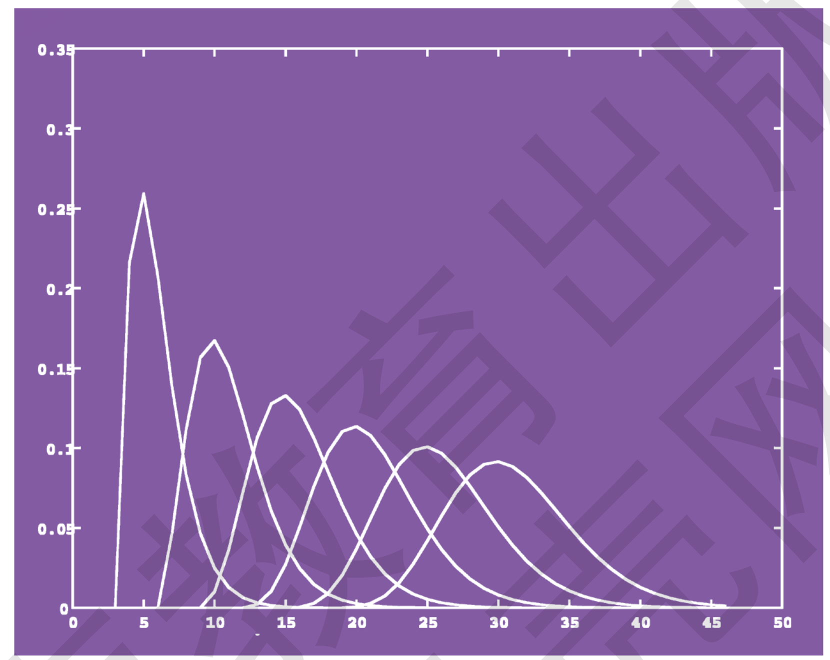
经全国中小学教材审定委员会 2005 年初审通过

Mathematics

普通高中课程标准实验教科书

数学

选修 2-3（理科）



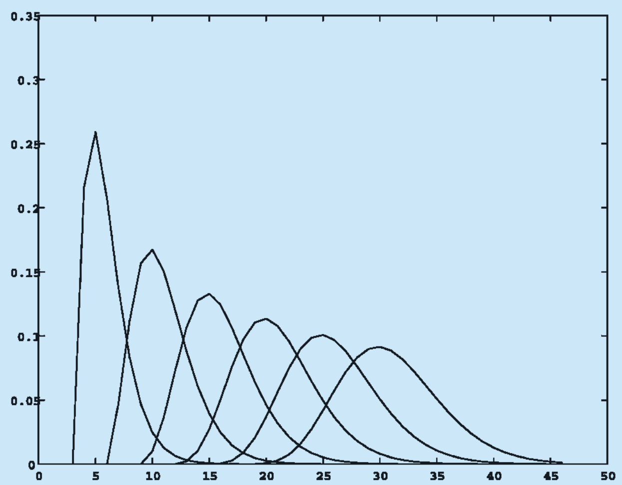
湖南教育出版社

普通高中课程标准实验教科书

数学

选修 2-3（理科）

湖南教育出版社



中国环境标志

绿色印刷产品

ISBN 978-7-5355-4612-8

9 787535 546128

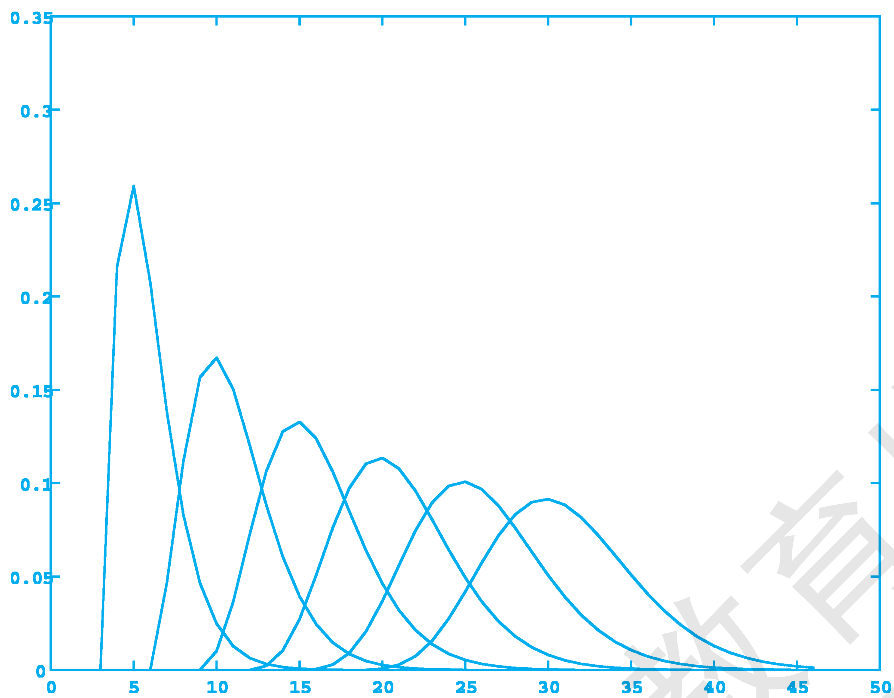
定价：6.65 元

Mathematics

普通高中课程标准实验教科书

数学

选修 2–3（理科）



湖南教育出版社

主　　编　张景中　黄楚芳
执行主编　李尚志

本册主编　何书元
编　　委　郑志明　查建国　文志英
　　　　　袁宏喜

普通高中课程标准实验教科书
数　　学
选修 2－3（理科）
责任编辑：邹楚林
责任校对：刘　源
湖南教育出版社出版发行（长沙市韶山北路 443 号）
电子邮箱：hnjycbs@sina.com
客　　服：电话 0731－85486979
湖南出版中心重印
广西区新华书店经销
湖南天闻新华印务邵阳有限公司印刷
890×1240　16 开　印张：7.25　字数：170000
2005 年 8 月第 1 版　2019 年 7 月第 2 版第 17 次印刷
ISBN 978—7—5355—4612—8
定　价：6.65 元

让数据说话

在人们的日常生活和工作中，经常遇到诸如“肺癌与吸烟有关吗”，“植物学家是如何对植物分类的”，“昆虫学家是如何给昆虫分类的”等等问题．解决这些问题的数学基础是和计数方法、统计与概率有关的．

在计数问题中，分类加法计数原理和分步乘法计数原理是解决计数问题的最基本、最重要的原理，也称为基本计数原理，它为许多实际问题的解决提供了思想方法和工具．通过学习基本计数原理，我们将逐步展开排列数、组合数和二项式定理及其应用的学习，从而了解计数方法和现实生活的联系．

现代社会已经成为信息化的社会，人们常常需要收集和整理大量的数据，统计就是研究如何合理收集、整理和分析数据，如何从数据中提取有用信息的科学，它可以为人们制定相关决策提供依据．

不像物理、化学、医学、社会学或心理学，它们都有自己确定的研究对象．统计学没有自己的基于试验的专门研究对象，但是可以为物理学家、化学家、医生、社会学家、心理学家等提供一套研究它们的问题的有效方法．这套方法可以帮助各个领域的研究工作者更快地走上成功之路．

统计学的英文名字是 Statistics．Statistics 是一个多义词，300 多年前首次应用时，指政府部门记录人们出生的日期等．时至今日，统计已经是世界上各个

层次的政府机构的重要技术支柱之一.

数据中含有许多重要的信息，用正确的统计方法从数据中提取的信息可以帮助人们制定更加合理的决策和行为规则，减少决策的盲目性和有偏性. 这就是人们经常说的一句话：让数据说话.

随机现象在日常生活中随处可见，统计与概率就是研究随机现象规律的科学，它为人们认识客观世界提供了重要的思维方式和解决问题的方法. 时至今日，统计与概率的基础知识已经成为一个未来公民必备的知识.

作　者

2008 年 5 月

目 录

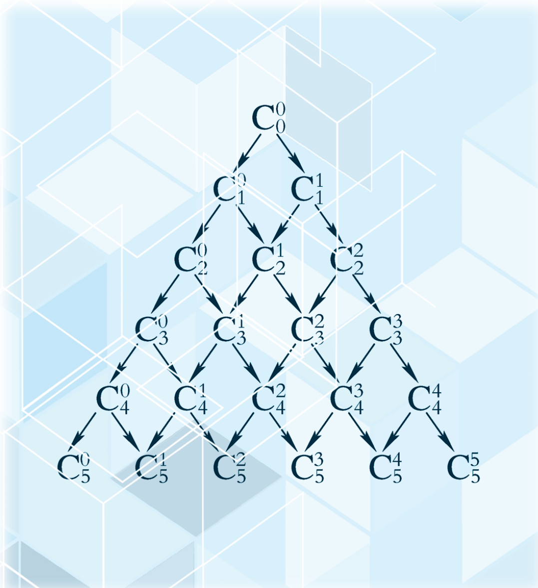
第 7 章 计数原理

第 8 章　统计与概率

第 7 章 计数原理

设赌摸球骗局深，迷图八阵费搜寻.

神机妙算杨辉数，组合排列有乾坤.



计数问题大量存在于我们的学习和日常生活中．分类加法计数原理和分步乘法计数原理是两个最基本的计数原理，排列数和组合数就是这两个计数原理的直接应用．在所有的计数问题中，排列数和组合数共同扮演了十分重要的角色.

问题探索

运气还是欺骗？（一）

王蒙先生是我国当代著名作家，他在《王蒙自述：我的人生哲学》中谈到在海滨旅游城市北戴河遇到的一件事情：一个经营游戏的人在袋中放入 20 个质地完全相同的球，这 20 个球分 4 种颜色，每种颜色的球各 5 个，参加游戏者从中随机摸出 10 个球.

为了叙述简单，我们用字母 $abcd$ 表示摸出的 10 个球中 4 种颜色的球的个数分别是 a，b，c，d. 按照这种表示规则，5500 表示摸出的球只有两种颜色；5410 表示摸出的球只有 3 种颜色，其中有 5 个颜色相同，其余 5 个中有 4 个颜色相同；……

经营者规定的游戏规则如下：

1. 摸出 5500，得一等奖，奖励一台摄像机；
2. 摸出 5410，得二等奖，奖励一条进口香烟；
3. 摸出 5311，得三等奖，奖励一个玩具机器人；
4. 摸出 4033 或 4411，得四等奖，奖励一盒进口香烟；
5. 摸出 4222，得五等奖，奖励一个小海螺；
6. 摸出 1234 或 3331，交游戏费 2 元；
7. 摸出 3322，交游戏费 5 元；
8. 摸出其他结果不奖不罚.

王蒙先生冷眼旁观，发现十之八九摸出的是 3322，十之一二摸出的是 1234 或 4033（见王蒙的原文第 143 页）.

初看起来，得奖的机会多于付钱的机会，所以有很多人怀着侥幸的心理参加游戏，其结果是拱手送上 5 元或 2 元钱. 游戏经营者一天的收入颇丰.

为什么是这样的结果呢？要回答这个问题，必须把从 20 个球

中抽取 10 个球的所有不同结果计算清楚，然后再计算 5500，5410，…，3322 在其中所占的比例.

解决以上问题的方法在数学上称为计数方法. 学习完计数方法，你可能就不会参加这种有欺骗性的游戏了.

7.1 两个计数原理

分类加法计数原理与分步乘法计数原理是计数方法中的最基本原理. 学好这两个原理是非常重要的.

7.1.1 分类加法计数原理

在日常生活中，计数问题是非常普遍的. 我们先考虑如下几个简单的例子，然后从中总结出一般的规律.

问题 1 从北京到长春可乘飞机、火车和长途汽车三类交通工具，如果一天内有 4 个航班飞往长春，有 3 列火车和 5 趟长途汽车开往长春，从北京到达长春有多少种不同的选择?

分析 如图 7－1，从北京到长春有飞机、火车和长途汽车这三类交通工具可供选择，其中乘飞机有 4 种选择，乘火车有 3 种选择，乘长途汽车有 5 种选择，所以一共有 $4+3+5=12$ 种选择.

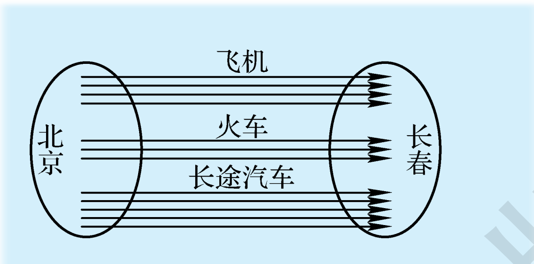


图 7－1

问题 2 书架上有语文书 8 册，数学书 8 册，英语书 7 册，小说 12 册，历史书 3 册，这些书互不相同. 从中选择一本，有多少种不同的选择方式?

分析 书架上一共有 5 类图书，分别是语文书、数学书、英语书、小说、历史书. 这 5 类书的数量分别是 8 册、8 册、7 册、12 册、3 册，一共是 $8+8+7+12+3=38$ 册. 从中选择时可以有 38 种不同的选择方式.

根据上面对问题 1 和问题 2 的分析，可以总结出下面的结论.

分类加法计数原理：如果完成一件事有 n 类办法，在第一类办法中有 m_1 种不同的方法，在第二类办法中有 m_2 种不同的方法，…，在第 n 类办法中有 m_n 种不同的方法，每种方法都能完成这件事，那么完成这件事共有

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

种不同的方法.

我们把分类加法计数原理简称为**分类计数原理**，或**加法原理**. 其特点是各类中的每一方法都可以完成要做的事情. 我们用图 7－2 表示分类计数原理. 强调每一类中的一个方法就可以完成要做的事情.

第一类有 m_1 种方法	第二类有 m_2 种方法	……	第 n 类有 m_n 种方法

一共有 $m_1+m_2+\cdots+m_n$ 种方法.

图 7－2 分类计数原理

分类时，首先要根据问题的特点确定一个适当的分类标准，然后根据这个分类标准进行分类. 分类时还要注意两条基本原则：一是完成这件事的任何一种方法必须分入相应的类；二是不同类的方法必须是不同的方法. 只要满足这两条基本原则就可以使计数不重不漏.

练 习

1. 一项工作可以用两种方法完成. 有 5 人会用第一种方法完成，另有 4 人会用第二种方法完成. 从这 9 个人中选出一个完成这件工作，共有多少种选法？
2. 在读书活动中，一个学生要从互不相同的 2 本科技书、2 本政治书、3 本文艺书里任选一本，共有多少种不同的选法？

习题 1

学而时习之

1. 一栋住宅楼共有 6 层，第一层有 8 个住户，其余每层有 12 个住户. 从中随机挑选一户进行抽样调查，会有多少种不同的挑选结果？
2. 北京的有线电视可以接收中央台 12 个频道、北京台 10 个频道和其他省市 46 个频道的节目.
   (1) 这些频道播放的节目互不相同时，一台电视机可以选看多少个节目？
   (2) 如果有 3 个频道正在转播同一场球赛，其余频道正在播放互不相同的节目，一台电视机可以选看多少个不同的节目？

7.1.2 分步乘法计数原理

问题 1 从 A 到 B 有 3 条不同的路径，从 B 到 C 有 4 条不同的路径，从 A 到 C 共有多少条不同的路径？

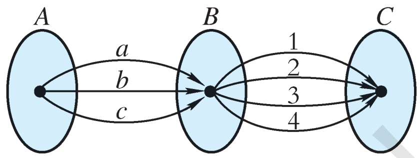


图 7－3

分析 如图 7－3，假定从 A 到 B 的三条路径分别为 a，b，c，从 B 到 C 的四条路径分别为 1，2，3，4，则从 A 到 C 的路径为

$$a1，a2，a3，a4，b1，b2，b3，b4，c1，c2，c3，c4，$$

共有 $3\times4=12$ 种.

