ 品牌的表质量要好.

例 3　掷一颗均匀的骰子，用 X 表示所得的点数. 求方差 DX.

解　在前面我们已得到 $EX=\frac{7}{2}$. DX 是随机变量 $(X-EX)^2$ 的期望，表 2-8 给出了它的分布列.

表 2-8

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$X-\frac{7}{2}$	$-\frac{5}{2}$	$-\frac{3}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{5}{2}$
$\left(X-\frac{7}{2}\right)^2$	$\frac{25}{4}$	$\frac{9}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{9}{4}$	$\frac{25}{4}$

由此易得 $E(X-EX)^2$，即

$$DX=\frac{25}{4}\times\frac{1}{6}+\frac{9}{4}\times\frac{1}{6}+\frac{1}{4}\times\frac{1}{6}+\frac{1}{4}\times\frac{1}{6}+\frac{9}{4}\times\frac{1}{6}+\frac{25}{4}\times\frac{1}{6}$$
$$=\frac{1}{6}\left(\frac{25}{4}+\frac{9}{4}+\frac{1}{4}+\frac{1}{4}+\frac{9}{4}+\frac{25}{4}\right)$$
$$=\frac{35}{12}\approx 2.92.$$

习 题 2—5

A 组

1. 已知 X 是掷两个均匀骰子的点数之和，求 X 的均值.
2. 设 X 只取 0，1 两个值，并且 $P(X=0)=1-p$，$P(X=1)=p$，试计算 X 的方差.
3. 有甲、乙两种棉花，从中各抽取等量的样品进行检验，结果如下：

$X_{甲}$	28	29	30	31	32
P	0.1	0.15	0.5	0.15	0.1

$X_{乙}$	28	29	30	31	32
P	0.13	0.17	0.4	0.17	0.13

其中 X 表示纤维长度(单位：mm)，根据纤维长度的期望和方差比较两种棉花的质量.

4. 某商场要根据天气预报来决定节日是在商场内还是在商场外开展促销活动. 统计资料表明，每年国庆节商场内的促销活动可获得经济效益 2 万元，商场外的促销活动如果不遇到有雨天气可获得经济效益 10 万元，遇到有雨天气则带来经济损失 4 万元. 9 月 30 日气象台预报国庆节当地有雨的概率是 40%，商场应该选择哪种促销方式?
5. 汽车保险公司每年向顾客收 500 元的保险费，公司通过调查历史档案知道，每年约 8%的顾客要求索赔，而平均赔款额为 1 200 元. 公司每年从每位顾客那里得到的平均收益是多少?

B 组

某住宅小区将要新建一所中学和一所小学. 一建筑公司考虑投标. 由于种种原因该建筑公司只能完成其中一项工程. 假设它投标建中学，需花费 4 000 元的投标准备费，中标的机会为$\frac{1}{5}$，若中标可获得 20 万元的收益；若投标建小学，需 2 000 元准备费，中标的概率是$\frac{1}{4}$，若中标可获得 16 万元的收益. 该公司应向哪一项工程投标?

* §6　正态分布

6.1　连续型随机变量

问题提出

前面讨论了离散型随机变量，它们的取值是可以一一列举的.但在实际应用中，还有许多随机变量可以取某一区间中的一切值.例如：

1. 某一自动装置无故障运转的时间 X 是一个随机变量，它可以取区间$(0,+\infty)$内的一切值.

2. 某种产品的寿命(使用时间)X 是一个随机变量，它可以取$[0,b]$或$[0,+\infty)$内的一切值.

怎样描述这样的随机变量的分布情况呢?

分析理解

我们来看一个产品寿命的例子.设 x 表示某产品的寿命(单位：h)，如果人们对该产品有如下的了解：寿命小于 500 h 的概率为0.71，寿命在 500～800 h 之间的概率为0.22，寿命在 800～1 000 h 之间的概率为 0.07，则我们可以画出图 2-11.但是此图是比较粗糙的，例如，它没有告诉我们产品寿命在 200～400 h 之间的概率到底是多少.如果了解得更多，图中的区间会分得更细，如图 2-12.为了完全了解产品寿命的分布情况，需要将区间无限细分，最终得到一条曲线，如图 2-13 所示.这条曲线称为随机变量 X 的分布密度曲线，这条曲线对应的函数称为 X 的分布密度函数，记为 $f(x)$.

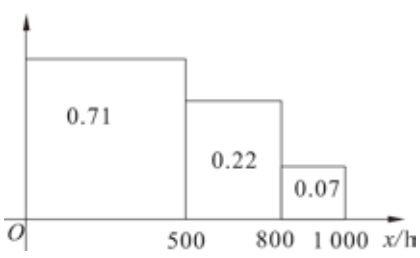


图 2-11

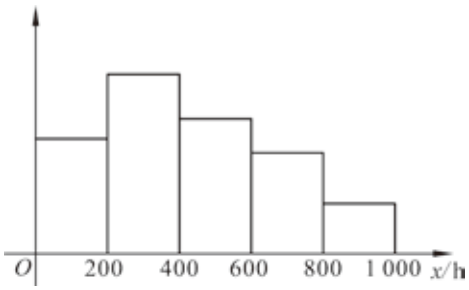


图 2-12

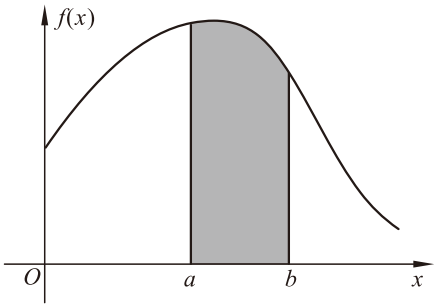


图 2-13

如果知道了 X 的分布密度曲线，则 X 取值于任何范围（例如 $\{a<X<b\}$）的概率，都可以通过计算该曲线下相应部分的面积而得到，因此，我们说 X 的分布密度函数 $f(x)$ 完全描述了 X 的规律. 计算面积，实际上是计算分布密度函数 $f(x)$ 在一个区间上的定积分.

6.2 正态分布

正态分布❶是现实中最常见的分布，它有两个重要的参数：均值 μ 和方差 $\sigma^2(\sigma>0)$，通常用 $X\sim N(\mu,\sigma^2)$ 表示 X 服从参数为 μ 和 σ^2 的正态分布. 当 μ 和 σ^2 给定后，就是一个具体的正态分布. 图 2-14 分别是 $N(0,0.5^2)$，$N(0,1)$，$N(0,2^2)$ 的分布密度曲线.

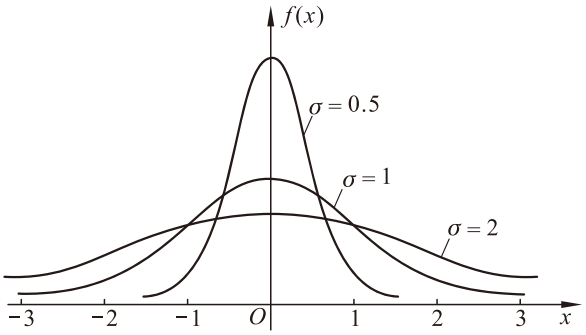


图 2-14

❶正态分布的分布密度函数为：$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$，$-\infty<x<\infty$. 其中 $\exp\{g(x)\}=e^{g(x)}$.

抽象概括

正态分布密度函数满足以下性质：

(1)函数图像关于直线 $x=\mu$ 对称；

(2)$\sigma(\sigma>0)$ 的大小决定函数图像的“胖”“瘦”；

(3)如图 2-15 所示.

$$P(\mu-\sigma<X<\mu+\sigma)=68.3\%,$$

$$P(\mu-2\sigma<X<\mu+2\sigma)=95.4\%,$$

$$P(\mu-3\sigma<X<\mu+3\sigma)=99.7\%.$$

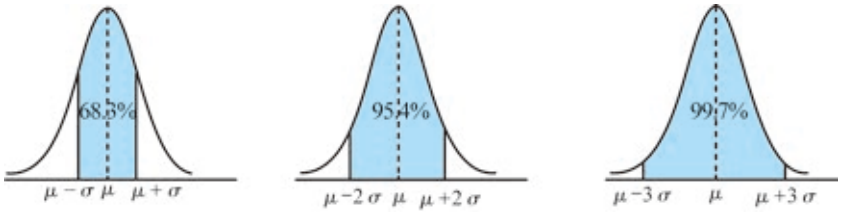


图 2-15

从上面的结论中可以看到，随机变量服从正态分布，则它在区间$(\mu-2\sigma,\mu+2\sigma)$外取值的概率只有 4.6%，而在区间$(\mu-3\sigma,\mu+3\sigma)$外取值的概率只有 0.3%，由于这些概率值很小，通常称这些情况发生为小概率事件. 也就是说，通常认为这些情况在一次试验中几乎不可能发生.

正态分布之所以特别重要，原因是正态分布是现实中最常见的分布，而且，正态分布有许多很好的性质. 各种误差多是服从正态分布的，这使得现实中的许多量，如长度、质量、噪声等的随机波动都可以用正态分布来描述. 甚至当 n 很大时，二项分布也可以用正态分布来近似描述. 在实际应用中，有时人们会不加检验地认为他们讨论的随机变量(特别是长度、质量等)是服从正态分布的.

例　某设备在正常运行时，产品的质量服从正态分布，其参数为：$\mu=500$ g，$\sigma^2=1$. 为了检查设备运行是否正常，质量检查员需要随机地抽取产品，测量其质量. 当检查员随机地抽取一个产品，测得其质量为 504 g，他立即要求停止生产，检查设备. 他的决定是否有道理呢?

解　如果设备正常运行，产品质量服从正态分布. 由于正态分布的参数为：$\mu=500$ g，$\sigma^2=1$. 根据正态分布的性质(3)可知，产品质量在 $\mu-3\sigma=500-3=497$(g)和 $\mu+3\sigma=500+3=503$(g)之间的概率为 0.997，而质量超出这个范围的概率只有 0.003，这是一个几乎不可能出现的事件. 但是，检查员随机抽取的产品为 504 g，这说明设备的运行极可能不正常，检查员的决定是有道理的.

这个例子反映了质量控制的基本思想，它在实际中有广泛的应用.

阅读材料

正态分布小史及其他

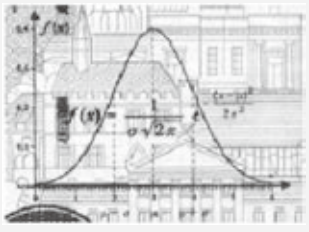
正态分布最早是由棣莫弗在 1733 年的一篇文章中引入的，当时是作为二项分布当试验次数增加时的近似分布. 我们现在借助高尔顿发明的钉板可以直观地看到怎样从二项分布到正态分布.

高尔顿钉板上有多排交错成三角状的钉子，当小球（如豌豆）从上方往下落时，在碰到每一排的钉子后都有两种可能的结果：向左或向右拐后继续往下落，直至落入下方的容器中，因此小球落入各容器的情形可以用二项分布来刻画. 当许多小球从上往下落时，各容器中的小球的多少（累积高度）便反映了二项分布取值的比例. 随着落入的球的增加，我们可以发现容器中小球呈中间多，两端少的“钟形”，或说近似于正态曲线.

在此后的几十年间，棣莫弗的工作并没有受到人们的关注. 直至 1778 年拉普拉斯才重新发现了正态分布. 拉普拉斯推广了棣莫弗的结果，证明了当每一个小的误差与总的误差相比可以忽略不计时，不管小的误差的分布是什么，总的误差将近似服从正态分布. 这一著名的结论说明了为什么现实中如此众多的随机现象可以用正态分布来描述其规律. 例如，自动火炮命中目标的误差一般认为是服从正态分布的，这个误差是风速、射击的方向和角度、弹药的质量等许许多多的因素共同影响的结果. 其中每一种因素，人们都努力去控制以至于都不起主要作用，但这些微小的误差数量之多，使得其总和仍起作用，最终造成了命中目标的误差.

正态分布的一个最早的应用是用来分析天文观测中的误差. 在十七八世纪，由于不完善的仪器以及观测人员缺乏经验等原因，天文观察误差是一个重要的问题，有许多重要的科学家都进行过研究. 1809 年，高斯指出正态分布可以很好地“拟合”误差分布. 基于误差分布服从正态分布的假设，高斯奠定了他此前使用过的最小二乘法的数学基础. 为纪念他的贡献，正态分布也称为高斯分布.

19 世纪，比利时统计学家魁特奈特了解到正态分布后，倡导并身体力行将正态分布用于数据的分析. 由于他的这一努力，正态分布在 19 世纪的统计应用中大为流行.



纸币中央的正态曲线

◆ 本章小结建议

一、学习要求

1. 理解离散型随机变量及其分布的概念，体会分布列对于刻画随机现象的重要性.

2. 理解超几何分布及其导出过程，并能进行简单的应用.

3. 了解条件概率和两个事件相互独立的概念，理解 n 次独立重复试验的模型及二项分布，并能解决一些简单的实际问题.

4. 理解离散型随机变量均值、方差的概念，能计算简单离散型随机变量的均值、方差，并能解决一些实际问题.

5. 借助直观（如实际问题的直方图），认识正态分布曲线的特点及曲线所表示的意义.

二、复习本章知识，整理笔记，建议就以下问题思考、归纳、概括，写出复习小结报告，并与同学互相交流

1. 为什么要引入随机变量的概念？举例说明离散型随机变量的含义.

2. 离散型随机变量的分布列对于刻画随机现象有什么重要意义？

3. 二项分布、超几何分布是如何定义的？这两个分布中分别包括哪些参数？它们的含义分别是什么？

4. 举出一些符合二项分布、超几何分布的实例，并说明为什么所举实例是符合某一分布的. 与同学交流对“二项分布、超几何分布是两个重要的概率模型”的体会.

5. 什么是离散型随机变量的均值和方差？请找一个生活中的例子说明它们的用处.

6. 正态分布曲线有什么特点？借助书中的图加以说明.

7. 学习完本章后，你有哪些印象深刻的例子？还有哪些不理解的问题？你能在生活中找一个例子，并尝试用本章知识来解决它吗？

8. 尝试运用自己的语言，借助实例，说明随机现象的特点及研究随机现象规律的意义.

复 习 题 二

A 组

1. 连续投掷一枚均匀的硬币 3 次，用 X 表示这 3 次投掷中反面朝上的次数，则 X 是一个随机变量. 分别说明下列集合所代表的随机事件：

   (1) $\{X=0\}$；　　(2) $\{X=3\}$；　　(3) $\{X\leqslant 2\}$；　　(4) $\{X>0\}$.

2. 商店经理要合理地安排售货员的人数. 安排多少名售货员依赖于顾客的人数，而顾客的人数是随机的，事先无法确定. 如果假定商店经理知道任一时刻来到 k 名顾客的概率 p_k 如下：

k	0	1	2	3	4	5	6	7	>7
p_k	0.03	0.10	0.14	0.19	0.21	0.19	0.09	0.04	0.01

   (1) 安排 3 名售货员能以多大概率使顾客不用等待？

   (2) 安排多少名售货员能以 99%的概率使顾客不用等待？

3. 连续投掷一枚均匀的硬币 6 次，全是正面朝上的概率有多大？第 7 次掷出正面的概率是多少？

4. 4 个人玩一副扑克牌(去掉大、小王，共 52 张)，求某个人手中正好抓到 6 张黑桃的概率.

5. 在一个纸箱中装有 12 个电灯泡，其中有 3 个是有缺陷的. 从纸箱中任意挑出 5 个电灯泡，用 X 表示其中有缺陷产品的个数. 试求 X 的分布列、均值、方差.

6. 在一个池塘中有 1 000 条鱼，其中有 100 条草鱼. 现从中捕出 20 条鱼，试计算其中至少有两条草鱼的概率.

7. 袋中装有 1 个红球和 4 个黑球，这些球除颜色外完全相同.

   (1) 从袋中任意摸出 1 个球，摸出红球的概率是多少？

   (2) 现在有放回地摸 5 次，“恰摸出 1 次红球”的概率是多少？

8. 某射手独立地进行 5 次射击，设各次中靶的概率都是 0.8. 试求下列各事件的概率：

   (1) 5 次都中靶；　　(2) 5 次都没中靶；

   (3) 前 3 次中靶，后 2 次没中靶；　　(4) 恰有 3 次中靶.

9. 一气球制造公司生产的气球 95%是合格的(充气后不爆破). 假设在你的生日聚会上准备了 20 个该公司生产的气球.

   (1) 这些气球充气后没有一个爆破的概率是多少？

   (2) 恰好有两个气球爆破的概率是多少？

   (3) 超过三个气球爆破的概率是多少？

10. 投掷一枚不均匀硬币出现正面的概率是 0.45. 若用它赌博，掷出正面时，赢 3 元；掷出反面时，输 2 元. 求掷一次的平均赢利.

11. 一位司机从饭店到火车站途中要经过 6 个交通岗，假设他在各交通岗遇到红灯是相互独立的，并且概率都是 1/3. 求这位司机在途中遇到红灯数 X 的期望和方差.

B　组

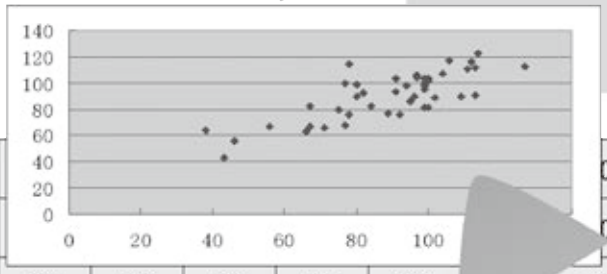
1. 一位国王的铸币大臣在每箱 100 枚的硬币中各掺入了一枚劣币. 国王怀疑大臣作弊，他准备在 10 箱硬币中各任意抽查一枚. 国王能发现至少一枚劣币的概率是多少？如果他在 5 箱硬币中各任意抽查两枚呢？
2. 赌博时，赢 a 元钱的概率为 p_1，输 b 元钱的概率为 p_2，不输不赢的概率为 p_3，这里 $p_1+p_2+p_3=1$.
   (1) 求赌博的平均赢利；
   (2) 当平均赢利等于 0 时，称赌博是公平的. 当 $p_1=\frac{1}{3}$，$p_2=\frac{2}{3}$，$p_3=0$，$a=1$ 时，b 取何值赌博是公平的？

第三章 统计案例

“一个家族中兄弟或姐妹智商的相关性如何?”“吸烟与患肺癌是否有关?”……这些都是日常生活中常见的一些问题.