问题 2 投掷两枚不同颜色的骰子，共有多少个不同的结果？

分析 这里我们可以把投掷第 1 枚骰子视为第一步，第一步有 6

种结果．把投掷第 2 枚骰子视为第二步，第二步有 6 种结果．第一步的每个结果都可以和第二步的 6 个结果搭配，所以一共有 $6\times6=36$ 个不同的结果．

根据上面对问题 1 和问题 2 的分析，可以总结出下面的结论．

分步乘法计数原理：如果完成一件事需要分成 n 个步骤，第一步有 m_1 种不同的方法，第二步有 m_2 种不同的方法，…，第 n 步有 m_n 种不同的方法，那么完成这件事共有

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

种不同的方法．

我们把分步乘法计数原理简称为**分步计数原理**，或**乘法原理**．其特点是每一步中都要使用一个方法才能完成要做的事情．可以用图 7－4表示分步计数原理．图中的“➡”强调要依次完成各步骤才能完成要做的事情．

第一步有 m_1 种方法	➡	第二步有 m_2 种方法	➡……➡	第 n 步有 m_n 种方法

一共有 $m_1\times m_2\times\cdots\times m_n$ 种方法．

图 7－4　分步计数原理

使用分步计数原理时，首先要根据问题的特点确定一个合理的分步标准．其原则是：如果分成 n 个步骤，那么需要而且只需要依次完成这 n 个步骤，这件事就最终完成．

分类计数原理和分步计数原理的区别在于一个与分类有关，一个与分步有关．

如果完成一件事有 n 类办法，这些办法之间是相互独立的，无论哪一类办法中的哪一种方法都能完成这件事，求完成这件事的方法数时，用分类计数原理．

如果完成一件事需要分成 n 个不可缺少的步骤，即只有依次完成所有的步骤，才能完成这件事，而完成每一个步骤都有若干不同的方

法，求完成这件事的方法数时，用分步计数原理.

例 1 书架上层放有 6 本不同的数学书，下层放有 5 本不同的语文书.

(1) 从中任取一本，有多少种不同的取法?

(2) 从中任取数学书与语文书各一本，有多少种不同的取法?

解 (1) 从书架上任取一本书，有两类办法：第一类办法是从上层取数学书，可以从 6 本书中取任一本，有 6 种方法；第二类办法是从下层取语文书，可以从 5 本书中取任一本，有 5 种方法. 因为无论哪一种方法都可以完成拿书的事情，所以用分类加法计数原理. 不同的取法有 $6+5=11$ 种.

(2) 从书架上任取数学书与语文书各一本，可以分成两个步骤完成：

第一步，取一本数学书，有 6 种方法；

第二步，取一本语文书，有 5 种方法.

这两个步骤缺一不可，根据分步乘法计数原理，得到不同的取书方法数是 $6\times5=30$.

例 2 办展览时有 6 幅国画和 4 幅油画供选择使用.

(1) 从中挑选一幅时，有多少种不同的选法?

(2) 从中各挑选一幅时，有多少种不同的选法?

解 (1) 由分类加法计数原理得，共有 $6+4=10$ 种选法.

(2) 一共两个步骤：

第一步，选国画，有 6 种选法；

第二步，选油画，有 4 种选法.

用分步乘法计数原理得到共有 $6\times4=24$ 种选法.

例 3 允许重复使用时，用数字 1，2，3，4，5 可以组成多少个三位数?

解 要组成一个三位数可以分成三个步骤完成：

第一步，确定百位上的数字，从 5 个数字中任选 1 个数字，共有 5 种选法；

第二步，确定十位上的数字，由于数字允许重复，仍有 5 种选法；

第三步，确定个位上的数字，也有 5 种选法.

根据分步乘法计数原理，得到三位数的个数是 $5\times5\times5=125$.

例 4 某农场要在 4 种不同类型的土地上，分别试验种植 A，B，C，D 四个不同品种的小麦，问有多少种不同的试验方案?

解 第一步，先考虑 A 种小麦，可在 4 种不同类型的土地中任选一种，有 4 种选法；

第二步，考虑 B 种小麦，可在剩下的 3 种不同类型的土地中任选一种，有 3 种选法；

第三步，考虑 C 种小麦，可在剩下的 2 种不同类型的土地中任选一种，有 2 种选法；

第四步，考虑 D 种小麦，这时只剩下 1 种土地，所以有 1 种选法.

以上四步依次完成后，试验方案才算完成. 依据分步乘法计数原理，可知有 $4\times3\times2\times1=24$ 种不同的试验方案.

例 5 乘积 $(x_1+x_2+x_3)(y_1+y_2+y_3)(z_1+z_2+z_3+z_4)$ 展开后共有多少项?

解 第一步，选出 x_1，x_2，x_3 中的一个，有 3 种方法；第二步，选出 y_1，y_2，y_3 中的一个，有 3 种方法；第三步，选出 z_1，z_2，z_3，z_4 中的一个，有 4 种方法. 根据分步乘法计数原理，所有不同的选法有 $3\times3\times4=36$ 个. 因为每个选法恰好对应展开式中的一项，所以展开式共有 36 项.

例 6 我国的邮政编码由 6 位数字组成，如果每个数字可以是 0，1，…，9 中的一个，最多可以编排多少个不同的邮政编码?

6 位数字不是六位数. 例如 010122 是一个 6 位数字，但不是一个六位数.

解 每个数字可以是 0，1，…，9 这 10 个数字中的一个. 我们可以从左往右依次安排出一个 6 位数字.

第一步，选取最左面的数字，有 10 种方法；

第二步，选取第 2 位数字，有 10 种方法；

……

第六步，选取第 6 位数字，有 10 种方法.

根据分步乘法计数原理，知道最多可以安排 $10\times10\times10\times10\times10\times10=10^6$ 个不同的邮政编码.

小结 在使用分类加法计数原理或分步乘法计数原理解决问题时，一定要分清完成这件事，是有 n 类办法还是有 n 个步骤. 分类要做到“不重不漏”. 分类后再分别对每一类进行计数，最后用分类加法计数原理求和，得到总数. 分步要做到“步骤完整”——完成了所有步骤，恰好完成任务，当然步与步之间要相互独立. 分步后再计算每一步的方法数，最后根据分步乘法计数原理求积，得到总数.

练习

1. 若 x，y 都可以是 1，2，3，4，5 中的任一个，则不同的点 (x,y) 有多少个?
2. 由 A 村去 B 村的道路有 3 条，由 B 村去 C 村的道路有 2 条. 从 A 村经 B 村去 C 村，共有多少种不同的走法?
3. 投掷一枚 5 角的硬币 10 次，依次记录正面或反面的出现情况，最多可以得到多少种不同的记录结果?
4. 在一副扑克的 54 张中有放回地每次抽取 1 张，一共抽取 3 次并依次排列结果，最多有多少个不同的排列结果?

习题 2

学而时习之

1. 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?
2. 有三个袋子，第一个袋子装有标号 1～20 的红色小球 20 个，第二个袋子装有标号 1～15 的白色小球 15 个，第三个袋子装有标号 1～8 的蓝色小球 8 个.

   (1) 从三个袋子里选取一个小球，有多少种不同的选法?

   (2) 从每个袋子里选取一个小球，有多少种不同的选法?
3. 从甲地到乙地有 3 条公路可走，从乙地到丙地有 5 条小路可走，又从甲地不经过乙地到丙地有 3 条水路可走.

(1) 从甲地经乙地到丙地有多少种不同的走法?

(2) 从甲地到丙地共有多少种不同的走法?

4. 罐中装有编号 $1\sim n$ 的小球 n 个，从中摸出一个，记下球号后放回. 摸球 m 次时，依次记录摸到的球号，最多得到多少种球号的排列?

5. 某省的体育彩票中，把有顺序的 7 个数字组成一个号码，称为一注. 7 个数字中的每个数字都选自 0，1，2，…，9 这 10 个数字且可以重复. 如果全体不同号码的彩票中只有一个大奖.

(1) 不同号码的彩票一共有多少注?

(2) 在不同号码的所有彩票中购买一张，计算中奖率.

6. 家住北京的李老师每周一要乘上午的火车或汽车到天津讲课一次. 如果每天上午有 6 次列车和 8 趟汽车开往天津，计算去天津三次时，一共有多少种不同的选择.

7.2 排 列

7.2.1 排列与排列数公式

问题 1 从 a，b，c，d 这 4 个字母中，取出 3 个排成一排，共有多少种不同的结果?

分析 解决这个问题，需要分 3 个步骤.

第一步，先确定左边的字母，在 4 个字母中任取 1 个，有 4 种方法;

第二步，确定中间的字母，从余下的 3 个字母中取，有 3 种方法;

第三步，确定右边的字母，只能从余下的 2 个字母中取，有 2 种方法.

根据分步乘法计数原理，共有 $4\times3\times2=24$ 种不同的排法. 它们是

$$
\begin{array}{cccccc}
abc & acb & abd & adb & acd & adc \\
bac & bca & bad & bda & bcd & bdc \\
cab & cba & cbd & cdb & cad & cda \\
dbc & dcb & dab & dba & dac & dca
\end{array}
$$

可以看出，上述排列的特点是无重复、有次序.

问题 2　某公司有 5 艘远洋货轮，现在要派遣 3 艘执行运输任务，在派遣的先后有次序时，有多少种派遣方法？

分析　第一步，从 5 艘远洋货轮中选取一艘，有 5 种方法；

第二步，从其余 4 艘远洋货轮中选取一艘，有 4 种方法；

第三步，从其余 3 艘远洋货轮中选取一艘，有 3 种方法.

根据分步乘法计数原理，知道一共有 $5\times4\times3=60$ 种派遣方法.

从对问题 1 和问题 2 的分析可以看出如下的规律.

应当记住 A_n^m 的最后一项是 $(n-m+1)$.

排列：从 n 个不同元素中取出 $m(m\leqslant n)$ 个不同的元素，按照一定的顺序排成一列，叫作从 n 个不同元素中取出 m 个元素的一个排列(permutation). 用符号 A_n^m 表示排列的个数时，有

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

证　第一步，从 n 个元素中选取一个，有 n 种方法；

第二步，从余下的 $n-1$ 个元素中选取一个，有 $n-1$ 种方法；

……

第 m 步，从余下的 $n-m+1$ 个元素中选取一个，有 $n-m+1$ 种方法.

根据分步乘法计数原理，知道一共有

$$n(n-1)(n-2)\cdot\cdots\cdot(n-m+1)$$

种方法.

为了方便地表示连乘积，对于自然数 n，我们定义

$$n!=1\times2\times3\times\cdots\times n,$$

并且称 $n!$ 是 n 的**阶乘**（factorial).

特别还规定 $0!=1$.

根据上面阶乘的定义得

$$n(n-1)(n-2)\cdot\cdots\cdot(n-m+1)=\frac{n!}{(n-m)!}.$$

于是，从 n 个不同元素中取出 $m(m\leqslant n)$ 个元素，按照一定的顺序排成一列，共有

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

个不同的排列结果.

根据排列的定义，一个排列包含两个方面的意义：一是“取出元素”；二是“按照一定顺序排列”. 因此，两个排列相同，当且仅当这两个排列的元素及其排列顺序完全相同.

换句话说，如果两个排列所含的元素不完全一样，那么肯定是不同的排列；如果两个排列所含的元素完全一样，但排列的顺序不同，也是不同的排列.

在上面定义的排列里，如果 $m<n$，表示只选一部分元素进行排列，因此又叫作**选排列**. 从 n 个不同元素中取出 $m(m<n)$ 个元素的选排列个数是 A_n^m.

如果 $m=n$，表示将全体元素进行排列，所以又叫作**全排列**. n 个不同元素的全排列个数是

$$A_n^n=n(n-1)\cdot\cdots\cdot 3\cdot 2\cdot 1=n!.$$

例 1 我国的邮政编码由 6 位数字组成，如果每个数字可以是 0，1，…，9 中的一个，最多可以编排多少个数字互不相同的邮政编码？

解 一个数字互不相同的邮政编码恰是从 0，1，2，…，9 这 10 个数字中取出的 6 个数字的一个排列. 这样的数字一共有

$$A_{10}^6=\frac{10!}{(10-6)!}=151\ 200$$

个，所以最多可以编排 151 200 个数字互不相同的邮政编码.

例 2 从 8 名同学中选 4 人参加 4×100 米接力赛，有多少种不同的参赛方案？

解 每一种参赛方案恰是从 8 个同学中选取 4 个同学的一个排列. 这样的排列数是

$$A_8^4=8\times7\times6\times5=1\ 680$$

个，所以参赛方案一共有 1 680 个.

例 3 某青年志愿者协会组织者将 n 棵树苗随机地分发给参加

"义务植树活动"的 n 名青年志愿者，会有多少种不同结果？

解 将树苗从 1 到 n 进行编号，分别记为 a_1，a_2，…，a_n.

第 1 名志愿者得到的树苗的号码是 a_1；

第 2 名志愿者得到的树苗的号码是 a_2；

……

第 n 名志愿者得到的树苗的编号是 a_n.

组织者分发树苗的每个结果恰好对应一个排列 a_1，a_2，…，a_n. 这样的排列一共有

$$n!=n\cdot(n-1)\cdot\cdots\cdot2\cdot1$$

个，所以一共会有 $n!$ 个不同的结果.

练习

1. 5 个人排成一排，共有________种不同的排法.
2. 某信号兵用红、黄、蓝三面旗从上到下挂在竖直的旗杆上表示信号，每次可以挂一面、二面或三面，并且用不同的顺序表示不同的信号，一共可以表示多少种不同的信号？

习题 3

学而时习之

1. 叙述排列的定义.
2. 写出排列数 A_n^m 的公式.
3. 写出 $n!$的计算公式.
4. 由数字 2，3，4，5，6 可以组成

   (1) 没有重复数字的五位数（　　　　）个；

   (2) 允许有重复数字的五位数（　　　　）个；

(3) 没有重复数字的自然数（　　　）个；

(4) 没有重复数字的三位数（　　　）个.

5. 判断下列问题是否是排列问题：

(1) 从 7 名同学中选派 3 人去完成 3 种不同的工作，每人完成一种，有多少种不同的选派方法；

(2) 从 7 名同学中选 3 人去某地参加一个会议，有多少种不同的选派方法.

6. 一台晚会有 6 个节目，其中有 2 个小品，如果 2 个小品不连续演出，共有不同的演出顺序多少种？

7.2.2 排列数的应用

例 1 验证排列数 A_n^m 满足

(1) $A_n^1=n$；　　　　(2) 当 $n>m>1$ 时，$A_n^m=nA_{n-1}^{m-1}$.

解 (1) 因为 $n-1+1=n$，所以 $A_n^1=n$.

$A_n^1=n$ 的解释：从 n 个不同的元素中选出一个进行排列，一共有 n 个选法.

(2) 因为 $A_{n-1}^{m-1}=(n-1)(n-2)\cdot\cdots\cdot(n-m+1)$，所以

$$A_n^m=n[(n-1)(n-2)\cdot\cdots\cdot(n-m+1)]=nA_{n-1}^{m-1}.$$

这个公式的解释：从 n 个不同的元素中选出 m 个进行排列，相当于第一步选出一个排在第 1 位，有 n 种方法；第二步在其余的 $n-1$ 个元素中选择 $m-1$ 个，依次排在第 2，第 3，…，第 m 位，有 A_{n-1}^{m-1} 种排法. 由分步乘法计算原理可知，不同的排列总数是 nA_{n-1}^{m-1} 个.

例 2 计算：

(1) A_6^4；

(2) $A_3^2+A_4^3+A_5^3$.

解 (1) 由 $6-4+1=3$ 知道

$$A_6^4=6\times5\times4\times3=360.$$

(2) 由 $3-2+1=2$，$4-3+1=2$ 和 $5-3+1=3$ 得到

$$A_3^2+A_4^3+A_5^3=3\times2+4\times3\times2+5\times4\times3=90.$$

例 3 解方程：

(1) $3A_x^3=2A_{x+1}^2+6A_x^2$；

(2) $3A_8^x=4A_9^{x-1}$.

解 (1) 因为

$$A_x^3=x(x-1)(x-2),\ A_{x+1}^2=(x+1)x,\ A_x^2=x(x-1),$$

所以由 $3A_x^3=2A_{x+1}^2+6A_x^2$，得

$$3x(x-1)(x-2)=2(x+1)x+6x(x-1).$$

从 A_x^3 有意义知道 $x\geqslant3$，故上式两边可以约去 x，得到方程

$$3(x-1)(x-2)=2(x+1)+6(x-1).$$

整理后得到

$$3x^2-17x+10=0.$$

解方程得 $x=5$ 和 $x=\frac{2}{3}$（舍去）．所以 $x=5$.