本章将在必修课程学习统计的基础上,通过对以上问题的讨论,了解和使用一些常见的统计方法,进一步体会运用统计方法解决实际问题的基本思想以及统计方法使用的广泛性.

兄弟智商散点图



x	78						0	113
y	114						0	91
x	99	97	80	84	89	100	75	94
y	95	106	99	82	77	81	111	98
x	67					27	113	91
y	82					113	112	93
x	96					77	109	99
y						100	90	103
x						66	78	95	38
y						63	76	86	64

§1　回归分析

通过必修阶段统计内容的学习，我们已经认识到了现实生活中存在着某些有关系的不同变量，这些变量之间的关系不是可以用函数表示的确定性关系．例如，父母的身高与他们孩子的身高，食物中所含的脂肪与所含的热量，模拟测验的成绩与实际考试的成绩，农作物的施肥量与产量．它们之间是一种非确定性关系，称为相关关系．例如，施肥量无疑是影响农作物产量的重要因素，但不是唯一因素，农作物的产量还与农作物的栽培方式、气候等因素有关，有些因素是人为可以控制的，而有些则无法控制．

由于一个变量与另一个变量之间往往不是确定的关系，人们也不可能把握与某个变量有关的所有变量，因此变量间的关系往往会表现出某种不确定性．回归分析就是研究这种变量之间的关系的一种方法，通过对变量之间关系的研究，从而发现蕴涵在事物或现象中的某些规律．

1.1　回归分析

在必修课程中，我们已经学习了最小二乘法，并会用它建立变量之间的线性回归方程．

例　始祖鸟是一种已经灭绝的动物．在一次考古活动中，科学家发现了始祖鸟的化石标本共 6 个，其中 5 个同时保有股骨（一种腿骨）和肱骨（上臂的骨头）．科学家检查了这 5 个标本股骨和肱骨的长度，得到表 3-1 中的数据：

表 3-1

编　号	1	2	3	4	5
股骨长度 x/cm	38	56	59	64	74
肱骨长度 y/cm	41	63	70	72	84

（1）求出肱骨长度 y 对股骨长度 x 的线性回归方程；

（2）还有 1 个化石标本不完整，它只有股骨，而肱骨不见了．现测得股骨的长度为 50 cm，请预测它的肱骨长度．

分析理解

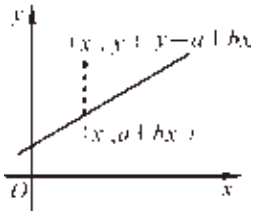


图 3-1

必修课程中，我们已经会用最小二乘法求变量之间的线性回归方程. 假设样本点为$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，设线性回归方程为$y=a+bx$，我们的想法就是要求$a,b$，使这$n$个点与直线$y=a+bx$的“距离”平方之和最小，即使得

$$Q(a,b)=(y_1-a-bx_1)^2+(y_2-a-bx_2)^2+\cdots+(y_n-a-bx_n)^2$$

达到最小.（如图 3-1 所示）

在统计上，我们使用$\bar{x}$表示一组数据$x_1,x_2,\cdots,x_n$的平均值，即$\bar{x}=\dfrac{x_1+x_2+\cdots+x_n}{n}$，为了简化表示，我们引进求和符号，记作$\bar{x}=\dfrac{x_1+x_2+\cdots+x_n}{n}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$. 同理有$\bar{y}=\dfrac{y_1+y_2+\cdots+y_n}{n}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i$.

这样就有：

$$\sum_{i=1}^{n}(x_i-\bar{x})=\sum_{i=1}^{n}x_i-n\bar{x}=n\bar{x}-n\bar{x}=0,$$

$$\sum_{i=1}^{n}(y_i-\bar{y})=\sum_{i=1}^{n}y_i-n\bar{y}=n\bar{y}-n\bar{y}=0.$$

为了简化上面的表示，我们引入以下记号：

$$l_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2,$$

$$l_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y},$$

$$l_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-n\bar{y}^2.$$

从而

$$\begin{aligned}
&Q(a,b)\\
&=(y_1-a-bx_1)^2+(y_2-a-bx_2)^2+\cdots+(y_n-a-bx_n)^2\\
&=\sum_{i=1}^{n}(y_i-a-bx_i)^2\\
&=\sum_{i=1}^{n}\{(y_i-\bar{y})+[\bar{y}-(a+b\bar{x})]-b(x_i-\bar{x})\}^2\\
&=\sum_{i=1}^{n}(y_i-\bar{y})^2+n[\bar{y}-(a+b\bar{x})]^2+b^2\sum_{i=1}^{n}(x_i-\bar{x})^2+2[\bar{y}-(a+\\
&\quad b\bar{x})]\sum_{i=1}^{n}(y_i-\bar{y})-2b\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})-2b[\bar{y}-(a+\\
&\quad b\bar{x})]\sum_{i=1}^{n}(x_i-\bar{x})\\
&=l_{yy}+n[\bar{y}-(a+b\bar{x})]^2+l_{xx}b^2-2l_{xy}b
\end{aligned}$$

$$= l_{yy} + n[\bar{y} - (a + b\bar{x})]^2 + l_{xx}\left(b - \frac{l_{xy}}{l_{xx}}\right)^2 - \frac{{l_{xy}}^2}{l_{xx}}.$$

当 $\bar{y} - (a + b\bar{x}) = 0$ 且 $b - \frac{l_{xy}}{l_{xx}} = 0$ 时，$Q(a,b)$ 取最小值，此时

$$b = \frac{l_{xy}}{l_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n}{x_i}^2 - n\bar{x}^2},$$

$$a = \bar{y} - b\bar{x}.$$

解：(1) 从散点图 3-2 可以看出，表 1-1 中的两个变量呈现出近似的线性关系，我们可以建立肱骨长度 y 对股骨长度 x 的线性回归方程.

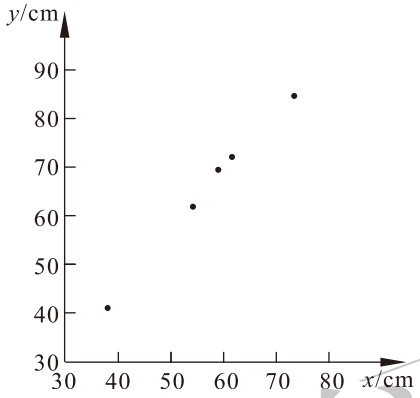


图 3-2

根据上面的分析，我们将数据列出如表 3-2 所示：

表 3-2

i	x_i	y_i	${x_i}^2$	$x_i y_i$
1	38	41	1 444	1 558
2	56	63	3 136	3 528
3	59	70	3 481	4 130
4	64	72	4 096	4 608
5	74	84	5 476	6 216
$\sum$	291	330	17 633	20 040

由此可得：$\bar{x} = \frac{291}{5} = 58.2$，$\bar{y} = \frac{330}{5} = 66$. 进而可以求得

$$b = \frac{20\,040 - 5 \times 58.2 \times 66}{17\,633 - 5 \times 58.2^2} \approx 1.197,$$

$$a = 66 - \frac{20\,040 - 5 \times 58.2 \times 66}{17\,633 - 5 \times 58.2^2} \times 58.2 \approx -3.660.$$

于是，y 对 x 的线性回归方程为

$$y = -3.660 + 1.197x.$$

回归直线的斜率 $b=1.197$ 的意思是，对于这次发现的始祖鸟的化石标本来说，股骨的长度每增加 1 cm，肱骨的长度平均增加 1.197 cm.

(2)由上面最小二乘法得到的线性回归方程可知，当股骨的长度为 50 cm 时，肱骨长度的估计值为

$$-3.660+1.197\times 50=56.19\approx 56(\text{cm}).$$

练　习

研究某灌溉渠道水的流速 y 和水深 x 之间的关系，测量得到的数据如下：

水深 x/m	1.40	1.50	1.60	1.70	1.80	1.90	2.00	2.10
流速 y/(m/s)	1.70	1.79	1.88	1.95	2.03	2.10	2.16	2.21

(1) 求出流速 y 对水深 x 的线性回归方程；

(2) 预测水深为 1.85 m 时水的流速是多少.

1.2　相关系数

问题提出

我们知道，任何数据，不管它们的线性相关关系如何，都可以用最小二乘法求出线性回归方程. 为使建立的线性回归方程有意义，在利用最小二乘法求线性回归方程之前，我们先要对变量之间的线性相关关系作一个判断，通常可以作数据的散点图.

但在某些情况下，从散点图中不容易判断变量之间的线性关系，另外，当数据量较大时，画散点图比较麻烦，此时我们有没有其他方法来刻画变量之间的线性相关关系呢?

为了解决以上问题，我们可以通过计算两个随机变量间的线性相关系数 r，来判断它们之间线性相关程度的大小.

假设两个随机变量的数据分别为 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，则变量间线性相关系数 r 的计算公式如下：

$$r=\frac{l_{xy}}{\sqrt{l_{xx}l_{yy}}}=\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}\sqrt{\sum_{i=1}^{n}y_i^2-n\bar{y}^2}}.$$

根据前面的分析，回归方程的系数 a，b 使误差达到最小.