(2) 利用

$$3A_8^x=3\,\frac{8!}{(8-x)!},\ 4A_9^{x-1}=4\,\frac{9!}{(9-x+1)!},$$

得到

$$\frac{3\times8!}{(8-x)!}=\frac{4\times9!}{(10-x)!}.$$

利用 $(10-x)!=(10-x)(9-x)(8-x)!$，将上式化简后得到

$$(10-x)(9-x)=4\times3.$$

再化简得到

$$x^2-19x+78=0.$$

解方程得 $x_1=6$，$x_2=13$．由于 A_8^x 和 A_9^{x-1} 有意义，所以 x 满足 $x\leqslant8$ 和 $x-1\leqslant9$．于是将 $x_2=13$ 舍去，原方程的解是 $x=6$.

例 4 由数字 1，2，3，4，5 组成没有重复数字的五位数，其中小于 50 000 的偶数共有多少个？

解 第一步排个位数，因为要求是偶数，所以只能排 2 或 4，排法有 A_2^1 种；

第二步排万位数，小于 50 000 的五位数，万位数只能用 1，3 或

用排个位数时余下的 2，4 中的一个，排法有 A_3^1 种；

在首末两位数排定后，第三步排中间 3 个数字时，排法有 A_3^3 种.

根据分步计数原理，要求的偶数有

$$A_2^1A_3^1A_3^3=2\times3\times3\times2\times1=36\text{（个）}.$$

例 5 解答下面的问题.

（1）从 5 种不同的书（每种不少于 3 本）中买 3 本送给 3 名同学，每人各 1 本，共有多少种不同的送法？

（2）4 个读者到 4 个服务台排队还书，有且只有一个服务台没有这 4 个读者还书的排队有多少种？

解 （1）送给第一个同学有 5 种不同的选购方法，送给第二、第三个同学各 1 本书，仍各有 5 种不同的选购方法. 因此，根据分步计数原理，一共有 $5^3=125$ 种不同送法.

（2）(捆绑法)第一步，从 4 个读者中选出 2 个“捆绑”在一起，视为 1 个“读者”，因为有排队的先后，所以有 A_4^2 种方法；第二步，从 4 个服务台中取定 3 个，将以上的“3 个读者”依次排列在这 3 个服务台，共有 A_4^3 种方法. 根据分步计数原理，一共有 $A_4^2A_4^3=288$ 种方法.

在某些元素要求必须相邻时，可以先将这些元素排列，并看做一个元素，然后与其他元素排列，这种方法称为“捆绑法”.

练 习

1. 计算：（1）$4A_4^2+5A_5^3$；（2）$A_4^1+A_4^2+A_4^3+A_4^4$.
2. 0，1，2，…，9 这 10 个数字能组成多少个没有重复数字的三位数？
3. 三个男生和四个女生按下列条件排成一排，有多少种排法？

（1）男生排在一起，女生排在一起；

（2）男、女生间隔相排；

（3）男生互不相邻；

（4）甲、乙两人必须相邻.

习题 4

学而时习之

1. 计算 $A_6^1+A_6^2+A_6^3$.
2. 从 x 个不同元素中取 3 个的排列数为 720，x 是多少？
3. 已知 $A_n^n+A_{n-1}^{n-1}=xA_{n+1}^{n+1}$，求 x 的值.
4. 用0，1，2，3，4，5 可以组成多少个无重复数字的五位数？五位奇数？五位偶数？
5. 写给 8 个人的信笺随意装入 8 个写好地址的信封，会有多少不同的结果？
6. 6 个同学排成一排照相，其中甲、乙两人必须相邻的排法有多少种？

7.3 组　合

7.3.1 组合与组合数公式

考虑如下的问题.

问题 1　全年级要举行班级篮球赛，如果全年级 8 个班中的任何两个班都比赛一次，需要安排多少场比赛？

问题 2　列车从始发站到达终点站中途停车 8 站，单程需要制作多少种不同的火车票？

问题 3　汽车公司从 12 辆客车中选派 3 辆客车运送高二年级同学参加秋游，有多少种选法？

上面的所有问题都有一个特点：选出的单位如两个班、两个站、三辆客车都是和次序无关的，所以以上问题不是排列问题. 它们与排列的区别在于抽取元素时不考虑顺序. 我们称这样的问题为组合问题.

从 n 个不同的元素中取出 $m(m\leqslant n)$ 个不同的元素，不论次序地构成一组，称为一个组合(combination). 我们用符号 C_n^m 表示所有不同的组合个数，称 C_n^m 为从 n 个不同的元素中取 m 个元素的组合数.

上述定义中“取出 m 个不同的元素”的“不同”是强调取出的元素不能有重复，也就是指无放回地抽取.

例如从 8 个不同的元素中取 5 个元素的组合数是 C_8^5，从 7 个不同元素中取 5 个元素的组合数是 C_7^5，从一副扑克的 54 张中抽取 13 张的组合数是 C_{54}^{13}.

排列与组合的相同点是都是从 n 个不同元素中取 m 个元素，元素无重复；不同点是组合与顺序无关，排列与顺序有关. 两个组合相同，当且仅当这两个组合的元素完全相同.

组合数 C_n^m 还常常有下面例 1 中的表述方法.

例 1 把 n 个不同的元素分成有顺序的两组，第一组有 m 个元素，第二组有 $n-m$ 个元素，证明共有 C_n^m 种分法.

证 从 n 个不同的元素中取出 $m(m\leqslant n)$ 个不同的元素，放入第一组，得到一个组合. 每个组合恰好是一个分组. 因为组合数是 C_n^m，所以共有 C_n^m 种不同的分组方法.

例 2 先回答以下问题是组合问题还是排列问题，然后再计算所问的结果.

(1) 集合 $\{0, 1, 2, 3, 4\}$ 的含三个元素的子集的个数是多少？

(2) 用没有任何三点共线的五个点可以连成多少条线段？如果连成有向线段，共有多少条？

(3) 某小组有 9 名同学，从中选出正、副班长各一人，有多少种不同的选法？若从中选出 2 名代表参加一个会议，有多少种不同的选法？

解 (1) 由于集合中的元素是不讲次序的，一个含三个元素的集合就是一个从 $\{0, 1, 2, 3, 4\}$ 中取出 3 个数的组合. 这是一个组合问题，组合的个数是 C_5^3，所以子集的个数是 C_5^3.

(2) 由 5 个点中取两个点恰好连成一条线段，不用考虑这两个点的次序，所以是组合问题，组合数是 C_5^2，连成的线段共有 C_5^2 条．再考虑有向线段问题．这时两个点的先后排列次序对应两个不同的有向线段，所以是排列问题，排列数是 $A_5^2=5\times4=20$，所以有向线段共有 20 条．

(3) 选正、副班长时要考虑次序，所以是排列问题．排列数是 $A_9^2=9\times8=72$，所以正、副班长共有 72 种选法．选代表参加会议是不用考虑次序的，所以是组合问题．组合数是 C_9^2，所以不同的选法有 C_9^2 种．

下面从研究组合与排列的关系入手，找出组合数 C_n^m 的计算公式．

计算从 n 个不同的元素中取出 m 个元素的排列数可以按以下两步来完成：

第一步，先从这 n 个不同元素中取出 m 个元素，不考虑次序构成一个组合，共有 C_n^m 个组合；

第二步，将每一个组合中的 m 元素进行全排列，全排列数是 $A_m^m=m!$．

由于第二步得到的全排列恰好是从 n 个不同元素中取出 m 个元素的选排列，所以根据分步计数原理，得到 $A_n^m=C_n^m A_m^m$．

因此得到组合数 C_n^m 的计算公式：

$$C_n^m=\frac{A_n^m}{A_m^m}=\frac{n(n-1)\cdot\cdots\cdot(n-m+1)}{m!},\ 0\leqslant m\leqslant n.$$

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

$$C_n^m=\frac{n!}{m!(n-m)!},\quad 0\leqslant m\leqslant n$$

我们把上面的公式叫作**组合数公式**．这个公式经常被用于和组合数有关的等式证明．

在组合数公式中，我们规定 $C_n^0=1$．这样对任何正整数 n 和 m，$m=0,1,\cdots,n$，都有

$$C_n^m = C_n^{n-m}$$

上面公式的解释：从 n 个不同元素中选取 m 个元素的组合数与从 n 个不同元素中选取 $n-m$ 个元素的组合数相等.

公式 $C_n^m = C_n^{n-m}$ 的证明：因为

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

又有 $C_n^m = \dfrac{n!}{m!(n-m)!}$，所以 $C_n^m = C_n^{n-m}$ 成立.

有了上面的组合数公式，就可以计算本节一开始提出的问题了.

问题 1 的解 全年级 8 个班举行班级篮球赛，由于每两个班都比赛一次，比赛的两个班无次序问题，所以需要安排

$$C_8^2 = \frac{8\times7}{2!} = \frac{8\times7}{2\times1} = 28$$

场比赛.

问题 2 的解 列车从始发站到达终点站中途停车 8 站，一共是 10 站，每两站之间要制作车票. 只考虑单程，相当于不考虑次序. 于是单程需要制作

$$C_9^2 = \frac{9\times8}{2!} = \frac{9\times8}{2\times1} = 36$$

种火车票.

问题 3 的解 汽车公司从 12 辆客车中选派 3 辆客车运送高二年级同学秋游，不用考虑 3 辆车的次序，因而有

$$C_{12}^3 = \frac{12\times11\times10}{3!} = \frac{12\times11\times10}{3\times2\times1} = 220$$

种选法.

练 习

1. 计算 $C_9^6 + C_8^5$.
2. 全班 40 个同学互相通话一次，一共要通话多少次？

习题 5

学而时习之

1. 计算 $C_6^1+C_6^2+C_6^3$.
2. 文具盒中有 6 支不同的圆珠笔，3 支不同的铅笔. 从中取出 3 支借给同学，有多少种借法？
3. 在全班的 18 名女生中挑选 9 名参加作文比赛，有多少种选法？
4. 列举从 4 个不同元素 a，b，c，d 中取出 3 个元素的所有组合和排列.
5. （1）平面内有 10 个点，以其中 2 个点为端点的线段共有多少条？

   （2）平面内有 10 个点，以其中 2 个点为端点的有向线段共有多少条？

7.3.2 组合数的性质和应用

组合数的性质 1 如果 $C_n^m=C_n^k$，则

$m=k$ 或者 $m=n-k$.

例 1 解方程 $C_{10}^x=C_{10}^{3x-2}$.

解 利用性质 1 得到

$$x=3x-2 \text{ 或 } x=10-(3x-2).$$

解上述方程得到 $x=1$ 或 $x=3$. 这两个解都符合题意.

组合数的性质 2

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

证 利用

$$m(m-1)!=m!,$$

$$(n-m+1)(n-m)!=(n-m+1)!$$

得到

$$
\begin{aligned}
& \mathrm{C}_n^m+\mathrm{C}_n^{m-1} \\
= & \frac{n!}{m!(n-m)!}+\frac{n!}{(m-1)![n-(m-1)]!} \\
= & \frac{n!(n-m+1)}{m!(n-m+1)!}+\frac{mn!}{m!(n-m+1)!} \\
= & \frac{(n-m+1)n!+mn!}{m!(n-m+1)!} \\
= & \frac{(n+1)!}{m!(n-m+1)!} \\
= & \mathrm{C}_{n+1}^m .
\end{aligned}
$$

所以

$$\mathrm{C}_{n+1}^m=\mathrm{C}_n^m+\mathrm{C}_n^{m-1}.$$

注意上面公式的特征：等式右边组合数的下标都是 n，上标的差是 1.

我们也可以从组合的定义出发证明上述等式．等式左边的 C_{n+1}^m 是从 $n+1$ 个不同的元素中取 m 个元素的组合数．等式右边可以看成是相同问题的另一种方法：把上述的组合分成两类，一类含有元素 a，一类不含元素 a．不含元素 a 的组合共有 C_n^m 个，含有元素 a 的组合共有 C_n^{m-1} 个．由分类计数原理得到 $\mathrm{C}_{n+1}^m=\mathrm{C}_n^m+\mathrm{C}_n^{m-1}$.

例 2 计算 $\mathrm{C}_6^3+\mathrm{C}_6^4+\mathrm{C}_7^5+\mathrm{C}_8^6$.

解 利用性质 2 得到

$$\mathrm{C}_6^3+\mathrm{C}_6^4=\mathrm{C}_7^4,$$
$$\mathrm{C}_7^4+\mathrm{C}_7^5=\mathrm{C}_8^5,$$
$$\mathrm{C}_8^5+\mathrm{C}_8^6=\mathrm{C}_9^6.$$

于是

$$
\begin{aligned}
& \mathrm{C}_6^3+\mathrm{C}_6^4+\mathrm{C}_7^5+\mathrm{C}_8^6 \\
= & \mathrm{C}_7^4+\mathrm{C}_7^5+\mathrm{C}_8^6 \\
= & \mathrm{C}_8^5+\mathrm{C}_8^6 \\
= & \mathrm{C}_9^6 \\
= & \frac{9\times8\times7}{3\times2\times1}=84.
\end{aligned}
$$

例 3 12 件产品中有 3 件次品，9 件正品，从中抽取 5 件.

(1) 5 件产品中没有次品的取法有多少种?

(2) 5 件产品中有 2 件次品的取法有多少种?

解 (1) 5 件产品中没有次品的取法就是从 9 件正品中取 5 件的取法，有 $\mathrm{C}_9^5=126$ 种.

（2）第一步，先从 3 件次品中取 2 件，有 C_3^2 种取法；第二步，从 9 件正品中取 3 件，有 C_9^3 种取法. 利用分步计数原理，知道一共有

$$C_3^2C_9^3=252$$

种取法.

例 4 从 4 台纯平彩电和 5 台超平彩电中选购 3 台，要求至少有纯平彩电与超平彩电各 1 台，问有多少种不同的选法？

解 完成满足条件的工作有两类方法：第一类是纯平彩电中选一台，超平彩电中选两台；第二类是纯平彩电中选两台，超平彩电中选一台. 这两类工作完成一类即可.

从 4 台纯平彩电中选出 1 台，有 C_4^1 种选法；再在 5 台超平彩电中选出 2 台，有 C_5^2 中选法. 于是选出的 3 台彩电中纯平彩电 1 台、超平彩电 2 台的选法有 $C_4^1C_5^2$ 种.

同样计算选出纯平彩电 2 台、超平彩电 1 台的选法有 $C_4^2C_5^1$ 种.

用分类计数原理得到选法总数是

$$\begin{aligned}&C_4^1C_5^2+C_4^2C_5^1\\&=\frac{4!}{1!3!}\cdot\frac{5!}{2!3!}+\frac{4!}{2!2!}\cdot\frac{5!}{1!4!}\\&=40+30=70\ (\text{种}).\end{aligned}$$

例 5 6 本不同的书，按下列要求各有多少种不同的分法：

（1）分给甲、乙、丙三人，每人 2 本；

（2）分为三份，每份 2 本；

（3）分为三份，一份 1 本，一份 2 本，一份 3 本；

（4）分给甲、乙、丙三人，一人 1 本，一人 2 本，一人 3 本.

解 （1）先从 6 本书中选 2 本给甲，有 C_6^2 种选法；

再从其余的 4 本中选 2 本给乙，有 C_4^2 种选法；

最后从余下的 2 本书中选 2 本给丙，有 C_2^2 种选法.

根据分步计数原理得到一共是

$$C_6^2C_4^2C_2^2=(5\times3)\times(2\times3)\times1=90$$

种分法.

所以分给甲、乙、丙三人，每人 2 本有 $C_6^2C_4^2C_2^2=90$ 种方法.

（2）这个过程可以分两步完成：

第一步，将 6 本书分为三份，每份 2 本，设有 x 种方法；

第二步，将上面三份分给甲、乙、丙三名同学有 A_3^3种方法.

根据（1）的结论和分步计数原理得到 $C_6^2C_4^2C_2^2=xA_3^3$，所以

$$x=\frac{C_6^2C_4^2C_2^2}{A_3^3}=15.$$

因此分为三份，每份 2 本一共有 15 种方法. 本题称为“均匀分组”问题.

（3）这是“不均匀分组”问题，按照（1）的方法得到一共有

$$C_6^1C_5^2C_3^3=6\times(5\times2)\times1=60$$

种方法.

（4）在（3）的基础上再进行全排列，所以一共有 $C_6^1C_5^2C_3^3\times A_3^3=360$ 种方法.

例 6 某省的福利彩票中，不考虑次序的 7 个数码组成一注，7 个数码中没有重复，每一个数码都选自数码 1，2，…，36. 如果电视直播公开摇奖时只有一个大奖，计算：

（1）公开摇奖时最多可以摇出多少不同的注；

（2）购买一注时的中奖率.

解 （1）摇奖时是从数码 1，2，…，36 中无重复地抽取 7 个数码，不计次序时，所有不同的结果有

$$C_{36}^7=8\ 347\ 680$$

个，于是可以摇出 8 347 680 个不同的注.

（2）购买一注的中奖率是

$$\frac{1}{8\ 347\ 680}\approx0.000\ 000\ 1.$$

练 习

1. 利用 $C_{n+1}^m=C_n^m+C_n^{m-1}$，计算 $C_7^3+C_7^4+C_8^5+C_9^6$.
2. 凸六边形有多少条对角线？

习题 6

学而时习之

1. 机场有 10 架飞机，要调用 7 架去执行任务，有多少种调法？

2. 机场有 10 架飞机，要调用 7 架排成一列，有多少种排法？

3. 分派 5 个同学中的 3 人擦教室玻璃，2 人扫地，有多少种分派方法？

4. 将全班 18 名男生分在两个不同的兴趣班上课，每个班 9 人，有多少种分法？

5. 10 件产品中有合格品 8 件，次品 2 件，从中抽取 4 件，计算：

   (1) 都不是次品的取法有多少种？

   (2) 至少有 1 件次品的取法有多少种？

6. 利用组合数的性质 2，计算：

   (1) $C_m^5-C_{m+1}^5+C_m^4$；

   (2) $C_{97}^{94}+C_{97}^{95}+C_{98}^{96}+C_{99}^{97}$.

7. 解方程 $C_{13}^{x+1}=C_{13}^{2x-3}$.

8. 凸 n 边形有多少条对角线？

问题探索

运气还是欺骗？（二）

下面我们解决本章开始时，王蒙先生所述的北戴河的游戏问题.

问题回顾 袋中装有 20 个质地完全相同的球，这 20 个球分 4 种颜色，每种颜色的球各 5 个. 参加游戏者从中随机摸出 10 个球.

为了叙述的简单，我们用字母 $abcd$ 表示摸出的 10 个球中 4 种颜色的球的个数分别是 a，b，c，d. 按照这种表示，5500 表示摸出的球只有两种颜色；5410 表示摸出的球只有 3 种颜色，其中有 5 个颜色相同，其余 5 个中有 4 个颜色相同；……

计算从 20 个球中摸出 10 个球的组合数，再计算以下各种结果出现的数目和各种事件发生的概率.

1. 摸出 5500，得一等奖，奖励一台摄像机；
2. 摸出 5410，得二等奖，奖励一条进口香烟；
3. 摸出 5311，得三等奖，奖励一个玩具机器人；
4. 摸出 4033 或 4411，得四等奖，奖励一盒进口香烟；
5. 摸出 4222，得五等奖，奖励一个小海螺；
6. 摸出 1234 或 3331，交游戏费 2 元；
7. 摸出 3322，交游戏费 5 元；
8. 摸出其他结果不奖不罚.

下面解决这个问题.

从 20 个球中抽取 10 个球，不计次序的组合数是

$$\mathrm{C}_{20}^{10}=\frac{20!}{10!10!}=184\ 756.$$

这也是从 20 个球中抽取 10 个球的等可能结果的总数. 我们把每个结果看成一个元素，用 Ω 表示这些元素构成的全集.

1. 摸出 5500 是在 4 种颜色中取定两种颜色，不计次序的组合

数是

$$C_4^2=\frac{4!}{2!2!}=6.$$

这也是事件 $A_1=$“得一等奖”作为 Ω 的子集时，含有的元素个数. 根据概率的定义，

$$P(A_1)=\frac{C_4^2}{C_{20}^{10}}=\frac{6}{184\ 756}\approx 0.000\ 032.$$

这个概率几乎是 0，说明得一等奖几乎是不可能的.

2. 计算摸出 5410 的总数目时，我们用分步计数原理.

第一步，将 4 种颜色进行排列，共 4! 种排法；

第二步，在第 1 种颜色的球中取 5 个球，共 C_5^5 种取法；

第三步，在第 2 种颜色的球中取 4 个球，共 C_5^4 种取法；

第四步，在第 3 种颜色的球中取 1 个球，共 C_5^1 种取法；

第五步，在第 4 种颜色的球中取 0 个球，共 C_5^0 种取法.

根据分步计数原理，摸出 5410 的总数目是

$$4!\times C_5^5\times C_5^4\times C_5^1\times C_5^0=4!\times 5\times 5=600.$$

这也是事件 $A_2=$“得二等奖”作为 Ω 的子集时，含有的元素个数. 根据概率的定义，

$$P(A_2)=\frac{600}{184\ 756}\approx 0.003\ 2.$$

这又是一个概率极小的事件，说明得二等奖也是不大可能的.

3. 仍用分步计数原理计算摸出 5311 的总数目.

第一步，选出充当 53 的颜色进行排列，共有 A_4^2 种方法；对于每种排列有 $C_5^5C_5^3$ 种选球方法. 这一步共有 $A_4^2C_5^5C_5^3$ 种方法.

第二步，在其余两种颜色中的每 5 个球中选取 1 个，共 $C_5^1C_5^1$ 种选法.

根据分步计数原理知道摸出 5311 的总数目是

$$A_4^2C_5^5C_5^3C_5^1C_5^1=(4\times 3)\times 1\times(5\times 2)\times 5\times 5=3\ 000.$$

于是得三等奖的概率是

$$\frac{3\ 000}{184\ 756}\approx 0.016\ 2.$$

这个概率也很小，得三等奖也不大可能.

用类似的方法可以计算出其他结果如下：

4. 摸出 4033 的总数目是

$$A_4^2C_5^4C_5^0C_5^3C_5^3=(4\times3)\times5\times1\times(5\times2)\times(5\times2)=6\ 000.$$

摸出 4411 的总数目是

$$C_4^2C_5^4C_5^4C_5^1C_5^1=6\times5\times5\times5\times5=3\ 750.$$

得四等奖的概率是

$$\frac{6\ 000+3\ 750}{184\ 756}\approx0.053.$$

5. 摸出 4222 的总数目是

$$C_4^1C_5^4C_5^2C_5^2C_5^2=4\times5\times10\times10\times10=20\ 000.$$

得五等奖的概率是

$$\frac{20\ 000}{184\ 756}\approx0.108\ 3.$$

6. 摸出 1234 的总数目是

$$4!C_5^1C_5^2C_5^3C_5^4=60\ 000.$$

摸出 3331 的总数目是

$$C_4^1C_5^3C_5^3C_5^3C_5^1=20\ 000.$$

于是交 2 元游戏费的概率是

$$\frac{60\ 000+20\ 000}{184\ 756}\approx0.433.$$

这是一个概率较大的事件了．所以摸一次球交 2 元钱是很有可能发生的．

7. 摸出 3322 的总数目是

$$C_4^2C_5^3C_5^3C_5^2C_5^2=60\ 000.$$

于是交 5 元钱的概率是

$$\frac{60\ 000}{184\ 756}\approx0.324\ 8.$$

结论：摸一次球交 2 元钱或交 5 元钱的概率总共是

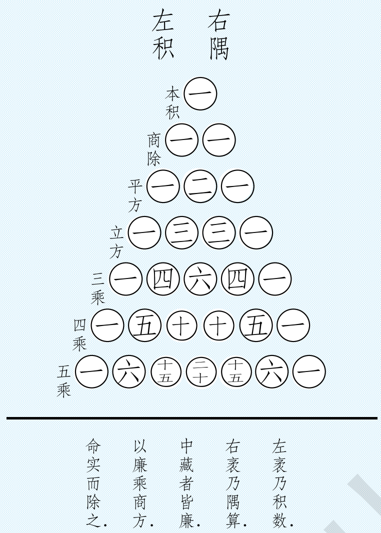
$$0.433+0.324\ 8=0.757\ 8.$$

这个概率相当大，摸 4 次球平均发生 3 次多．所以我们说这种游戏带有欺骗性．

数学文化

杨辉三角

杨辉是我国宋朝的数学家，公元 1261 年他在一本名为《详解九章算法》的书中使用了下面的图，后人称之为“杨辉三角”．杨辉三角最早的出现应当在公元 1200 年以前，确切年代已经很难考证了．类似的图表在欧洲被称为帕斯卡三角，因为许多人认为这是帕斯卡在公元 1654 年发明的．其实在帕斯卡之前已经有人在公元 1527 年将类似杨辉三角的图表印在了算术书的封面上，但是比杨辉三角的出现晚多了．



我们将杨辉三角的前 6 行简写成：

1

1　1

1　2　1

1　3　3　1

1　4　6　4　1

1　5　10　10　5　1

你在杨辉三角中能发现什么规律？你能继续写出杨辉三角的第 7 行、第 8 行吗？

7.4 二项式定理

我们已经学过了

$(a+b)^1=a+b$ 系数是 1 1

$(a+b)^2=a^2+2ab+b^2$ 系数是 1 2 1

$(a+b)^3=a^3+3a^2b+3ab^2+b^3$ 系数是 1 3 3 1

可以看出，上面 3 个式子的系数正好对应杨辉三角的第 2，3，4 行．对照杨辉三角，你能写出$(a+b)^4$，$(a+b)^5$，…的展开式吗？

杨辉三角的特点是**两条斜边上的数字都是 1，其余的数都是它“肩上”的两个数的和．**例如，

第 3 行：2=1+1，

第 4 行：3=1+2，3=2+1，

第 5 行：4=1+3，6=3+3，4=3+1，

……

如果我们注意到杨辉三角的第 1 行是 1，第 2 行是 C_1^0，C_1^1，再利用公式 $C_n^m+C_n^{m-1}=C_{n+1}^m$($m=0$，1，…，$n$)就可以把杨辉三角排成如下形式：

$$
\begin{array}{lccccccccccc}
(a+b)^0 & & & & & & 1 & & & & & \\
(a+b)^1 & & & & & C_1^0 & & C_1^1 & & & & \\
(a+b)^2 & & & & C_2^0 & & C_2^1 & & C_2^2 & & & \\
(a+b)^3 & & & C_3^0 & & C_3^1 & & C_3^2 & & C_3^3 & & \\
(a+b)^4 & & C_4^0 & & C_4^1 & & C_4^2 & & C_4^3 & & C_4^4 & \\
(a+b)^5 & C_5^0 & & C_5^1 & & C_5^2 & & C_5^3 & & C_5^4 & & C_5^5 \\
\cdots\cdots & & & & & & \cdots\cdots & & & & &
\end{array}
$$

根据杨辉三角的性质，可以得到如下的二项式定理．

二项式定理 对于正整数 n，

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

我们用组合的方法证明上述定理.

证明　$(a+b)^n$ 是 n 个 $(a+b)$ 相乘，每个 $(a+b)$ 和其他项相乘时，有两种选择——选 a 或 b.

用 a 表示红球，用 b 表示黑球. 考虑 n 个盒子中的每个盒子里放有红球和黑球各一个. 现从每个盒子中取一个球，分以下情况进行：

a，b	a，b	a，b	…	a，b	a，b

第 1 种情况：a^n 是在每个盒子中都取红球的结果，共有 $C_n^n=C_n^0$ 种取球方法，所以一共有 C_n^0 项 a^n.

第 2 种情况：$a^{n-1}b$ 是在 $n-1$ 个盒子中取红球，在 1 个盒子中取黑球的结果，共有 C_n^1 种取球方法，所以共有 C_n^1 项 $a^{n-1}b$.

……

第 $r+1$ 种情况：$a^{n-r}b^r$ 是在 $n-r$ 个盒子中取红球，在 r 个盒子中取黑球的结果，共有 C_n^r 种取球方法，所以共有 C_n^r 项 $a^{n-r}b^r$. 证毕！

我们称 $C_n^r a^{n-r}b^r$ 是二项展开式的第 $r+1$ 项，其中 C_n^r 称作第 $r+1$ 项的二项式系数. 把

$$T_{r+1}=C_n^r a^{n-r}b^r \text{（其中 } 0\leqslant r\leqslant n，r\in \mathbf{N}，n\in \mathbf{N}_+\text{）}$$

叫作二项展开式的**通项公式**.

下面是从二项式定理中发现的一些基本性质：

1. 二项展开式一共有 $n+1$ 项.

2. 第一个字母 a 按降幂排列，第二个字母 b 按升幂排列.

3. a 的幂加 b 的幂等于 n.

4. 在二项展开式中，与首末两端“等距离”的两项的二项式系数相等，即 $C_n^m=C_n^{n-m}$.

5. 二项式系数从两端向中间逐渐增大，且当 n 是偶数时，中间的一项的二项式系数取得最大值；当 n 是奇数时，中间的两项的二项式系数 $C_n^{\frac{n-1}{2}}$，$C_n^{\frac{n+1}{2}}$ 相等，且同时取得最大值.

6. $C_n^0+C_n^1+C_n^2+\cdots+C_n^n=2^n$. 这可以在二项式定理中取 $a=1$，$b=1$ 得到.

7. $C_n^0-C_n^1+C_n^2+\cdots+(-1)^nC_n^n=0$. 这可以在二项式定理中取 $a=1$，$b=-1$ 得到.

例 1 展开 $\left(3\sqrt{x}-\dfrac{1}{\sqrt{x}}\right)^4$.

解 $\left(3\sqrt{x}-\dfrac{1}{\sqrt{x}}\right)^4=\left(\dfrac{3x-1}{\sqrt{x}}\right)^4=\dfrac{(3x-1)^4}{x^2}$

也可以直接用二项式定理展开.

$$=\frac{1}{x^2}[(3x)^4-C_4^1(3x)^3\cdot 1+C_4^2\cdot(3x)^2\cdot 1^2-C_4^3\cdot(3x)\cdot 1^3+C_4^4]$$

$$=\frac{1}{x^2}(81x^4-108x^3+54x^2-12x+1)$$

$$=81x^2-108x+54-\frac{12}{x}+\frac{1}{x^2}.$$

例 2 计算 $(x+2y)^9$ 的展开式中第 5 项的系数和二项式系数.

解 $(x+2y)^9$ 的展开式的第 5 项是

$$T_5=T_{4+1}=C_9^4\cdot x^{9-4}\cdot(2y)^4=C_9^4\cdot 2^4\cdot x^5y^4,$$

所以展开式的第 5 项的二项式系数是 $C_9^4=126$，展开式的第 5 项的系数是 $C_9^4\cdot 2^4=2\ 016$.

例 3 用二项式定理及性质求解下列问题.

(1) 若 $(1+x)^n$ 的展开式中，x^3 的系数是 x 的系数的 7 倍，求 n 及二项式系数的最大值，其中 $n\in\mathbf{N}_+$；

(2) 已知在 $(ax-1)^5$ 的展开式中，x^3 的系数是 -80，求 a.

解 (1) x^3 的系数是 C_n^3，x 的系数是 C_n^1. 依题意有 $C_n^3=7C_n^1$，即

$$\frac{n(n-1)(n-2)}{3!}=7n.$$

整理得 $$n^2-3n-40=0,$$

解得 $n=8$.（舍去 $n=-5$.）

$\therefore$ 二项式系数在 $r=4$ 时取得最大值，即 $C_8^4=70$.

(2) $\because$ x^3 的系数是 $C_5^2a^3$，

$\therefore$ $C_5^2a^3=-80$，

即 $10a^3=-80$.

解得 $a=-2$.