误差可以表示为

$$\begin{aligned}Q(a,b)&=\sum_{i=1}^{n}[y_i-(a+bx_i)]^2\\&=l_{yy}+n[\bar{y}-(a+b\bar{x})]^2+l_{xx}\left(b-\frac{l_{xy}}{l_{xx}}\right)^2-\frac{l_{xy}^{\ 2}}{l_{xx}}.\end{aligned}$$

其最小值为

$$Q=l_{yy}-\frac{l_{xy}^{\ 2}}{l_{xx}}=l_{yy}\left(1-\frac{l_{xy}^{\ 2}}{l_{yy}l_{xx}}\right)=l_{yy}(1-r^2).$$

由于 $Q\geqslant 0$，从而 $r^2\leqslant 1$. 故变量之间线性相关系数 r 的取值范围为 $[-1,1]$. $|r|$ 值越大，误差 Q 越小，变量之间的线性相关程度越高；$|r|$ 值越接近 0，Q 越大，变量之间的线性相关程度越低. 当 $r>0$ 时，$l_{xy}>0$，从而 $b=\frac{l_{xy}}{l_{xx}}>0$，两个变量的值总体上呈现出同时增减的趋势，此时称两个变量正相关；当 $r<0$ 时，$b<0$，一个变量增加，另一个变量有减少的趋势，称两个变量负相关；当 $r=0$ 时，称两个变量线性不相关.

思考交流

(1) 如何求出例题(73 页)中变量的线性相关系数 r?

对于例题，由表 3-2 可得：$\sum_{i=1}^{n}x_i^2=17\ 633$，$\sum_{i=1}^{n}y_i^2=22\ 790$，$\sum_{i=1}^{n}x_iy_i=20\ 040$；$\bar{x}=\frac{291}{5}=58.2$，$\bar{y}=\frac{330}{5}=66$. 进而可以求得

$$\begin{aligned}r&=\frac{\sum_{i=1}^{n}x_iy_i-n\,\bar{x}\,\bar{y}}{\sqrt{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}\sqrt{\sum_{i=1}^{n}y_i^2-n\bar{y}^2}}\\&=\frac{20\ 040-5\times 58.2\times 66}{\sqrt{17\ 633-5\times 58.2^2}\times\sqrt{22\ 790-5\times 66^2}}\approx 0.994\ 1.\end{aligned}$$

由此可以得出，肱骨长度 y 和股骨长度 x 有较强的线性相关程度.

(2) 请计算表 3-3 中变量的线性相关系数 r，通过计算，发现了

什么？

表 3-3

x	-5	-4	-3	0	3	4	5
y	0	3	4	5	4	3	0

根据表 3-3 的数据，列表如表 3-4 所示：

表 3-4

i	x_i	y_i	x_i^2	y_i^2	x_iy_i
1	-5	0	25	0	0
2	-4	3	16	9	-12
3	-3	4	9	16	-12
4	0	5	0	25	0
5	3	4	9	16	12
6	4	3	16	9	12
7	5	0	25	0	0
$\sum$	0	19	100	75	0

由此可得：$\sum\limits_{i=1}^{n}x_i^2=100$，$\sum\limits_{i=1}^{n}y_i^2=75$，$\sum\limits_{i=1}^{n}x_iy_i=0$，$\bar{x}=0$，$\bar{y}=2.71$. 进而可以求得

$$r=\frac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}}{\sqrt{\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2}\sqrt{\sum\limits_{i=1}^{n}y_i^2-n\bar{y}^2}}=\frac{0-7\times0\times2.71}{\sqrt{100-7\times0^2}\times\sqrt{75-7\times2.71^2}}=0.$$

从散点图 3-3 容易看出，表格中的数据都在同一个半圆上，此时建立线性回归方程是没有任何意义的，这与线性相关系数 r 的计算结果是一致的.

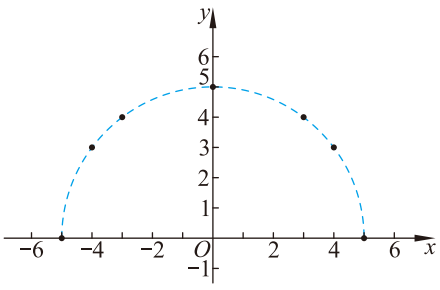


图 3-3

实际上，线性相关系数 r 越大，变量之间的线性关系就越强，用直线拟合的效果就越好. 相关系数 r 的值究竟大到什么程度就认为线性关系较强，同学们可以参考有关的统计书籍进一步学习！

练　习

许多先进国家对驾驶员的培训，大多采用室内模拟教学和训练，而后再进行实地训练并考试，这种方法可以大大节约训练的费用. 问题是这种方法有效吗？下表是 12 名学员的模拟驾驶成绩 x 与实际考试成绩 y 的记录(单位：分)：

x	98	55	50	87	77	89
y	95	60	45	85	75	87
x	79	98	94	83	74	73
y	75	97	92	80	71	72

试问：两者的相关性如何？请画出散点图，并求出 y 与 x 间的线性相关系数.

1.3　可线性化的回归分析

问题提出

表 3-5 按年代给出了 1981～2001 年我国出口贸易量(亿美元)的数据，你能根据此表预测 2008 年我国的出口贸易量吗？

表 3-5　1981～2001 年我国出口贸易量数据

年份	1981	1982	1983	1984	1985	1986	1987
出口贸易量/亿美元	220.1	223.2	222.3	261.4	273.5	309.4	394.4
年份	1988	1989	1990	1991	1992	1993	1994
出口贸易量/亿美元	475.2	525.4	620.9	719.1	849.4	917.4	1 210.1
年份	1995	1996	1997	1998	1999	2000	2001
出口贸易量/亿美元	1 487.8	1 510.5	1 827.9	1 837.1	1 949.3	2 492.0	2 661.0

要预测 2008 年我国的出口贸易量，首先应根据表中的数据找到出口贸易量与年份之间的关系.

我们在必修阶段已经详细讨论了用直线方程 $y=a+bx$ 来描述两个变量的关系. 然而由图 3-4 可以看出，出口贸易量与年份之间呈现出一种非线性的相关性. 如果用直线来描述，其结果如图 3-5 所

示，显然效果不太好，若用它来作预测，误差将会很大. 于是，我们考虑用非线性函数来描述图 3-4 中数据的变化关系.

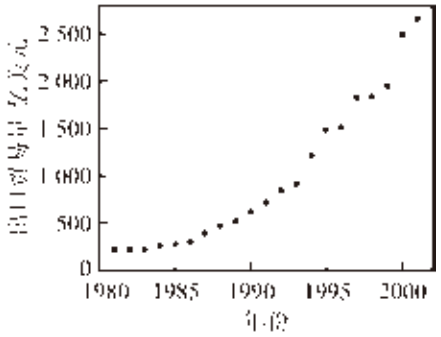


图 3-4

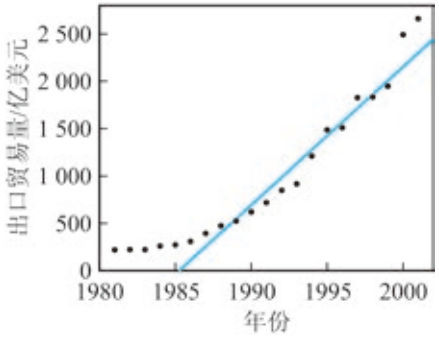


图 3-5

分析理解

从原始数据的散点图（图 3-4）可以看出，图像近似一个指数函数，我们可以考虑用函数

$$y=ae^{bx}$$

来拟合数据的变化关系，但如何进行拟合呢？我们可以先将其转化成线性函数，这就需要事先作个变换：对上式两边取对数，得

$$\ln y=\ln a+bx.$$

若记

$$u=\ln y,\ c=\ln a,$$

则上式就变成了

$$u=c+bx,$$

即回到我们的线性回归方程，于是就可以用最小二乘法来进行计算. 为了方便建立回归方程，我们把年份 1981 记为 $x=1$，年份 1982 记为 $x=2$，……因此，我们实际上是对 $(x,u=\ln y)$ 作线性回归，其中 y 表示原始观测值. 将表 3-5 中的数据经过上述变换后即可得到表 3-6 中的数据.

表 3-6

x	1	2	3	4	5	6	7
u	5.394	5.408	5.404	5.566	5.611	5.735	5.977
x	8	9	10	11	12	13	14
u	6.164	6.264	6.431	6.578	6.745	6.822	7.098
x	15	16	17	18	19	20	21
u	7.305	7.320	7.511	7.516	7.575	7.821	7.886

我们对表 3-6 中的(x,u)求线性回归方程，可得：$c=5.056$，$b=0.138$，即 $u=5.056+0.138x$(如图 3-6 所示).

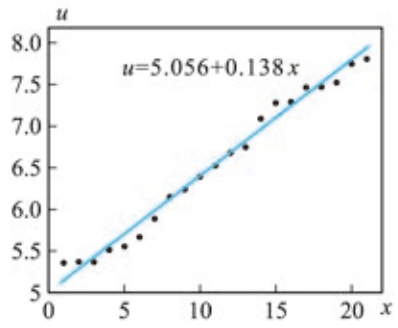


图 3-6

由此可得：$y=e^{u}=e^{5.056}\cdot e^{0.138x}$，在图 3-4 中画出函数曲线，如图 3-7 所示.