例 4 已知 $(1-2x)^7=a_0+a_1x+a_2x^2+\cdots+a_7x^7$，计算 a_1+

$a_2+\cdots+a_7$.

解 取 $x=0$ 得到 $a_0=1$. 再取 $x=1$，得 $-1=a_0+a_1+a_2+\cdots+a_7$，于是 $a_1+a_2+\cdots+a_7=-1-a_0=-2$.

例 5 在 $\left(1+\frac{2}{x}\right)(1-x)^9$ 的展开式中求 x^5 的系数.

解 原式可化为 $(-x+1)^9+\frac{2}{x}(-x+1)^9$，

所以含 x^5 的项为

$$C_9^4(-x)^5+\frac{2}{x}C_9^3(-x)^6=-126x^5+168x^5=42x^5.$$

因此，x^5 的系数是 42.

例 6 用二项式定理证明 $(n+1)^n-1$ 可以被 n^2 整除 $(n\in\mathbf{N}_+)$.

解 用二项式定理和 $C_n^1=n$，得到

$$\begin{aligned}(n+1)^n-1&=C_n^0+C_n^1n+C_n^2n^2+\cdots+C_n^nn^n-1\\&=n^2+C_n^2n^2+\cdots+C_n^nn^n\\&=n^2(1+C_n^2+\cdots+C_n^nn^{n-2}).\end{aligned}$$

所以 $(n+1)^n-1$ 是 n^2 的倍数，即可以被 n^2 整除.

例 7 在二项式 $\left(\sqrt{x}+\frac{1}{2\sqrt[4]{x}}\right)^n$ $(n\in\mathbf{N}_+)$ 的展开式中，前三项的系数成等差数列，求展开式中所有的有理项.

解 此二项式的展开式的通项公式为

$$T_{r+1}=C_n^r(\sqrt{x})^{n-r}\left(\frac{1}{2\sqrt[4]{x}}\right)^r=C_n^r\frac{1}{2^r}x^{\frac{2n-3r}{4}}.$$

分别取 $r=0$，1，2，得前三项的系数分别为

$$t_1=1,\ t_2=C_n^1\frac{1}{2}=\frac{1}{2}n,\ t_3=C_n^2\frac{1}{4}=\frac{1}{8}n(n-1).$$

又由题意可得 $2t_2=t_1+t_3$，即 $n=1+\frac{1}{8}n(n-1)$，解得 $n=8$（$n=1$ 舍去）.

因此，此二项式的展开式的通项公式为 $T_{r+1}=C_8^r\frac{1}{2^r}x^{\frac{16-3r}{4}}$，$r=0$，1，2，…，8.

设 T_{r+1} 为有理项，则 $16-3r$ 是 4 的倍数，$\therefore r=0$，4，8.

因此，展开式中的所有有理项分别为 $T_1=x^4$，$T_5=C_8^4\,\frac{1}{2^4}x=\frac{35}{8}x$，$T_9=C_8^8\,\frac{1}{2^8}x^{-2}=\frac{1}{256x^2}$.

练习

1. 展开二项式 $\left(1+\frac{1}{x}\right)^4$.

2. 已知 $\left(\sqrt{x}+\frac{3}{\sqrt[3]{x}}\right)^n$ $(n\in\mathbf{N}_+)$ 的展开式中，各项系数的和与其各项二项式系数的和之比为 64∶1，则 $n=$________.

3. 已知 $(1+2x)^n$ $(n\in\mathbf{N}_+)$ 的展开式中第 7 项和第 8 项的二项式系数相等，求展开式中系数最大的项及二项式系数最大的项.

习题 7

学而时习之

1. 在 $(1-x)^{10}$ 的展开式中，x^5 的系数是多少？
2. 在 $x^2(1-x)^{10}$ 的展开式中，x^5 的系数是多少？
3. 在 $(1+x+x^2)(1-x)^{10}$ 的展开式中，x^5 的系数是多少？
4. （1）计算二项式 $(3x+1)^8$ 的展开式中 $(3x)^k$ 的系数和 x^k 的系数；

   （2）计算二项式 $(3x+1)^8$ 的系数之和.
5. 已知 $(1+3x)^8=a_0+a_1x+a_2x^2+\cdots+a_7x^7+a_8x^8$，计算：

   (1) $a_0+a_1+\cdots+a_7+a_8$；

   (2) $a_0-a_1+a_2-a_3+\cdots-a_7+a_8$；

   (3) $a_1+a_2+\cdots+a_8$.
6. 用二项式定理证明 $99^{10}-1$ 可以被 1 000 整除.
7. 设 $f(x)=(1+x)^m+(1+x)^n$ $(m,\ n\in\mathbf{N}_+)$ 的展开式中 x 的系数是 19.

   （1）求 $f(x)$ 展开式中 x^2 的系数的最小值及 m，n 的值；

   （2）当 $f(x)$ 展开式中 x^2 的系数取最小值时，求 x^7 的系数.

数学文化

地图染色和四色定理

地图上标有各国的疆域，为了区分明确，相邻的国家需要采用不同的颜色加以区别．当我们仔细观察地图就会发现，一般只需要四种不同的颜色给地图着色就够了．

如果区域更多一些，四种颜色还够不够呢？这个看似简单的问题首先被格斯里在1852年提出．格斯里在对英国地图上色时发现，无论多么复杂的地图，只需要四种颜色就可以使得相邻的区域着上不同的颜色．这件事情引起了他的兴趣，他感到其中可能隐藏着某种科学道理．他把想法告诉了哥哥费特里，费特里又把这个问题转交给了著名数学家**德·摩尔根**(De Morgan)，德·摩尔根也解释不了，就写信给著名的数学家哈密顿．这位著名的数学家也被此问题弄得一筹莫展，直到逝世也无结果．

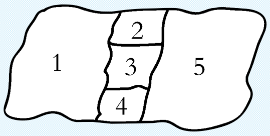
1876年，著名数学家凯莱在数学年会上把这个问题归纳为“四色猜想”提出，于是“四色猜想”开始引起人们的关注．但是其难度并未引起大家的注意．

闵可夫斯基是位为人谦虚、成就不凡的数学家，偏偏在给大学生上课时，提到了这个问题．当时他一时兴起，失言道：“四色猜想之所以没有解决，是因为当今世界上第一流的数学家没有研究它．”说着拿起粉笔在课堂上即兴推演，没想到越写越多，越写越复杂，终于挂了黑板(指讲不下去了)．

首先宣布“证明”了四色猜想的是一个叫肯普的律师，他于1879年发表了自己的证明方法，可是过了11年，年仅29岁的年轻数学家希伍德指出肯普的证明不能成立，接着希伍德成功地使用了肯普的技巧，证明出平面地图最多用五种颜色着色就够了．

直到 1976 年，美国的数学家阿佩尔和哈肯设计出一个计算机程序，他们同时启动三台超高速电子计算机，耗用 1 200 个机时，终于用计算机证明了“四色定理”，但不借助计算机的纯数学证明至今尚未作出.

四色定理是一个和组合、拓扑、图论有关的问题，有兴趣的同学可以考虑以下问题：在一个具有五个行政区域的地图（如下图）上，用四种颜色给这五个行政区着色，当相邻的区域不能使用同一种颜色时，共有多少种着色方法？



五个行政区的地图

多知道一点

利用计算机或计算器计算组合数

使用软件 MatLab 计算排列数、组合数和阶乘

1. 计算 n^k 时用语句 n^k. 例如计算 5^{11} 时，输入

   N=5^11

   N=48828125（计算的结果）

2. 计算 C_n^k 时用语句 nchoosek（n，k）. 例如计算 C_{15}^9 时，输入

   C=nchoosek（15，9）

   C=5005.00（计算的结果）

3. 计算 A_n^k 用语句 nchoosek（n，k）* factorial（k）. 例如计算 A_{15}^9 时，输入

   A=nchoosek（15，9）* factorial（9）

   A=1816214400.00（计算的结果）

4. 计算 $n!$ 用语句 factorial（n）. 例如计算 11! 时，输入

   G=factorial（11）

   G=39916800.00（计算的结果）

通过例子学习用计算器计算排列数、组合数和阶乘

1. 计算 5^{11} 时，输入

   5 $\boxed{x^y}$ 11 $\boxed{=}$ 48828125 （计算的结果）

2. 计算 C_{15}^9 时，输入

   15 $\boxed{nC_r}$ 9 $\boxed{=}$ 5005 （计算的结果）

3. 计算 A_{15}^9 时，输入

   15 $\boxed{\text{SHIFT}}$ $\boxed{nP_r}$ 9 $\boxed{=}$ 1816214400 （计算的结果）

4. 计算 11! 时，输入

11 SHIFT $x!$ = 39916800 （计算的结果）

用“Z+Z 超级画板”程序工作区计算排列数、组合数和阶乘

1. 计算 5^{11} 时，键入

   5^11；（按 Ctrl+Enter 键执行，下同）

   ≫48828125＃（计算的结果，下同）

2. 计算 C_{15}^{9} 时，键入

   C(15,9)；

   ≫5005＃

3. 计算 A_{15}^{9} 时，键入

   P(15,9)；

   ≫1816214400＃

4. 计算 11！时，键入

   Factorial(11)；

   ≫39916800＃

小结与复习

我们把排列、组合的有关结论总结如下：

1. **分类计数原理**：如果完成一件事有 n 类办法，在第一类办法中有 m_1 种不同的方法，在第二类办法中有 m_2 种不同的方法，…，在第 n 类办法中有 m_n 种不同的方法，每种方法都可以完成这件事，那么完成这件事共有

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

种不同的方法.

2. **分步计数原理**：如果完成一件事需要分成 n 步，第一步有 m_1 种不同的方法，第二步有 m_2 种不同的方法，…，第 n 步有 m_n 种不同的方法，那么完成这件事共有

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

种不同的方法.

3. **排列**：从 n 个不同元素中取出 $m(m\leqslant n)$ 个不同的元素，按照一定的顺序排成一列，叫作从 n 个不同元素中取出 m 个元素的一个排列，简称为排列. 用符号 A_n^m 表示排列的个数时，有

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

4. **全排列**：把 n 个不同元素排成一列，所有不同的排列数有 $A_n^n=n!$ 个.

5. **组合**：从 n 个不同的元素中取出 $m(m\leqslant n)$ 个不同的元素，不论次序地构成一组，称为一个组合. 用符号 C_n^m 表示所有不同的组合个数时，

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

6. **组合数的性质**：

(1) 如果 $C_n^m=C_n^k$，有 $m=k$ 或者 $m=n-k$.

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

7. 二项式定理： 对于正整数 n，有

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

复习题七

学而时习之

1. 从甲地到乙地，可以乘火车，也可以乘汽车，还可以乘轮船. 一天中，火车有 3 班，汽车有 5 班，轮船有 3 班. 那么一天中乘坐这些交通工具从甲地到乙地共有多少种不同的走法？

2. 判断下列问题是排列问题还是组合问题.

   (1) 从 9 种不同的小麦良种中选出 4 种，有多少种选法？

   (2) 从 50 件不同的产品中随机抽出 5 件来检查，有多少种不同的等可能结果？

   (3) 5 个人互送贺年卡 1 张，共送了多少张贺年卡？

3. 有 11 个队参加排球赛，比赛时分成两组，第一组 5 个队，第二组 6 个队. 各组都进行单循环赛(即每两队都要比赛一场)，共需要比赛多少场？

4. 在产品质量检验问题中，需要从 100 件产品中随机抽取 6 件进行检查. 设这 100 件产品中有 5 件次品，95 件正品，根据以下的要求，计算各有多少种等可能结果？(只需用组合数表达结果，不必将组合数计算出来.)

   (1) 任取 6 件；

   (2) 抽到的全是正品；

   (3) 抽到 2 件正品；

   (4) 抽到至少 1 件次品；

   (5) 抽到 2 件次品.

5. 12 件产品中，有 5 件一等品，4 件二等品，3 件三等品. 现在从中抽出 4 件，一等品有 2 件，二等品有 1 件，三等品有 1 件的不同结果有多少个？

6. 从 1，2，3，4，5，6 这六个数字中无重复地取两个，和为偶数的取法有多少种？

7. 某校高中一年级有 8 个班，高二年级有 7 个班，高三年级有 8 个班．各年级分别进行班与班的排球单循环赛，一共需要比赛多少场？

8. 7 名同学站成一排，共有多少种不同的排法？

9. 求 $\left(x-\dfrac{1}{x}\right)^{10}$ 的展开式中 x^4 的系数．

10. $(a+b)^2(b+c)^3$ 的展开式中 ab^3c 的系数为（　　）

　A．2　　B．3　　C．4　　D．6

11. $(1+x)^n(n\in\mathbf{N}_+)$的二项展开式中，若只有 x^5 的系数最大，则 n 等于（　　）

　A．7　　B．8　　C．10　　D．11

12. 在$(1-x^2)^{20}$的展开式中，如果第 $4r$ 项和第 $r+2$ 项的二项式系数相等．

（1）求 r 的值；

（2）写出展开式中的第 $4r$ 项和第 $r+2$ 项．

温故而知新

13. 5 个不同的小球放入 5 个不同的盒中，恰有一个空盒的放法有多少种？

14. 全班有 38 名同学，其中正、副班长各 1 名．现选派 8 名同学参加某项学习竞赛，在下列情况下，各有多少种不同的选法？

（1）无任何限制条件；

（2）两名班长必须入选；

（3）两名班长有且只能有一人入选；

（4）两名班长都不入选；

（5）两名班长至少有一人入选；

（6）两名班长至多有一人入选．

15. 圆上有两两不同的 8 个点．

（1）过两点可画一条弦，一共可画多少条弦？

（2）过三点可画一个圆内接三角形，一共可画多少个内接三角形？

16. 由数字 1，2，3，4，5，6 可以组成多少个没有重复数字的正整数？

17. 新华书店有语文、数学、英语辅导书各 10 种．

（1）买其中一本有多少种方法？

(2) 买两本不同类的书有多少种方法?

18. 某工厂有三个车间,第一车间有 4 个小组,第二车间有 5 个小组,第三车间有 6 个小组,有一个新工人分配到该工厂工作,有几种不同的安排?

19. 完成一件产品需要三道工序,这三道工序分别由第一、第二、第三车间来完成. 第一车间有 3 个小组,第二车间有 4 个小组,第三车间有 5 个小组,每一个车间的小组都只能完成该车间规定的工序,问完成这件产品有几种不同的分派方案?

20. 9 名同学站成两排,前 4 后 5,共有多少种不同的排法?

21. 11 名同学站成一排,甲站在中间的位置,共有多少种不同的排法?

22. 从 9 个不同的文艺节目中选 6 个编成一个节目单,如果演员甲的独唱不能排为第一个节目,共有多少种不同的排法?

23. 6 本不同的书全部送给 5 人,每人至少 1 本,有多少种不同的送书方法?

24. 已知 $f(x)=(2x+1)^m+(6x+1)^n$(m, $n\in\mathbf{N}$)的展开式中含 x 项的系数为 24,求展开式中含 x^2 项的系数的最小值.

25. 已知 $(1+2x)^n=\sum\limits_{k=0}^{n}\mathrm{C}_n^k(2x)^k=\sum\limits_{k=0}^{n}a_kx^k$ $(n\in\mathbf{N}_+)$. $(1+2x)^n$ 的展开式中末三项的二项式系数的和为 92,判断展开式系数组成的数列 a_0, a_1, $\cdots$, a_n 的单调性,并求其最大项.

上下而求索

26. 从 9 名男生和 6 名女生中挑选 5 人,最多选到 2 名女生的选法有多少种?

27. 从 5 名男生和 4 名女生中选出 5 名代表.

(1) 要求男生 2 名,女生 3 名且女生李晶必须在内的选法有多少种?

(2) 要求男生不少于 2 名的选法有多少种?

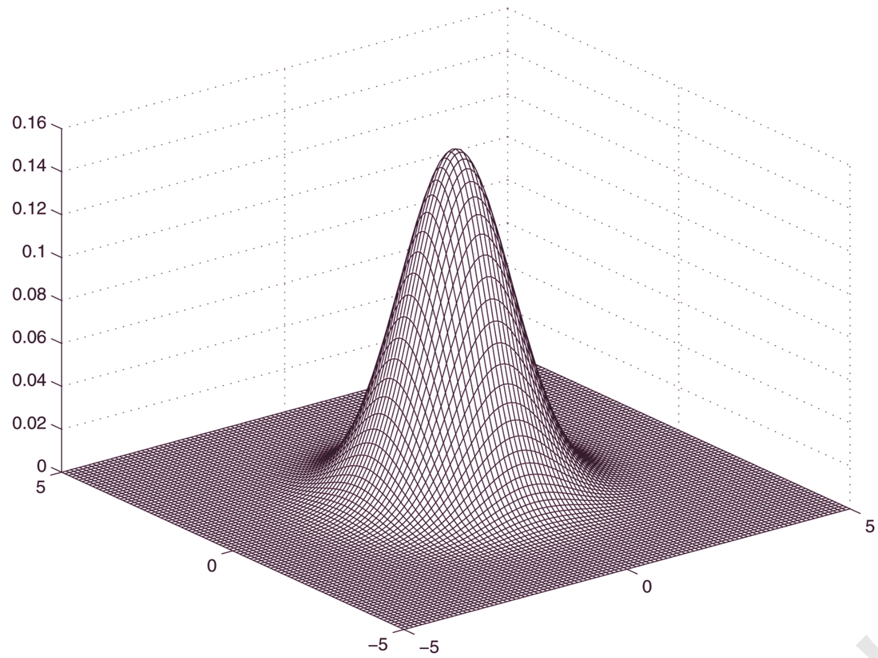
28. 8 名同学站成一排.

(1) 甲、乙两同学必须相邻的排法共有多少种?

(2) 甲、乙、丙三个同学必须相邻的排法共有多少种?

第 8 章

统计与概率



统计学是研究如何从数据中提取有用信息的科学，内容包括如何收集和分析数据. 基于统计学的数据处理方法称为统计方法. 在科学研究、工农业生产、新产品开发、产品质量的提高乃至政治、教育、社会科学等各个领域，使用统计方法和不使用统计方法获得的结果是大不相同的. 只要统计方法使用得当，就能够得到事半功倍的效果. 这也是统计学能随着科学技术和国民经济的发展而快速发展的重要原因.

8.1 随机对照试验

收集数据的方法之一是从总体中进行抽样，另外一个方法是在试验中得到观测数据．为了能根据试验的数据对试验进行合理的分析，需要对试验进行合理的安排．

案例 1 （坏血病的研究）17 世纪初期，长期在海上航行的水手经常患坏血病．坏血病的症状是牙龈肿大出血，皮肤上出现青灰的斑点．英国海军部试图考察坏血病的起因．他们怀疑这是因为水手体内缺少柑橘类水果中的某种成分造成的．当此想法提出时，刚好有 4 艘军舰要远航．为了调查是否由于水手缺乏柑橘类的水果而导致坏血病，海军部设计了一次试验：随机地安排了一艘军舰上的水兵每天喝柑橘汁，另外 3 艘军舰不供应柑橘汁．

试验的结果是：航行还没有结束，没有喝柑橘汁的水兵多数得了坏血病，而提供柑橘汁的军舰上的水兵没有发现坏血病．最后，提供柑橘汁的军舰不得不把携带的柑橘汁分给其他的军舰，以帮助他们顺利返航．

尽管本次试验的计划还可以从各个方面进行改进，但是试验的结果成功地证实了最初的怀疑．

在案例 1 中，我们称喝柑橘汁的水兵为**试验组**（experimental group），称不喝柑橘汁的水兵为**对照组**（control group）．

试验组由随机选择出的对象构成，试验组的成员要接受某种特殊的待遇或治疗等．而对照组由那些没有接受这种特殊待遇的对象构成．一个好的试验设计都应当有一个试验组和一个对照组．

在案例 1 中，如果没有对照组，为 4 艘军舰都提供柑橘汁，就没有水兵患上坏血病，海军部就不能确认他们的最初怀疑．因为不能确定是否是其他的食品或治疗避免了坏血病．

为什么试验组要随机抽取呢？

设想在案例 1 中，如果安排喜欢喝柑橘汁的水兵在试验组，喜欢喝啤酒的水兵在对照组，就不能确定研究开始前这两组水兵的身体状

无论人们意识与否，统计学存在于工农业生产、国民经济和日常生活之中．不懂统计可能会造成不知不觉的损失．

本文讲述随机对照试验的目的不是为了引出概率话题，而是想让同学们知道进行随机对照试验的重要性：没有随机对照试验的试验很可能导致严重的错误后果．

这里的案例尽管都较长，但是没有任何数学上的困难，同学们可以像读故事一样学习到随机对照试验的基本原理和方法．

况是否有差异．水兵身体状况的差异也可能影响是否容易得坏血病．随机选择试验组能够有效地减少个体差异造成的对试验结果的影响．

随机选择试验对象是英国统计学家费歇尔(Fisher)的贡献，在 20 世纪初，他用此方法致力于农业试验的研究．从此随机选择试验组成为安排试验的基本原则．

案例 2 （静脉吻合分流术)在一些肝硬化病例中，许多病人会因肝出血而导致死亡．历史上有一种称为“静脉吻合分流术”的外科手术用于治疗肝硬化，其原理是运用外科手术的方法使血流改变方向．这种手术花费很大，并且有很大的危险性．值得做这样的手术吗？

为了解决上述问题，一共进行了三批共 51 次手术试验．第一批进行了 32 次无对照组的手术试验．结果如下：

设计方法	试验次数	显著有效	中等有效	无效
	32	24	7	1
所占比例		75％	21.9％	3.1％

试验说明有 75％的手术显著有效，21.9％的手术中等有效，看来手术是值得做的．

第二批共进行了 15 次手术试验，这批试验有对照组，但是对照组的病人不是随机选取的．医生根据病人的临床诊断情况决定将病人编入试验组做手术或编入对照组不做手术．结果如下：

设计方法	试验次数	显著有效	中等有效	无效
	15	10	3	2
所占比例		66.7％	20％	13.3％

这次试验的结果是 66.7％的手术显著有效，20％的手术中等有效，13.3％的手术无效．这个试验结果也是对“静脉吻合分流术”的肯定．这次的结果与无对照组的试验结果差别不是很大．

再看有随机选取的对照组的第三批试验．这批试验只有 4 次手术．随机选取的方式可以是掷硬币，如果硬币正面朝上就将病人选入试验组做手术，否则放入对照组不做手术．这次试验的结果如下：

以上数据来源于 Journal of Gastroenterology，Vol. 50 (1966)，pp. 646～691.

设计方法	试验次数	显著有效	中等有效	无效
	4	0	1	3
所占比例		0％	25％	75％

随机对照试验的结果显著地否定了“静脉吻合分流术”.

结果显示：没有随机选取对照组的前两批试验过分夸大了“静脉吻合分流术”的价值. 经过认真设计的带有随机选取对照组的试验显示“静脉吻合分流术”几乎没有什么价值.

为什么会出现如此大的差别呢?

在无对照组和非随机选取对照组的试验中，试验者根据病人的临床诊断决定是否将他编入试验组进行手术. 这样做就出现一种自然的倾向：试验人员更倾向于将那些身体状态较好的病人选入试验组，以减少手术风险. 其结果有利于对手术的肯定评价，这种结果是不真实的.

对上述试验的跟踪观测发现，做手术的51个病人中3年后大约有60%仍然活着，随机对照组中(没做手术的病人)3年后大约也有60%的病人仍然活着. 这就说明手术基本是无效的. 而在非随机对照组中，只有45%的病人存活期超过3年，这就说明了非随机对照组中的病人健康情况较差，验证了健康情况较好的病人更容易被选入试验组做手术.

随机安排对照组是十分必要的，否则可能得出错误的结论.

我们称随机选取试验组的对照试验为**随机对照试验**.

在随机对照试验中，为了得到更真实的结果，有时还需要其他的手段配合.

案例3 1916年小儿麻痹症(脊髓灰质炎)袭击了美国，以后的40年间，受害者成千上万. 20世纪50年代，人们开始发现预防疫苗. 当时**萨凯**(Salk)培育的疫苗最有希望. 他的疫苗在实验室中表现良好：安全，产生对脊髓灰质炎病毒的抗体. 但是在大规模使用前必须进行临床试验，通过试验最后确定疫苗是否有效. 只有这样才能达到保护儿童的目的.

当时采用了随机对照的研究方案，对每个儿童用类似投掷一枚硬币的方法决定是否将他编入试验组：正面朝上分在试验组，否则分在对照组. 除了试验的设计人员，连医生也不知道哪个儿童分在试验组，哪个儿童分在对照组.

然后给分在试验组的儿童注射疫苗，给分在对照组的儿童注射生理盐水，让他们认为也被注射了疫苗．得到的结果如下：

	试验人数	试验后的发病率
试验组	20 万	28/100 000
对照组	20 万	71/100 000

试验结果显示，疫苗将小儿麻痹症的发病率从 71/100 000 降低到 28/100 000．由于 71 和 28 的差别超出了随机性本身所能解释的范围，所以宣布疫苗是成功的．

我们把对照组中的处理方法称为使用**安慰剂**，案例 3 中的安慰剂是注射生理盐水．给对照组的儿童使用安慰剂是为了避免儿童的心理作用影响试验的结果．尽管可以认为光靠精神作用不能抵抗小儿麻痹症，但是为了确认试验结果的可靠性，使用安慰剂是必要的．

不让医生知道儿童是来自试验组还是对照组是为了使医生能够作出更公正的诊断，避免在诊断儿童是否患有小儿麻痹症时受到心理因素的影响．

在许多场合，心理因素是不能忽视的．有资料显示，在医院中给那些手术后产生剧痛的病人服用由淀粉制成的“止痛片”后，大约有 1/3 的病人感觉剧痛减轻．

练习

某个社会团体的 18 名理事开会决定是否增加一名新理事甲某，哪种选举方式最能体现与会理事们的真实意愿（　　），哪种选举方式有利于甲入选理事会（　　）．

(A) 请同意增选甲为理事的举手

(B) 请不同意增选甲为理事的举手

(C) 采用记名投票

(D) 采用无记名投票

习题 1

学而时习之

在评价一种治疗高血压的磁疗手表时，调查了 100 名刚开始使用这种手表的高血压患者，他们中有 75 人回答磁疗手表对降低高血压有效.

(1) 能否作出这种磁疗手表对降低高血压的有效率是 75%的结论？

(2) 设计一种能够公正评价这种磁疗手表的试验方案.

(3) 你的设计中试验组和对照组是随机选取的吗？

(4) 在你的设计中，使用“安慰剂”了吗？安慰剂是什么？

(5) 在对参加试验的人进行高血压的测量时，你让医生知道被测者使用的是磁疗手表还是外观完全相同的普通手表吗？

8.2 概率

8.2.1 概率的加法公式

学习概率时，我们把一个试验的可能结果称为试验的元素，把该试验元素构成的集合称为试验的全集．用 Ω 表示试验的全集时，称 Ω 的子集为事件．当两个事件不能同时发生时，称这两个事件互斥．事件 A 和事件 B 互斥的条件是 $A\cap B=\varnothing$.

设全集 Ω 中有有限个元素，$A\subseteq\Omega$．如果 Ω 中每个元素发生的可能性相同，则称 $P(A)=\dfrac{A\text{ 中元素数}}{\Omega\text{ 中元素数}}$为 A 发生的概率，简称为 A 的概率.

在计算事件 A 的概率时，先计算全集 Ω 中元素的个数，然后计算事件 A 中元素的个数．特别要注意，全集 Ω 中每个元素发生的可

在考虑一个未来事件是否会发生的时候，人们常关心该事件发生的可能性的大小．概率就是用来测量该未来事件发生的可能性的度量.

元素作为试验的可能结果，又被统计学家们称为试验的**样本点**(sample outcome)或基本事件.

试验的全集 Ω 又被统计学家们称为**样本空间**(sample space).

能性必须相同.

对于全集 Ω 的事件 A_1，A_2，…，A_m，$A_1\cup A_2\cup\cdots\cup A_m$ 表示事件 A_1，A_2，…，A_m 中至少有一个发生.

概率的加法公式　如果 Ω 的事件 A_1，A_2，…，A_m 两两互斥，则

$$P(A_1\cup A_2\cup\cdots\cup A_m)=P(A_1)+P(A_2)+\cdots+P(A_m).$$

运用概率的加法公式的前提是全集 Ω 的事件 A_1，A_2，…，A_m 两两互斥.

证　集合 A_1，A_2，…，A_m 两两互斥，于是

$$\begin{aligned}&(A_1\cup A_2\cup\cdots\cup A_m)\text{ 中元素数}\\&=A_1\text{ 中元素数}+A_2\text{ 中元素数}+\cdots+A_m\text{ 中元素数}.\end{aligned}$$

最后得到

$$\begin{aligned}&P(A_1\cup A_2\cup\cdots\cup A_m)\\&=\frac{(A_1\cup A_2\cup\cdots\cup A_m)\text{ 中元素数}}{\Omega\text{ 中元素数}}\\&=\frac{A_1\text{ 中元素数}+A_2\text{ 中元素数}+\cdots+A_m\text{ 中元素数}}{\Omega\text{ 中元素数}}\\&=\frac{A_1\text{ 中元素数}}{\Omega\text{ 中元素数}}+\frac{A_2\text{ 中元素数}}{\Omega\text{ 中元素数}}+\cdots+\frac{A_m\text{ 中元素数}}{\Omega\text{ 中元素数}}\\&=P(A_1)+P(A_2)+\cdots+P(A_m).\end{aligned}$$

例 1　根据以往的经验，某家庭装修公司每月能够签订 k 份装修合同的概率是 p_k. 已知 $p_k=C_{20}^{k}0.6^{k}0.4^{20-k}$，$k=0$，1，…，20. 计算以下概率.

(1) 下月能够签订 11 份装修合同的概率；

(2) 下月能够签订 11～13 份装修合同的概率.

解　用 A_k 表示下月签订 k 份装修合同，

(1) 则 $P(A_{11})=p_{11}=C_{20}^{11}0.6^{11}0.4^{9}=0.1597$；

(2) A_{11}，A_{12}，A_{13} 两两互斥，因为签订 11 份装修合同发生，签订 12 份装修合同就不发生……

$A_{11}\cup A_{12}\cup A_{13}$ 表示签订的合同在 11～13 份之间. 计算得到

$$\begin{aligned}&P(A_{11}\cup A_{12}\cup A_{13})\\&=P(A_{11})+P(A_{12})+P(A_{13})\\&=p_{11}+p_{12}+p_{13}\\&=\mathrm{C}_{20}^{11}0.6^{11}0.4^{9}+\mathrm{C}_{20}^{12}0.6^{12}0.4^{8}+\mathrm{C}_{20}^{13}0.6^{13}0.4^{7}\\&=0.1597+0.1797+0.1659\\&\approx 0.5.\end{aligned}$$

签订 11～13 份装修合同的概率约等于 0.5.

用 $\overline{A}=\Omega\setminus A$ 表示 A 的对立事件，则 $\overline{A}$ 是 A 的补集.

例 2 计算例 1 中的家庭装修公司下个月至少签订 5 份装修合同的概率.

解 仍用 A_k 表示下月签订 k 份装修合同，则 A_0，A_1，…，A_{20} 两两互斥. $A=A_0\cup A_1\cup A_2\cup A_3\cup A_4$ 表示至多签订 4 份装修合同. A 的对立事件 $\overline{A}=\Omega\setminus A$ 表示至少签订 5 份装修合同，

$$\begin{aligned}P(A)&=P(A_0)+P(A_1)+\cdots+P(A_4)\\&=p_0+p_1+\cdots+p_4\\&=\mathrm{C}_{20}^{0}0.6^{0}0.4^{20}+\mathrm{C}_{20}^{1}0.6^{1}0.4^{19}+\cdots+\mathrm{C}_{20}^{4}0.6^{4}0.4^{16}\\&\approx 0.0017.\end{aligned}$$

于是 $$P(\overline{A})=1-P(A)\approx 0.9983.$$

说明该装修公司下月至少能够签订 5 份装修合同的概率约等于 99.83%.