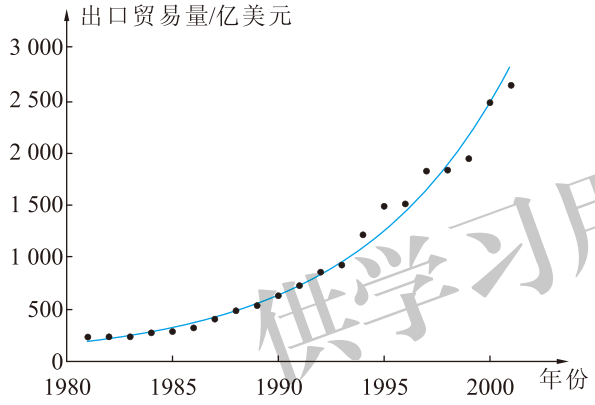


图 3-7

现在，你能预测出 2008 年我国的出口贸易量吗?

抽象概括

上面我们用非线性函数 $y=ae^{bx}$ 对我国 1981～2001 年的出口贸易量进行了拟合. 其方法是通过变换先将其转化成线性函数，利用最小二乘法得到线性回归方程，再通过相应变换得到非线性回归方程.

然而上述函数仅是非线性的一种，还有其他经过变换也可以变成线性函数的非线性函数. 我们如何将一些常见的非线性回归模型转化为线性回归模型，从而得到相应的回归方程呢?

1. 幂函数曲线 $y=ax^b$(如图 3-8 所示).

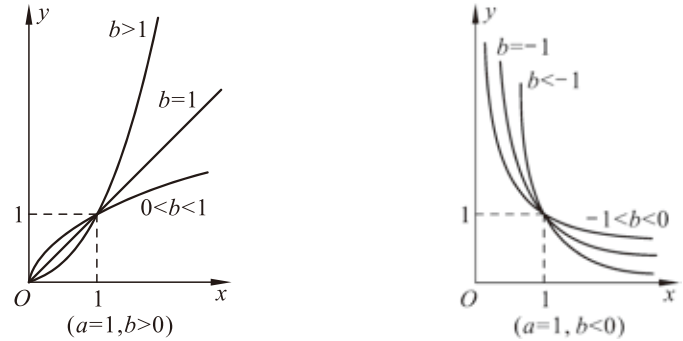


图 3-8

作变换 $u=\ln y, v=\ln x, c=\ln a$,得线性函数 $u=c+bv$.

2. 指数曲线 $y=ae^{bx}$(如图 3-9 所示).

图 3-9

作变换 $u=\ln y, c=\ln a$,得线性函数 $u=c+bx$.

3. 倒指数曲线 $y=ae^{\frac{b}{x}}$(如图 3-10 所示).

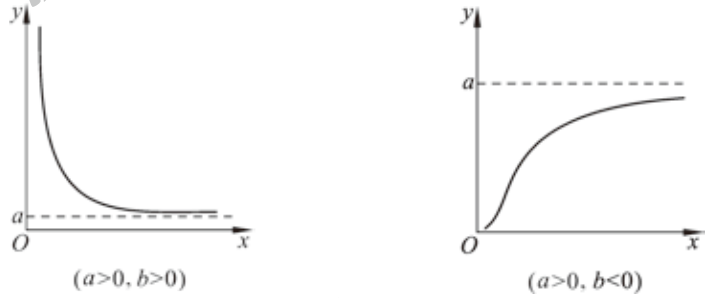


图 3-10

4. 对数曲线 $y=a+b\ln x$(如图 3-11 所示).

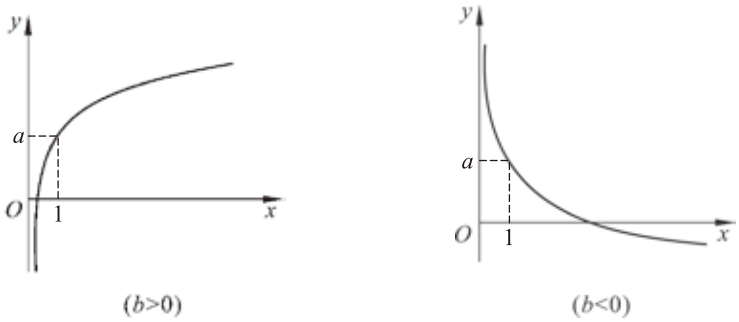


图 3-11

思考交流

对于上面的 3,4,你能通过适当的变换,分别将其转化成线性函数吗?与同伴进行交流.

在具体问题中,我们首先应该作出原始数据(x,y)的散点图,从散点图中看出数据的大致规律,再根据这个规律选择适当的函数进行拟合.

练　习

某矿脉中设有 13 个相邻样本点,现人为地设定一个原点,测得各样本点与原点的距离 x,与该样本点处某种金属的含量 y 的数据如下:

x	2	3	4	5	7	8	10
y	106.42	108.20	109.58	109.50	110.00	109.93	110.49
x	11	14	15	15	18	19	
y	110.59	110.60	110.90	110.76	111.00	111.20	

请按 $y=a+\frac{b}{x}$ 建立 y 对 x 的回归方程,并预测当样本点对原点的距离 x 为 20 时该种金属的含量.

阅读材料

高尔顿与回归

高尔顿(Francic Galton,1822—1911)

“回归”这一名称是英国生物学家兼统计学家高尔顿在1866年左右提出来的.人们大概都注意到,儿子的身高与其父亲的身高有关.高尔顿以父亲的平均身高 x 为自变量,其成年儿子的平均身高 y 为因变量,研究了 1 074 对父亲及其成年儿子的平均身高,将所得 (x,y) 值作为点的坐标,把这些点标在直角坐标系中,发现二者的关系近乎一条直线.总的趋势是当 x 增加时,y 也倾向于增加——这是意料中的结果.有意思的是,高尔顿对所得数据作了深入一层的考察,从而发现了一个有趣的现象.

高尔顿计算出这 1 074 个 x 的平均值为 $\bar{x}=68$ in❶,而1 074个 y 的平均值为 $\bar{y}=69$ in,子代身高平均增加了1 in.由此,人们会有这样的推测:平均身高为 72 in 的父亲,其儿子的平均身高应为 73 in;类似地,平均身高为 64 in 的父亲,其儿子的平均身高应为 65 in,等等.但高尔顿实际观察的结果与此不符.他发现:当父亲的平均身高为 72 in 时,其儿子的平均身高只有71 in,不仅达不到预计的 73 in,反而比父亲的平均身高矮了.反之,当父亲的身高为 64 in 时,观察数据显示其儿子的平均身高为 67 in,比预计的 65 in 要高.高尔顿对此的解释是:大自然有一种约束机制,使人类身高的分布保持某种稳定形态而不向两极分化.这就是一种使身高“回归于中心”的作用.例如,父亲身高平均为 72 in,比他们这一代平均身高 68 in 高出许多,“回归于中心”的力量把他们子代的身高拉回来一些:其平均值只有 71 in,反而比父亲平均身高矮,但仍超过子代全体平均值 69 in.反之,当父亲平均身高只有 64 in——远低于他们这一代的平均值68 in,而“回归于中心”的力量将其子代身高拉回去一些:其平均值达到 67 in,增长了 3 in,但仍低于子代全体平均值 69 in.

正是通过这个例子,高尔顿引入了“回归”这个词.自高尔顿起,“回归”一词一直沿用至今,作为变量关系统计分析的称呼.

---

❶in 表示英寸,1 in=2.54 cm.

习 题 3—1

1. 为研究鲈鱼身长与体重的关系，芬兰某渔业公司得出如下表所示的鲈鱼身长（单位：cm）与体重（单位：g）的记录：

身长 x/cm	30.0	31.2	31.1	33.5	34.0	34.7	34.5	35.0	35.1	36.2
体重 y/g	242.0	290.0	340.0	363.0	430.0	450.0	500.0	390.0	450.0	500.0
身长 x/cm	36.2	36.2	36.4	37.2	37.2	38.3	38.5	38.6	38.7	
体重 y/g	475.0	500.0	500.0	600.0	600.0	700.0	700.0	610.0	650.0	

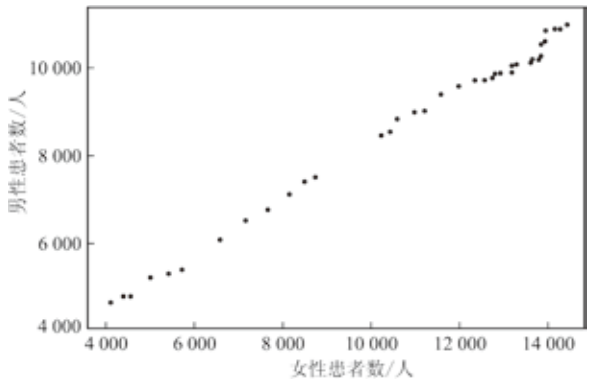
（1）请画出散点图，并求鲈鱼身长与体重间的线性相关系数；

（2）建立线性回归方程，并预测鲈鱼身长为 38 cm 时体重是多少.

2. 下表是德国 1955～1995 年男性与女性患肠癌的逐年病例数的记录. 从常识上看，男性与女性在同一国家，其生活饮食、环境有类似之处，两者患病的可能性也应有相关性. 然而也不是说同一家庭中丈夫得了肠癌，妻子也会得肠癌. 将表中的女性患病人数与对应的男性患病人数画在一张图上就是如图所示的样子. 从图中读者不难看出，随着女性患肠癌人数的增加，男性患者也在增加；但 1990 年和 1994 年两年女性患者人数几乎一样，而男性患者人数却相差很大.