练习

某人每天打出 k 次电话的概率 p_k 如下：

k	0	1	2	3	4	5	6	7	8
p_k	0.01	0.02	0.07	0.17	0.25	0.25	0.16	0.06	0.01

如果每打一个电话的话费是 0.3 元，计算：

（1）明天用 0.3 元电话费的概率；

（2）明天用 0.6 元电话费的概率；

（3）明天至少用 0.6 元电话费的概率；

（4）明天用 2.4 元电话费的概率.

习题 2

学而时习之

1. 某人的手机在一天内收到 k 条短信的概率 p_k 如下：

k	0	1	2	3	4	5	6	7	8
p_k	0.01	0.06	0.16	0.25	0.25	0.17	0.07	0.02	0.01

（1）计算该手机明天收到最多 3 条短信的概率；

（2）计算该手机明天收到至少 3 条短信的概率；

（3）计算该手机明天收到的短信在 3～6 条之间的概率；

（4）计算该手机明天收到的短信数是奇数的概率.

2. 一批产品有 10 件，其中含有 4 件次品，从中随机抽取 3 件. 计算：

（1）这 3 件产品都是次品的概率；

（2）这 3 件产品都是正品的概率；

（3）这 3 件产品都是次品或都是正品的概率.

8.2.2 条件概率

一个班有 23 名男生，20 名女生. 从班中随机选出一名数学课代表时，男生甲被选到的概率是 $\frac{1}{43}$. 当老师公布出选举的结果是女生时，甲被选到的概率是零. 我们称已知选取的是女生的条件下，男生甲被选到的概率是零. 如果老师公布选举的结果是男生，甲被选到的概率是多少呢？

已知选举出男生的条件下，甲被选到的概率是 1/23.

为了讲述条件概率，我们先分析下面的问题.

问题 掷一枚骰子，已知掷出了奇数，求这个奇数是 3 的概率.

答案是 1/3，结论是一目了然的. 但是为了探讨更复杂一些的问题，还是需要把问题分析清楚.

分析 已知掷出奇数后，试验的可能结果只有 3 个，它们是点数 1，3，5. 这是新的全集，用 $A=\{1, 3, 5\}$表示这个新的全集.

由于全集已经不是投掷一枚骰子的全集，我们称试验的条件已经改变，称 A 是新的试验条件下的全集. 这里新的试验指投掷一枚骰子和已知掷出了奇数.

A 中的 3 个元素处于相等的地位，所以发生的可能性是相同的，用 B 表示掷出点数 3，B 是 A 的子集，A 中元素数$=3$，B 中元素数$=1$. 所以用 $P(B \mid A)$表示已知掷出奇数的条件下，掷出 3 的概率时，有

$$P(B|A)=\frac{B\text{ 中元素数}}{A\text{ 中元素数}}=\frac{1}{3}.$$

此时 B 是 A 的子集. 若 B 不是 A 的子集，则 $P(B \mid A)=\frac{(A\cap B)\text{中元素数}}{A\text{ 中元素数}}$.

例 1 某校高中三个年级各派一名男生和一名女生参加市里的中学生运动会，每人参加一个不同的项目，且每人是否获得冠军是等可能的. 已知只有一名女生获得冠军，求高一的女生获得冠军的概率.

解 用 A 表示只有一名女生获得冠军，用 B 表示高一女生获得冠军. 已知 A 发生的条件下，A 成为试验的全集，A 的元素具有等可能性，B 是A 的子集，A 中元素数$=3$，B 中元素数$=1$. 所以，用 $P(B|A)$表示要求的概率时，$P(B|A)=\frac{B\text{ 中元素数}}{A\text{ 中元素数}}=\frac{1}{3}$. 我们称 $P(B|A)$是已知事件 A 发生的条件下，事件 B 发生的概率.

注意：因为 $P(A)=1/2$，$P(A\cap B)=1/6$，所以有

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

例 2 在一副扑克的 52 张(去掉两张王牌后)中任取 1 张，已知抽到草花的条件下，求抽到的是草花 5 的概率.

草花就是我们常说的梅花.

解 $A=$ “抽到草花”，$B=$ “抽到草花 5”. 已知 A 发生的条件下 A 成为试验的全集，A 中的元素发生的可能性相同，B 是 A 的子集，所以

$$P(B|A)=\frac{B\text{ 中元素数}}{A\text{ 中元素数}}=\frac{1}{13}.$$

因为 $P(A)=13/52$，$P(A\cap B)=1/52$. 所以也有

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

设 A，B 是事件，且 $P(A)>0$，以后总是用 $P(B|A)$表示在已知 A 发生的条件下 B 发生的条件概率，简称为**条件概率**. 下面是条件概率的计算公式.

条件概率公式：如果 $P(A)>0$，则

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

条件概率公式有时会带来许多计算的方便. 但有时候根据问题的特点可以直接得到结果.

例 3 把一副扑克的 52 张(去掉两张王牌后)随机均分给赵、钱、孙、李四家，A=“赵家得到 6 张草花”，B=“孙家得到 3 张草花”.

(1) 计算 $P(B|A)$；

(2) 计算 $P(A\cap B)$.

解 (1) 四家各有 13 张牌，已知 A 发生后，A 的 13 张牌已固定. 余下的 39 张牌中恰有 7 张草花，在另三家中的分派是等可能的.

问题已经转变成：39 张牌中有 7 张草花，将这 39 张牌随机分给钱、孙、李三家，求孙家得到 3 张草花的概率. 于是

$$P(B|A)=\frac{C_7^3C_{39-7}^{10}}{C_{39}^{13}}\approx 0.278.$$

(2) 在 52 张牌中任选 13 张牌有 C_{52}^{13} 种不同的等可能的结果. 于是 Ω 中元素数 $=C_{52}^{13}$，A 中元素数 $=C_{13}^6C_{39}^7$. 利用条件概率公式得到

$$P(A\cap B)=P(A)P(B|A)=\frac{C_{13}^6C_{39}^7}{C_{52}^{13}}\times 0.278\approx 0.012.$$

练习

投掷两枚骰子.

(1) 已知一枚是偶数点，求另一枚也是偶数点的概率；

（2）已知两枚骰子的点数相同，求点数都是 3 的概率；

（3）已知点数和是 6，求两枚骰子点数相同的概率.

习题 3

学而时习之

打扑克的赵、钱、孙、李四家各从一副扑克的 52 张(去掉两张王牌后)中随机抽取 13 张，A=“赵家没得到 2”，B=“孙家得到 1 张 2”.（结果精确到小数点后三位）

（1）计算 $P(B|A)$；　　（2）计算 $P(A|B)$；

（3）计算 $P(A\cap B)$；　　（4）计算 $P(A\cup B)$.

8.2.3 事件的独立性

投掷一枚骰子和一枚硬币，骰子的点数和硬币是否正面朝上是独立的．这时我们称投掷一枚骰子的试验和投掷一枚硬币的试验是独立的.

用 Ω_1 表示第一个试验的全集，用 Ω_2 表示第二个试验的全集．如果这两个试验是独立的，就称全集 Ω_1 和 Ω_2 独立(independent).

理论和试验都证明了以下结论.

> **当事件的全集 Ω_1 和 Ω_2 独立，对于 $A\subseteq\Omega_1$ 和 $B\subseteq\Omega_2$，有**
>
> $$P(A\cap B)=P(A)P(B).$$

这时我们也称事件 A，B 独立.

若 A 与 B 独立，则 A 与 $\overline{B}$，$\overline{A}$ 与 B，$\overline{A}$ 与 $\overline{B}$ 也都独立.

例 1　投掷一枚骰子和一枚硬币，计算骰子出现 2 或 4 点，硬币正面朝上的概率.

解　用 A 表示骰子的点数是 2 或 4，用 B 表示硬币正面朝上，则 A，B 独立，且 $A\cap B$ 表示骰子出现 2 或 4 点，硬币正面朝上，因此，

有 $$P(A\cap B)=P(A)P(B)=\frac{2}{6}\times\frac{1}{2}=\frac{1}{6}.$$

例 2 同学甲的数学作业得优的概率是 0.8，同学乙的语文作业得优的概率为 0.7. 今天同时留了数学和语文作业，计算甲的数学得优、乙的语文没得优的概率.

解 用 A 表示甲的数学作业得优，用 B 表示乙的语文作业没有得优，则 $$P(A)=0.8,\quad P(B)=1-0.7=0.3.$$
$A\cap B$ 表示甲的数学作业得优、乙的语文没有得优. A，B 独立，所以
$$P(A\cap B)=P(A)P(B)=0.8\times0.3=0.24.$$

全班 45 名同学同时随机地翻开数学书，用 A_j 表示第 j 名同学翻开的左面页数在第 30 页或以下，则事件 A_1，A_2，…，A_{45} 是相互独立的.

对于 $j=1$，2，…，n，用 Ω_j 表示第 j 个试验的全集. 如果这 n 个试验是相互独立的，就称这些试验的全集 Ω_1，Ω_2，…，Ω_n 是相互独立的.

理论和试验都证明了以下结果.

如果试验的全集 Ω_1，Ω_2，…，Ω_n 是相互独立的，则对
$$A_1\subseteq\Omega_1,\ A_2\subseteq\Omega_2,\ \cdots,\ A_n\subseteq\Omega_n,$$
有
$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)P(A_2)\cdot\cdots\cdot P(A_n).$$

这时，我们也称事件 A_1，A_2，…，A_n 是相互独立的.

例 3 高中每个年级三个班的羽毛球水平相当，各年级举办班级羽毛球比赛时，计算都是三班得冠军的概率.

解 用 A_i 表示第 i 年级的三班获得冠军，则 A_1，A_2，A_3 相互独立，且 $P(A_1)=P(A_2)=P(A_3)=\frac{1}{3}$.

事件 $A=A_1\cap A_2\cap A_3$ 表示都是三班得冠军，由于 A_1，A_2，A_3 相互独立，所以
$$\begin{aligned}P(A)&=P(A_1\cap A_2\cap A_3)\\&=P(A_1)P(A_2)P(A_3)\end{aligned}$$

$$=\frac{1}{3^3}\approx 0.037.$$

例 4 一服装店出售标价为 180 元的夹克. 售货员对前来问价的顾客以 180 元推销成功的概率是 0.8. 如果一小时内先后有两位顾客前来问价，计算服务员对这两位顾客都没有推销成功的概率.

解 用 A_1，A_2 分别表示对第一、第二位顾客没有推销成功，则 A_1，A_2 独立. $A=A_1\cap A_2$ 表示对这两位顾客都没有推销成功. 利用

$$P(A_1)=P(A_2)=1-0.8=0.2$$

得到 $P(A)=P(A_1\cap A_2)=P(A_1)P(A_2)=0.2\times 0.2=0.04$.

例 5 李浩的棋艺不如张岚，李浩每局赢张岚的概率只有 0.45. 假设他们下棋时各局的输赢是独立的.

本例中不考虑心理因素的影响.

(1) 计算他们的 3 局棋中李浩至少赢 1 局的概率；

(2) 计算他们的 6 局棋中李浩至少赢 1 局的概率.

解 (1) 用 A_1，A_2，A_3 分别表示第1，第2，第3局李浩输，则 $A=A_1\cap A_2\cap A_3$ 表示李浩连输 3 局，其对立事件 $\overline{A}=\Omega\setminus A$ 表示李浩至少赢 1 局.

因为事件 A_1，A_2，A_3 相互独立，并且

$$P(A_1)=P(A_2)=P(A_3)=1-0.45=0.55,$$

所以 $$P(A)=P(A_1)P(A_2)P(A_3)=0.55^3\approx 0.166\ 4.$$

于是 $P(\overline{A})=1-P(A)\approx 0.833\ 6$. 说明 3 局棋中李浩至少赢 1 局的概率还是很大的.

这个例子体现了有志者事竟成的道理.

(2) 用 A_1，A_2，…，A_6 分别表示第1，第2，…，第6局李浩输，则 $A=A_1\cap A_2\cap\cdots\cap A_6$ 表示李浩连输 6 局，其对立事件 $\overline{A}=\Omega\setminus A$ 表示李浩至少赢 1 局.

因为事件 A_1，A_2，…，A_6 相互独立，并且

$$P(A_1)=P(A_2)=\cdots=P(A_6)=1-0.45=0.55,$$

所以

$$P(A)=P(A_1)P(A_2)\cdot\cdots\cdot P(A_6)=0.55^6\approx 0.027\ 7.$$

于是 $P(\overline{A})=1-P(A)\approx 0.972\ 3$.

说明 6 局棋中李浩至少赢 1 局的概率大于 0.97.

例 6 幸运抽奖活动中，中奖的比例是 1%. 计算：

（1）随机抽取 1 张，没中奖的概率 p；

（2）有放回地随机抽取 $n=100$ 张，没中奖的概率 p_n；

（3）有放回地随机抽取 $n=100$ 张，至少中奖 1 次的概率.

解 （1）用 A_1 表示第 1 次没有抽中，$p=P(A_1)=0.99$.

（2）用 A_1，A_2，…，A_{100} 分别表示第 1 次，第 2 次，…，第 100 次没有抽中，由于是有放回地随机抽样，所以 A_1，A_2，…，A_{100} 相互独立，并且

$$P(A_j)=0.99，j=1，2，\cdots，100.$$

$A=A_1\cap A_2\cap\cdots\cap A_{100}$ 表示 100 次都没有抽中，

$$\begin{aligned}P(A)&=P(A_1\cap A_2\cap\cdots\cap A_{100})\\&=P(A_1)P(A_2)\cdot\cdots\cdot P(A_{100})\\&=0.99^{100}\approx 0.366.\end{aligned}$$

（3）用 B 表示至少中奖 1 次，B 是 A 的对立事件.

$$P(B)=1-P(A)\approx 1-0.366\approx 0.634.$$

这个结果和你想象中的结果是否一样呢？

例 7 在某幸运抽奖活动中，每张奖券的中奖率为千分之一. 计算有放回地随机抽取 n 张奖券不能中奖的概率.

解 用 A_j 表示抽出的第 j 张奖券不能中奖.

对正整数 n，A_1，A_2，…，A_n 相互独立，并且

$$P(A_j)=1-P(\overline{A}_j)=1-\frac{1}{1\ 000}=0.999，j=1，2，\cdots，n.$$

$A=A_1\cap A_2\cap\cdots\cap A_n$ 表示购买 n 张奖券不能中奖.

$$\begin{aligned}P(A)&=P(A_1\cap A_2\cap\cdots\cap A_n)\\&=P(A_1)P(A_2)\cdot\cdots\cdot P(A_n)=0.999^n\end{aligned}$$

是抽出 n 张奖券不能中奖的概率.

由上面的公式，可以计算出抽出 50，100，500，1 000，2 000 张奖券不能中奖的概率 p_n.

n	50	100	500	1 000	2 000
p_n	0.95	0.905	0.606	0.368	0.135

可以看出，抽取 1 000 张奖券时，仍有 36.8%的概率不中奖.

例 8 设某试验成功的概率是 p，$p\in(0,1)$. 现在将该试验独立重复 3 次，证明：恰好有两次成功的概率为 $C_3^2p^2(1-p)$.

证 用 A_i 表示第 i 次试验成功，则 $B_1=A_1A_2\overline{A}_3$ 表示第 1，2 次试验成功，第 3 次不成功，且 $P(B_1)=p^2(1-p)$.

$B_2=A_1\overline{A}_2A_3$ 表示第 1，3 次试验成功，第 2 次不成功，且$P(B_2)=p^2(1-p)$.

$B_3=\overline{A}_1A_2A_3$ 表示第 2，3 次试验成功，第 1 次不成功，且 $P(B_3)=p^2(1-p)$.

$C=B_1\cup B_2\cup B_3$ 表示 3 次试验中恰有两次成功. 因为 B_1，B_2，B_3 两两互斥，所以

$$P(C)=P(B_1)+P(B_2)+P(B_3)=3p^2(1-p)=C_3^2p^2(1-p).$$

练 习

两门高炮同时向一架敌机射击，每门高炮击中飞机的概率都是 0.8. 计算：

(1) 飞机没被击中的概率；

(2) 飞机被击中一炮的概率；

(3) 飞机被击中两炮的概率.