1955～1995 年德国患肠癌的男女的人数

年份	1955	1956	1957	1958	1959	1960	1961	1962	1963
女性	3 936	4 138	4 443	4 594	5 019	5 439	5 710	6 558	7 122
男性	4 356	4 623	4 769	4 769	5 193	5 260	5 390	6 087	6 563
年份	1964	1965	1966	1967	1968	1969	1970	1971	1972
女性	7 641	8 125	8 459	8 719	10 220	10 444	10 588	10 995	11 228
男性	6 781	7 142	7 560	7 451	8 602	8 540	8 921	9 080	9 106
年份	1973	1974	1975	1976	1977	1978	1979	1980	1981
女性	11 581	12 012	12 379	12 771	12 835	13 210	12 612	12 951	12 781
男性	9 475	9 680	10 159	9 966	10 292	10 303	9 816	9 989	9 818
年份	1982	1983	1984	1985	1986	1987	1988	1989	1990
女性	12 837	13 315	13 209	13 684	13 626	13 865	13 821	14 186	13 965
男性	9 861	9 869	9 952	10 196	9 967	10 258	10 410	10 747	10 690
年份	1991	1992	1993	1994	1995				
女性	13 982	14 444	14 286	13 953	13 882				
男性	10 739	11 151	11 021	11 039	11 041				



（第 2 题）

请计算女性患者与男性患者人数的线性相关系数.

3. 一只红铃虫的产卵数 y 与温度 x 有关，下表是产卵数与温度相对应的一组数据，试求 y 与 x 间的回归方程，并预测当温度为 37℃时红铃虫的产卵数.

温度 x/℃	21	23	25	27	29	32	35
产卵数 y/个	7	11	21	24	66	115	325

4. 在彩色显像中，根据以往的经验，知道染料光学密度 y 与析出银的光学密度 x 之间有如下函数关系：$y=ae^{\frac{b}{x}}$ $(b<0)$.

我们通过 11 次试验得到如下数据：

x	0.05	0.06	0.07	0.10	0.14	0.20	0.25	0.31	0.38	0.43	0.47
y	0.10	0.14	0.23	0.37	0.59	0.79	1.00	1.12	1.19	1.25	1.29

试通过拟合，确定函数的参数，并预测当析出银的光学密度为 0.50 时形成染料的光学密度.

§2　独立性检验

2.1　独立性检验

问题提出

我们知道，“吸烟具有危害性”，为什么人们都认可这一观点呢？它的根据又是什么呢？

为了调查吸烟与患肺癌是否有联系，某机构随机调查了 6 578 人，得到表 3-7 中的数据（单位：人）：

表 3-7

吸烟情况 \ 患肺癌情况	患肺癌	未患肺癌
吸烟	56	1 932
不吸烟	23	4 567

上面是一张 2 行 2 列的表，在统计中称为 2×2 列联表. 在这个问题中，考虑两个变量：是否吸烟，是否患肺癌；每个变量取两个值：吸烟、不吸烟和患肺癌、未患肺癌.

表中数据是根据调查得到的结果，如：吸烟且患肺癌的人数是 56，不吸烟但患肺癌的人数是 23，等等. 我们的问题是：如何根据表格中的数据来判断吸烟与患肺癌是否独立，这一问题称为 2×2 列联表的独立性检验.

分析理解

为了讨论的方便，我们引入以下记号：

变量 A：A_1＝吸烟，　　$A_2=\overline{A_1}$＝不吸烟；

变量 B：B_1＝患肺癌，　　$B_2=\overline{B_1}$＝未患肺癌.

根据表 3-7，我们可以计算出吸烟与不吸烟的总人数分别是 1 988和 4 590，患肺癌和未患肺癌的总人数分别是 79 和 6 499，调查的总人数为 6 578. 得到表 3-8（单位：人）：

表 3-8

患肺癌情况 / 吸烟情况	患肺癌 B_1	未患肺癌 B_2	总计
吸烟 A_1	56	1 932	1 988
不吸烟 A_2	23	4 567	4 590
总计	79	6 499	6 578

我们假设吸烟与患肺癌是独立的，即吸烟不影响患肺癌．根据直观的经验，我们把吸烟人群中患肺癌的人所占百分比，与不吸烟人群中患肺癌的人所占百分比作比较. 如果吸烟不影响患肺癌，就意味着，无论吸烟与否，患肺癌的人所占的百分比应是基本一样的. 就此题而言：

吸烟人群中患肺癌的人所占百分比是：$\frac{56}{1\ 988}\approx 2.82\%$；

不吸烟人群中患肺癌的人所占百分比是：$\frac{23}{4\ 590}\approx 0.50\%$.

吸烟人群中患肺癌的人所占百分比，与不吸烟人群中患肺癌的人所占百分比不等，且相差较大. 由此我们可以推断，开始的假设是不成立的. 也就是说，患肺癌与吸烟是有关系的. 由吸烟人群中患肺癌的人所占的百分比较多，我们认为吸烟会对肺癌的发病率造成一定的影响.

另一方面，如果吸烟与患肺癌是独立的，那么有 $P(A_1B_1)=P(A_1)P(B_1)$，$P(A_1B_2)=P(A_1)P(B_2)$，$P(A_2B_1)=P(A_2)P(B_1)$，$P(A_2B_2)=P(A_2)P(B_2)$成立.

我们先讨论 $P(A_1B_1)=P(A_1)P(B_1)$的情况. 我们可以列出频率表，并用既吸烟又患肺癌的人的频率来估计 $P(A_1B_1)$，用吸烟的人的频率来估计 $P(A_1)$，用患肺癌的人的频率来估计 $P(B_1)$，得到表 3-9：

表 3-9

患肺癌情况 / 吸烟情况	患肺癌 B_1	未患肺癌 B_2	总计
吸烟 A_1	$\frac{56}{6\ 578}$	$\frac{1\ 932}{6\ 578}$	$\frac{1\ 988}{6\ 578}$
不吸烟 A_2	$\frac{23}{6\ 578}$	$\frac{4\ 567}{6\ 578}$	$\frac{4\ 590}{6\ 578}$
总计	$\frac{79}{6\ 578}$	$\frac{6\ 499}{6\ 578}$	1

既吸烟又患肺癌的人的频率是：$\frac{56}{6\ 578}\approx 0.85\%$，

吸烟的人的频率是：$\frac{1\ 988}{6\ 578}\approx 30.22\%$，

患肺癌的人的频率是：$\frac{79}{6\ 578}\approx 1.20\%$.

显然，$30.22\%\times1.20\%=0.36\%\neq0.85\%$，由于根据表中数据计算出的值是频率值，它只是概率的估计值，因此即使变量之间独立，这两个数一般也不一定恰好相等. 但是当两边相差很大时，就可以得出：患肺癌与吸烟有关.

抽象概括

设 A,B 为两个变量，每一个变量都可以取两个值，

变量 A：A_1，$A_2=\overline{A_1}$；

变量 B：B_1，$B_2=\overline{B_1}$.

通过观察得到表 3-10 所示数据：

表 3-10

A \ B	B_1	B_2	总计
A_1	a	b	$a+b$
A_2	c	d	$c+d$
总计	$a+c$	$b+d$	$n=a+b+c+d$

其中，a 表示变量 A 取 A_1，且变量 B 取 B_1 时的数据；b 表示变量 A 取 A_1，且变量 B 取 B_2 时的数据；c 表示变量 A 取 A_2，且变量 B 取 B_1 时的数据；d 表示变量 A 取 A_2，且变量 B 取 B_2 时的数据.

设 $n=a+b+c+d$，用 $\frac{a}{n}$ 估计 $P(A_1B_1)$，$\frac{a+b}{n}$ 估计 $P(A_1)$，$\frac{a+c}{n}$ 估计 $P(B_1)$.

若有式子

$$\frac{a}{n}=\frac{a+b}{n}\cdot\frac{a+c}{n},$$

则可以认为 A_1 与 B_1 独立.

同理，若 $\frac{b}{n}=\frac{a+b}{n}\cdot\frac{b+d}{n}$，则可以认为 A_1 与 B_2 独立；若 $\frac{c}{n}=\frac{c+d}{n}\cdot\frac{a+c}{n}$，则可以认为 A_2 与 B_1 独立；若 $\frac{d}{n}=\frac{c+d}{n}\cdot\frac{b+d}{n}$，则可以认为 A_2 与 B_2 独立.

但是，在 $\frac{a}{n}=\frac{a+b}{n}\cdot\frac{a+c}{n}$ 中，由于 $\frac{a}{n}$，$\frac{a+b}{n}$，$\frac{a+c}{n}$ 表示的是频率，不同于概率. 即使变量之间独立，式子两边也不一定恰好相等. 但是当两边相差很大时，变量之间就不独立.

练　习

为了调查吸烟是否对患慢性支气管炎有影响，某机构随机调查了 5 896 人，得到如下的数据（单位：人）：

吸烟情况 \ 患慢性支气管炎情况	患慢性支气管炎	未患慢性支气管炎
吸烟	54	1 896
不吸烟	28	3 918

请根据上面的数据分析吸烟是否对患慢性支气管炎有影响.

2.2　独立性检验的基本思想

在上一节研究吸烟是否对患肺癌有影响的问题中，我们表明了当 $\left|\frac{a}{n}-\frac{a+b}{n}\cdot\frac{a+c}{n}\right|$ 大时，变量之间不独立．同理，我们知道当 $\left|\frac{b}{n}-\frac{a+b}{n}\cdot\frac{b+d}{n}\right|$，$\left|\frac{c}{n}-\frac{c+d}{n}\cdot\frac{a+c}{n}\right|$，$\left|\frac{d}{n}-\frac{c+d}{n}\cdot\frac{b+d}{n}\right|$ 大时，变量之间也不独立．但这些量究竟要多大才能说明变量之间不独立呢？我们能不能选择一个量，用它的大小来检验变量之间是否独立呢？

统计学家选取以下统计量，用它的大小来检验变量之间是否独立：

$$\chi^2=n\left[\frac{\left(\frac{a}{n}-\frac{a+b}{n}\cdot\frac{a+c}{n}\right)^2}{\frac{a+b}{n}\cdot\frac{a+c}{n}}+\frac{\left(\frac{b}{n}-\frac{a+b}{n}\cdot\frac{b+d}{n}\right)^2}{\frac{a+b}{n}\cdot\frac{b+d}{n}}+\frac{\left(\frac{c}{n}-\frac{c+d}{n}\cdot\frac{a+c}{n}\right)^2}{\frac{c+d}{n}\cdot\frac{a+c}{n}}+\frac{\left(\frac{d}{n}-\frac{c+d}{n}\cdot\frac{b+d}{n}\right)^2}{\frac{c+d}{n}\cdot\frac{b+d}{n}}\right].$$

当 χ^2 较大时，说明变量之间不独立．上面的式子看起来很复杂，但是经过化简可以得到：

$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}. \qquad (*)$$

当数据量较大时，在统计中，用以下结果对变量的独立性进行判断.

（1）当 $\chi^2\leqslant 2.706$ 时，没有充分的证据判定变量 A,B 有关联，可

以认为变量 A,B 是没有关联的；

(2) 当 $\chi^2>2.706$ 时，有 90%的把握判定变量 A,B 有关联；

(3) 当 $\chi^2>3.841$ 时，有 95%的把握判定变量 A,B 有关联；

(4) 当 $\chi^2>6.635$ 时，有 99%的把握判定变量 A,B 有关联．

对于吸烟与患肺癌的问题，我们计算

$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$=\frac{6\ 578\times(56\times4\ 567-1\ 932\times23)^2}{1\ 988\times4\ 590\times79\times6\ 499}\approx62.698.$$

因为 $62.698>6.635$，所以有 99%以上的把握认为吸烟与患肺癌是有关的．

练　习

为了了解高中生是否喜欢参加体育锻炼和性别之间的关系，调查者随机调查了 500 名高中生的情况，调查结果如下(单位：人)：

性别 \ 参加体育锻炼情况	喜欢参加体育锻炼	不喜欢参加体育锻炼
男	197	48
女	135	120

试问：高中生是否喜欢参加体育锻炼和性别之间有关系吗？

2.3　独立性检验的应用

例 1　某组织对男女青年是否喜爱古典音乐进行了一个调查，调查者随机调查了 146 名青年，表 3-11 给出了调查的结果(单位：人)：

表 3-11

青年 \ 喜爱古典音乐情况	喜爱	不喜爱
男青年	46	30
女青年	20	50

试问：男女青年喜爱古典音乐的程度是否有差异？

解　问题是判断喜爱古典音乐是否与青年的性别有关．根据表 3-11 中的数据计算得到表 3-12(单位：人)：

表 3-12

喜爱古典音乐情况 / 青 年	喜爱	不喜爱	总计
男青年	46	30	76
女青年	20	50	70
总计	66	80	$n=146$

由(*)计算得:$\chi^2=\dfrac{146\times(46\times50-30\times20)^2}{76\times70\times66\times80}\approx15.021$.

因为 15.021>6.635,所以有 99%以上的把握认为是否喜爱古典音乐与青年的性别有关.

例 2 容易生气的人更有可能患心脏病吗?某机构随机调查了 2 796人,表 3-13 给出了调查的结果(单位:人):

表 3-13

患心脏病情况 / 是否易怒	患心脏病	未患心脏病
易怒	27	606
不易怒	53	2 110

试问:容易生气的人是否更有可能患心脏病?

解 问题是要判断患心脏病是否与易怒有关.根据表 3-13 中的数据计算得到表 3-14(单位:人):

表 3-14

患心脏病情况 / 是否易怒	患心脏病	未患心脏病	总计
易怒	27	606	633
不易怒	53	2 110	2 163
总计	80	2 716	$n=2\ 796$

由(*)计算得:$\chi^2=\dfrac{2\ 796\times(27\times2\ 110-606\times53)^2}{633\times2\ 163\times80\times2\ 716}\approx5.805$.

因为 5.805>3.841,所以有 95%以上的把握认为患心脏病与易怒有关.

例 3 生物学上对于人类头发的颜色与眼睛虹膜的颜色是否有关进行了调研,以下是一次调查结果,调查人数共 212 人.调查记录如表 3-15(单位:人):

表 3-15

眼睛虹膜颜色 / 头发颜色	蓝色	棕色
红/金黄色	156	12
黑色	20	24

试问：头发的颜色与眼睛虹膜的颜色有关吗？

解　问题是要判断头发的颜色是否与眼睛虹膜的颜色有关. 根据表 3-15 中的数据计算得到表 3-16(单位：人)：

表 3-16

眼睛虹膜颜色 / 头发颜色	蓝色	棕色	总计
红/金黄色	156	12	168
黑色	20	24	44
总计	176	36	$n=212$

由(*)计算得：$\chi^2=\dfrac{212\times(156\times24-12\times20)^2}{168\times44\times176\times36}\approx55.576$.

因为 $55.576>6.635$，所以有 99%以上的把握认为头发的颜色与眼睛虹膜的颜色有关.

练　习

某县有甲、乙两所规范化学校，教育主管部门为了检验两校初中三年级学生的数学水平，从甲、乙两校的初三学生中，分别随机抽取 55 人和 45 人(各占全校初三学生总数的 15%)，进行统一试题的数学测验. 测验结果如下表所示(单位：人)：

及格情况 / 学校	及格	不及格
甲校	47	8
乙校	30	15

试问：甲、乙两校初三学生的数学成绩的差异是否显著？

习 题 3—2

1. 请用 χ^2 检验解决 §2.1 练习(90 页)中的问题.

2. 为了考查研制出的新药对预防某种疾病的效果,科学家进行了试验,得到如下结果(单位:人):

患病情况 / 服用新药情况	患病	未患病
服用新药	12	58
未服用新药	22	28

问:新药对预防此种疾病是否有效?

3. 下面是对智商在 40～69 之间的人的出生季节所作的一个调查,结果如下(单位:人):

智商 / 季节	40～54	55～69
夏和秋	30	48
春和冬	40	30

问:智商在 40～69 之间的人其智商与出生季节有关吗?

4. 为了了解高中生数学考试成绩是否和吃早点有关,调查者随机调查了 50 名高中生的情况,调查结果如下(单位:人):

考试成绩 / 吃早点情况	及格	不及格
吃早点	17	10
不吃早点	15	8

试问:高中生的数学考试成绩是否和吃早点有关?

5. 下表是老一代和年青一代对某影片的评价的调查,所得数据如表所示(单位:人):

评价 / 年代	评价高	评价一般
老一代	45	60
年青一代	36	51

试问:老一代和年青一代对影片的评价是否一致?

6. 选择生活中的一个大家比较关心的问题(如男女同学对数学课的喜欢程度),利用独立性检验进行分析,并写出一份简明的分析报告.

统计活动

学习成绩与视力之间的关系

问题提出

在学习生活中，我们或许都有过这样的疑问：学习成绩和视力之间是否存在一定的线性相关关系？请同学们对这个问题设计一个调查方案并开展统计活动.

实践活动

我们可以按照如下的步骤来进行这个统计活动.

1. 确定调查对象

全班所有同学.

2. 收集数据

每位同学分别记录近期自己左右眼的视力情况，求出平均值，并记录上次期末考试的总分，得到下表：

姓名	左眼视力	右眼视力	左右眼视力平均值	上次期末考试总分

3. 整理数据

（1）先将本小组成员收集到的数据按下表汇总：

第____小组

	左眼视力	右眼视力	左右眼视力平均值	上次期末考试总分
小组成员 1				
小组成员 2				
⋮	⋮	⋮	⋮	⋮
小组成员 n				

（2）再把班上所有同学的数据按照小组进行汇总，得到下页表：

	左眼视力	右眼视力	左右眼视力平均值	上次期末考试总分
第 1 小组				
	⋮	⋮	⋮	⋮
第 2 小组				
	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
第 m 小组				
	⋮	⋮	⋮	⋮

4. 分析数据

(1) 画出左右眼视力平均值和上次期末考试总分的散点图，你能从散点图中大致判断出视力与学习成绩之间的线性相关关系吗？

(2) 计算视力与学习成绩之间的线性相关系数 r，并与散点图的结果进行比较.

5. 作出推断

从上面的数据分析，你能得到什么结论？它与你在从事这个统计活动之前的猜想一致吗？

思考交流

1. 在这个统计活动的过程中，调查的问题和目的分别是什么？你经历了哪些主要的步骤？

2. 根据调查的问题和目的，应当如何确定调查的对象？

3. 你认为如何收集数据比较方便？

4. 你将如何对收集到的数据快捷并且正确地进行整理？

5. 从统计数据中，要想得到一个比较好的结论，应当对数据做哪些分析？

附：

北京市延庆一中高二(2)班学生视力情况及中考成绩统计表

姓名	左眼视力	右眼视力	中考成绩
奚晶晶	4.0	4.0	596
李　佳	4.5	4.0	576
刘　莹	4.5	4.3	614
赵黎明	4.3	4.2	599
智　华	4.1	4.1	603
时俊明	4.0	4.0	597
张鹏冀	4.4	4.4	614
赵婀娜	4.3	4.3	584
高艳方	4.3	4.4	565
张　蕾	4.1	4.6	588
郎　琪	5.0	4.1	600
张博辉	4.3	4.3	592
李　雪	4.1	4.1	594
王立辉	4.3	4.1	596
张志锋	4.2	4.5	559
卢艳芬	4.6	4.5	590
梁　策	4.8	4.3	576
谢志鹏	4.1	4.1	590
陈　颖	4.4	4.6	605
周　玮	4.5	4.1	600
吴雅婧	5.1	5.0	602
张立东	5.0	4.9	582
高慧娟	4.5	4.5	573
李　妮	5.1	5.1	562
刘　昕	5.0	5.0	574
刘贵宾	5.1	5.0	587

续表

姓名	左眼视力	右眼视力	中考成绩
郑晓宇	5.2	5.2	604
蔡淑珍	4.0	4.1	563
张小文	4.4	4.6	598
张文龙	5.1	5.2	573
李　岩	5.1	5.1	589
赵文龙	5.1	5.0	561
吕　楠	5.2	5.2	582
乔立国	5.1	5.0	601
鲁俊杰	5.0	4.7	567
常洞赜	4.3	4.3	614
高宇峰	5.2	5.2	588

（数据来源：张鹏翼. 视力情况统计分析. 见：中学生研究性学习案例——中学数学建模论文选编（一）. 长春：东北师范大学出版社，2003）

◆ 本章小结建议

一、学习要求

1. 通过对典型案例的探究，进一步了解回归分析的基本思想、方法及初步应用.

2. 通过对典型案例的探究，了解独立性检验（只要求 2×2 列联表）的基本思想、方法及初步应用.

二、复习本章知识，就以下问题思考、归纳、总结，写出复习小结报告

1. 你可以通过什么方法来刻画变量之间的线性相关程度？

2. 当两个变量之间呈现出非线性的相关性时，你如何得到它们之间的回归方程？

3. 通过具体实例说明独立性检验的基本思想和主要过程.

4. χ^2 统计量在 2×2 列联表的独立性检验中起了什么作用？

复 习 题 三

1. 家族中兄弟或姐妹的智商是否有相关性一直是教育工作者、社会学家、生理学家关注的一个问题，日本学者 Ogasawara 在 1989 年曾对 45 对兄弟的智商进行测试，得出下表的结果，其中，x 表示兄，y 表示弟.

x	78	77	112	114	104	99	92	80	113
y	114	68	116	123	107	81	76	90	91
x	99	97	80	84	89	100	111	75	94
y	95	106	99	82	77	81	111	80	98
x	67	46	106	99	102	127	113	91	91
y	82	56	117	98	89	113	112	103	93
x	96	100	97	82	43	77	109	99	99
y	90	102	104	92	43	100	90	100	103
x	100	56	56	67	71	66	78	95	38
y	103	67	67	67	66	63	76	86	64

(1) 请画出散点图，并求 y 与 x 间的线性相关系数；

(2) 建立 y 关于 x 的线性回归方程，并预测当 x 为 110 时 y 的值.

2. 炼钢厂所用的盛钢水的钢包，由于钢水对耐火材料的侵蚀，容积会不断增大. 我们希望找出使用次数 x 与增大的容积 y 之间的关系，试验数据见下表.

x	2	3	4	5	6	7	8	9
y	6.42	8.20	9.58	9.50	9.70	10.00	9.93	9.99
x	10	11	12	13	14	15	16	
y	10.49	10.59	10.60	10.80	10.60	10.90	10.76	

试求 y 关于 x 的回归方程，并预测当使用次数为 20 时增大的容积量.

3. 为了了解对生活满意程度与婚姻状况的关系，调查者随机对 150 人进行了抽查，结果如下(单位：人)：

生活满意程度 / 婚姻状况	满意	不满意
已婚	82	38
丧偶	11	19

问：对生活满意程度是否与婚姻状况有关？

4. 一个调查机构向某大学的毕业生发放调查表，下面是回收情况(单位：人)：

寄回情况 / 学位	寄回	未寄回
学士	78	11
博士和硕士	61	13

问：调查表的寄回与否与学位高低有关吗？

附录 1

模拟“投掷一枚均匀的硬币 100 次”试验的程序

C 语言程序

```
/* 和*/之间的文字为注释.
#include <stdio.h>
#include <stdlib.h>
main(){
int N,i,j,z,a,b,c;
double randf;
N=100;                          /*100 次试验*/
a=0;                            /*a 记录 100 次试验中正面朝上的试验次数*/
srand((unsigned)time(NULL));    /*以当前时间作为随机函数的种子*/
for(i=0;i<N;i++)
{
z=0;                            /*z 记录两次投掷中正面朝上的次数*/
for(j=1;j<=2;j++)
{
randf=rand()/(double)RAND_MAX;  /*0~1 之间的随机数*/
if(randf<0.5) z++;
}
if(z==0) a++;
else if(z==1) b++;
else c++;
}
printf("%d,%d.%d",a,b,c);       /*输出结果*/
}
```


Maple 语言程序

```
"#" 后面的文字为说明.
>N:=50000:                      #模拟总次数
die1:=rand(1..6):               #模拟掷一颗骰子，得到 1 和 6 之间的随机整
                                  数值
die2:=rand(1..6):               #有两颗骰子需要模拟
n:=0:                           #记录有多少次模拟出现至少一对 6 点
for i from 1 to N do            #循环模拟 N 次
m:=0:                           #记录一次模拟的 24 次投掷中，出现一对 6 点的
```

```
                                        次数
for j from 1 to 24 do                   #同时投掷两颗骰子 24 次
  if die1()=6 and die2()=6              #判断是否出现一对 6 点
    then m:=m+1;                        #如果出现一对 6 点，m 增 1
  end if;                               #if 循环结束标志
end do;                                 #(24 次投掷)for 循环结束
if m>0
  then n:=n+1                           #如果出现一对 6 点的次数非 0，n 增 1
end if;
end do:                                 #(N 次模拟)for 循环结束
n;                                      #返回出现一对 6 点至少一次的模拟数
evalf(n/N);                             #及其在总模拟数中的比例
```

第二个 for 循环可以替换为 while 循环，因为如果确有一对 6 点出现了，则不必完成 24 次试验. 也就是说可以将上述代码的第 7 行到第 11 行替换成如下代码：

```
count:=0;
while (m=0 and count<=24) do
  if die1()=6 and die2()=6
    then m:=m+1;
  end if;
count:=count+1;
end do;
```

附录 2

部分数学专业词汇中英文对照表

中文	英文
组合	combination
排列	permutation
加法原理	principle of addition
乘法原理	principle of multiplication
组合公式	combination formula
二项式定理	binomial theorem
回归分析	regression analysis
随机变量	random variable
离散随机变量	discrete random variable
二项分布	binomial distribution
均值(数学期望)	mean(mathematical expectation)
方差	variance
正态分布	normal distributions

附录 3

信息检索网址导引

基础教育教材网

http://www.100875.com.cn/

简介：基础教育教材网是由北京师范大学出版社创建的一个综合性网站，内容主要涉及新课程标准改革研究、课题研究、教学研究、评价研究和教学资源等几个方面. 网站在提供教学实例、教学课件的同时，也给教师和学生提供了交流互动的宽松平台.

后　记

本套教材是按照教育部于2003年4月颁布的《普通高中数学课程标准(实验)》编写的.我们在编写过程中强调了数学课程的基础性和整体性,突出了数学的思想性和应用性,尊重学生的认知特点,创造多层次的学习活动,为不同的学生提供不同的发展平台,注意发挥数学的人文教育价值,好学好用.

教材的建设是长期、艰苦的任务,每一位教师在教学实践中要自主地开发资源,创造性地使用教材.我们殷切希望教材的使用者与我们携手合作,对教材的逐步完善提供有力的支持,促进基础教育课程改革的深入发展。

本套教材的编委会组成如下(按姓氏笔画排序):

王希平、王尚志、王建波、任志瑜、刘美仑、吕世虎、吕建生、李亚玲、李延林、汪香志、严士健、张丹、张饴慈、张思明、姚芳、赵大悌、徐勇、戴佳珉.

参加本册教材编写的还有(按姓氏笔画排序):

马芳华、白雪峰、刘卫锋、欧阳顺湘、赵冬歌、赵青.

由于时间仓促,教材中的错误在所难免,恳请广大使用者批评指正.

北京师范大学出版社