习题 4

学而时习之

1. 公共汽车一共要停靠 9 站. 甲、乙两名互不相识的乘客在始发站上车，如果他们在每站下车的概率是相同的，计算：

(1) 甲在第 2 站下车、乙在第 3 站下车的概率；

(2) 甲、乙都在第 3 站下车的概率；

(3) 甲、乙同时在第 3 站或第 4 站下车的概率；

(4) 甲、乙在同一站下车的概率.

2. 假设每个人的生日在一年的 365 天中是等可能的. 在全校随机选取两名同学，计算以下事件的概率：

(1) 两个同学的生日都在 5 号；

(2) 一个的生日在 5 号，另一个的生日在 7 号.

3. 甲、乙二人进行羽毛球比赛，采取三局两胜的规则. 如果每局甲胜的概率是 0.6，计算：

(1) 两局结束时甲获胜的概率；

(2) 甲第一局输，第二局和第三局赢的概率；

(3) 甲第一局赢，第二局输，第三局赢的概率.

4. 在本教材 P. 58 例 7 中，若中奖的概率是万分之一. 计算抽出 1 000 张奖券不能中奖的概率和能中奖的概率.

5. 设某试验成功的概率是 p，$p\in(0，1)$. 现在将该试验独立重复 4 次，证明：恰好有 2 次成功的概率为 $C_4^2p^2(1-p)^2$.

8.2.4 离散型随机变量及其分布

一个字母 A 只能表示一个事件，许多问题中遇到的事件数目很大，都用字母表示时就显得很烦琐，于是人们引入了随机变量. 随机变量的引入大大地节省了符号的使用，也使得问题的表达更加简单明确.

如果用 X 表示明天的最高气温，$\{X=30\}$就表示明天的最高气温是30 ℃. 由于 X 的取值在今天无法确定，所以称 X 是**随机变量**(random variable). 随机变量常用字母 X，Y，ξ，η，…表示.

随机变量和函数都是一种映射，随机变量把随机试验的结果映为实数，函数把实数映为实数. 在这两种映射之间，试验结果的范围相当于函数的定义域，随机变量的取值范围相当于函数的值域.

例 1 投掷一枚骰子，用 X 表示掷出的点数，计算：

(1) $P(X=5)$； (2) $P(4\leqslant X\leqslant 5)$.

解 (1) X 是随机变量，$\{X=5\}$表示掷出的点数是 5，于是

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

(2) $\{4\leqslant X\leqslant 5\}$表示掷出的点数是 4 或 5，于是有

$$\{4\leqslant X\leqslant 5\}=\{X=4\}\cup\{X=5\}.$$

$$P(4\leqslant X\leqslant 5)=P(\{X=4\}\cup\{X=5\})$$

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

例 2 投掷两枚硬币，用 X 表示掷出的正面数，X 是随机变量. 计算：

(1) $P(X=0)$； (2) $P(X=1)$.

还有 $P(X=2)$ 这种情况，建议读者自己去算一算.

解 (1) $\{X=0\}$ 表示没有硬币正面朝上，$P(X=0)=\frac{1}{4}$.

(2) $\{X=1\}$ 表示只有一个硬币正面朝上，这个事件含有两个元素，所以 $P(X=1)=\frac{2}{4}=0.5$.

像上述例题那样，所有取值可以一一列出的随机变量，称为**离散型随机变量**(discrete random variable).

离散型随机变量的例子很多. 例如某人射击一次可能命中的环数 X 是一个离散型随机变量，它的所有可能取值为 0，1，…，10；投掷一枚骰子后上面出现的点数 Y 也是一个离散型随机变量，它的所有可能取值为 1，2，…，6. 但电灯泡的寿命 ξ 的可能取值是任何一个非负实数，而所有非负实数不能一一列出，所以 ξ 不是离散型随机变量.

如果离散型随机变量 X 的取值是 x_1，x_2，…，x_n，则 $\{X=x_i\}$ 是事件，用 $p_i=P(X=x_i)$ 表示事件 $\{X=x_i\}$ 的概率，我们称

$$p_i=P(X=x_i),\ i=1,\ 2,\ \cdots,\ n$$

是离散型随机变量 X 的**概率分布**.

以下用 $\{p_i\}$ 表示 X 的概率分布. 概率分布 $\{p_i\}$ 有如下的性质：

1. $p_i\geqslant 0$，$i=1$，2，…，n；

2. $p_1+p_2+\cdots+p_n=1$.

当离散型随机变量 X 的概率分布 $\{p_i\}$ 的规律性不够明显时，我们还用下面的表格来表示 X 的分布.

X	x_1	x_2	x_3	…
P	p_1	p_2	p_3	…

例 3 全班有 40 个同学，某次数学作业的成绩如下：

分数	0	1	2	3	4	5
人数	0	1	3	12	20	4
所占比例	0	0.025	0.075	0.3	0.5	0.1

从班中任选一个同学，用 X 表示这个同学的作业成绩，求 X 的概率分布.

解 $P(X=0)=0/40=0$，　$P(X=1)=1/40=0.025$，

$P(X=2)=3/40=0.075$，　$P(X=3)=12/40=0.3$，

$P(X=4)=20/40=0.5$，　$P(X=5)=4/40=0.1$.

因此，X 的概率分布列是

X	0	1	2	3	4	5
P	0	0.025	0.075	0.3	0.5	0.1

由上可以看出，随机变量 X 的分布就是该班作业成绩的分布.

例 4 设随机变量 X 的分布列为 $P(X=k)=\frac{c}{k(k+1)}$，$k=1$，2，3，4，其中 c 为常数，求 $P\left(\frac{1}{2}<X<\frac{5}{2}\right)$的值.

解 由离散型随机变量的概率分布的性质可知 $\frac{c}{1\times2}+\frac{c}{2\times3}+\frac{c}{3\times4}+\frac{c}{4\times5}=1$，所以 $c\left(1-\frac{1}{2}+\frac{1}{2}-\frac{1}{3}+\frac{1}{3}-\frac{1}{4}+\frac{1}{4}-\frac{1}{5}\right)=1$，

所以 $c=\frac{5}{4}$.

因此，$P\left(\frac{1}{2}<X<\frac{5}{2}\right)=P(X=1)+P(X=2)$

$$=\frac{\frac{5}{4}}{1\times2}+\frac{\frac{5}{4}}{2\times3}=\frac{5}{6}.$$

练　习

1. 投掷一枚骰子，用 X 表示掷出的点数，对 $k=1$，2，…，6 计算 $P(X=k)$.

2. 离散型随机变量 X 的概率分布列如右表，则 p 的值为（　　）

X	0	1	2	4
P	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	p

A. $\frac{1}{2}$　　B. $\frac{1}{6}$　　C. $\frac{1}{3}$　　D. $\frac{1}{4}$

习题 5

学而时习之

1. 投掷两枚骰子，用 X 表示掷出的点数之和，计算 $P(X=3)$ 和 $P(X=10)$.

2. 设随机变量 X 的分布列为 $P\left(X=\frac{k}{5}\right)=ak$ （$k=1$，2，3，4，5），a 为常数.

（1）求常数 a 的值；

（2）求 $P\left(\frac{1}{10}<X<\frac{7}{10}\right)$ 的值.

3. 袋中有 4 只红球、3 只黑球，从袋中任取 4 只球，取到 1 只红球得 1 分，取到 1 只黑球得 3 分，设得分为随机变量 X，求 $P(X\leqslant 6)$ 的值.

4. 在打靶时，选手甲每次击中目标的概率是 p. 如果他独立重复射击 4 次，用 X 表示他击中目标的次数，计算 X 的分布.

8.2.5　几个常用的分布

1. 两点分布 $B(1, p)$.

如果 X 只取值 0 或 1，概率分布是

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

就称 X 服从两点分布，记作 $X\sim B(1, p)$.

任何试验，当只考虑成功与否时，就可以用服从两点分布的随机变量描述：

两点分布又称 0—1 分布．由于只有两个可能结果的随机试验叫伯努利(Bernoulli)试验，所以还称这种分布为伯努利分布.

$$X=\begin{cases}1，当试验成功；\\0，当试验不成功.\end{cases}$$

例 1 某试验成功的概率是 p，将该试验独立重复 4 次，用 X 表示 4 次试验中的成功次数，计算 $P(X=3)$.

解 用 A_j 表示第 j 次试验成功，则 $\overline{A}_j$ 表示第 j 次试验不成功. 事件 A_1，A_2，A_3，A_4 相互独立，而且 $P(A_j)=p$，$P(\overline{A}_j)=1-p$.

$B_1=\overline{A}_1\cap A_2\cap A_3\cap A_4$ 表示后 3 次试验成功，第 1 次试验失败，则

$$P(B_1)=P(\overline{A}_1)P(A_2)P(A_3)P(A_4)=p^3(1-p).$$

用 B_j 表示第 j 次试验失败，其余 3 次试验成功，同样有

$$P(B_j)=p^3(1-p)，j=1，2，3，4.$$

现在 $\{X=3\}=B_1\cup B_2\cup B_3\cup B_4$.

因为 B_1，B_2，B_3，B_4 两两互斥，所以

$$\begin{aligned}P(X=3)&=P(B_1\cup B_2\cup B_3\cup B_4)\\&=P(B_1)+P(B_2)+P(B_3)+P(B_4)\\&=4p^3(1-p)\\&=\mathrm{C}_4^3p^3(1-p)^{4-3}.\end{aligned}$$

完全类似地可以计算出

$$P(X=k)=\mathrm{C}_4^kp^k(1-p)^{4-k}，k=0，1，2，3，4.$$

上面的例子可以推广到 n 次独立重复试验的情况.

2. 二项分布 $B(n，p)$.

设某试验成功的概率为 p，$p\in(0，1)$. 将该试验独立重复 n 次，用 X 表示成功的次数，则 X 有概率分布：

$$P(X=k)=\mathrm{C}_n^kp^kq^{n-k}，k=0，1，\cdots，n，其中\ q=1-p，$$

这时，我们称 X 服从**二项分布**，记作 $X\sim B(n，p)$.

这里的 B 是 binomial（二项式）的缩写.

这里，称为二项分布的原因是 $\mathrm{C}_n^kp^kq^{n-k}$ 为二项展开式

$$(p+q)^n=\mathrm{C}_n^0p^0q^n+\mathrm{C}_n^1p^1q^{n-1}+\cdots+\mathrm{C}_n^np^n$$

的第 $k+1$ 项.

例 2 甲每次投资获利的概率是 $p=0.8$. 对他进行 6 次相互独立的投资，计算：

(1) 有 5 次获利的概率；

（2）有 6 次获利的概率；

（3）至少 5 次获利的概率.

解 用 X 表示甲在 6 次投资中获利的次数，X 服从二项分布 $B(6,0.8)$.

$$P(X=5)=C_6^5 0.8^5(1-0.8)\approx 0.39.$$

$$P(X=6)=C_6^6 0.8^6\approx 0.26.$$

（1）他 5 次获利的概率约等于 0.39.

（2）6 次都获利的概率约等于 0.26.

（3）$\{X\geqslant 5\}$表示他至少 5 次获利，即$\{X\geqslant 5\}=\{X=5\}\cup\{X=6\}$. 由于事件$\{X=5\}$和$\{X=6\}$互斥，所以

$$P(X\geqslant 5)=P(X=5)+P(X=6)\approx 0.39+0.26=0.65.$$

因此，至少 5 次获利的概率约等于 0.65.

例 3 某家庭装修公司和客户洽谈装修协议时，洽谈成功的概率是 0.4. 设一天内有 9 个客户前来洽谈装修协议. 用 X 表示这天洽谈成功的客户数，求 $P(X=5)$.

解 X 服从二项分布 $B(9,0.4)$，于是

$$P(X=5)=C_9^5 0.4^5(1-0.4)^{9-5}\approx 0.167.$$

因此，洽谈成功 5 个客户的概率约等于 0.167.

3. 超几何分布.

一般地，在含有 M 件次品的 N 件产品中，任取 n 件，其中恰有 X 件次品，则事件$\{X=k\}$发生的概率为

$$P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n},\ k=0,1,2,\cdots,m,$$

其中 $m=\min\{M,n\}$，且 $n\leqslant N$，$M\leqslant N$，n，M，$N\in\mathbf{N}^*$.

称分布列

X	0	1	…	m
P	$\frac{C_M^0 C_{N-M}^{n-0}}{C_N^n}$	$\frac{C_M^1 C_{N-M}^{n-1}}{C_N^n}$	…	$\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$

为**超几何分布列**. 如果随机变量 X 的分布列为超几何分布列，就称 X 服从**超几何分布**，记作 $X\sim H(N,M,n)$.

例 4 鱼塘中只有 80 条鲤鱼和 20 条草鱼，每条鱼被打捞的可能性相同. 捞鱼者一网打捞上来 4 条鱼，计算：

（1）其中有 1 条鲤鱼的概率；

（2）其中有 2 条鲤鱼的概率；

（3）其中有 3 条鲤鱼的概率；

（4）4 条都是鲤鱼的概率.

解 用 X 表示被打捞的 4 条鱼中的鲤鱼条数，X 服从超几何分布.

（1）$P(X=1)=\dfrac{C_{80}^{1}C_{20}^{3}}{C_{100}^{4}}\approx 0.0233.$

（2）$P(X=2)=\dfrac{C_{80}^{2}C_{20}^{2}}{C_{100}^{4}}\approx 0.1531.$

（3）$P(X=3)=\dfrac{C_{80}^{3}C_{20}^{1}}{C_{100}^{4}}\approx 0.4191.$

（4）$P(X=4)=\dfrac{C_{80}^{4}C_{20}^{0}}{C_{100}^{4}}\approx 0.4033.$

结论表明打捞到多条鲤鱼的概率要大一些，原因是鲤鱼的数目多于草鱼的数目.

例 5 在某班的春节联欢活动中，组织了一次幸运抽奖活动. 袋中装有 18 个除颜色外质地相同的小球，其中 8 个是红球，10 个是白球. 抽奖者从中一次抽出 3 个小球，抽到 3 个红球得一等奖，2 个红球得二等奖，1 个红球得三等奖，0 个红球不得奖. 分别计算得到一等奖、二等奖和三等奖的概率.

解 从 18 个小球中抽取 3 个时，有 C_{18}^{3} 种不同的等可能结果，这是元素的总数. 用 X 表示抽到的红球数，则 X 服从超几何分布，并且

$$P(\text{得一等奖})=P(X=3)=\frac{C_{8}^{3}}{C_{18}^{3}}=\frac{56}{816}\approx 0.0686.$$

$$P(\text{得二等奖})=P(X=2)=\frac{C_{8}^{2}C_{10}^{1}}{C_{18}^{3}}=\frac{280}{816}\approx 0.3431.$$

$$P(\text{得三等奖})=P(X=1)=\frac{C_{8}^{1}C_{10}^{2}}{C_{18}^{3}}=\frac{360}{816}\approx 0.4412.$$

从中看出，得三等奖的概率最大.

练习

1. 一副眼镜不慎落地被摔坏的概率是 0.6，计算：

   (1) 第 1 次落地被摔坏的概率；

   (2) 第 2 次落地被摔坏的概率；

   (3) 5 次落地还没摔坏的概率.

2. 在某年级的联欢会上设计了一个摸奖游戏，在一个口袋中装有 10 个红球和 20 个白球，这些球除颜色外完全相同. 一次从中摸出 5 个球，至少摸到 3 个红球就中奖. 求中奖的概率.

习题 6

学而时习之

1. 对一大批产品的验收方案如下：从中任取 10 件检验，无次品就接收这批产品. 设产品的次品率是 10%，计算产品被接收的概率.

2. 某收藏家在拍卖会上决定参加对 5 件艺术品的竞买，各拍品是否竞买成功是相互独立的. 如果他成功购得 1 件艺术品的概率是 0.2，计算：

   (1) 成功竞买 2 件的概率；

   (2) 成功竞买 5 件的概率；

   (3) 至少竞买 1 件成功的概率.

3. 年级学生会改选时要求各班先选出 5 名候选人. 一班有 23 名男生，20 名女生，如果每个人被选成候选人的概率相同，计算一班的候选人中有 3 名是女生的概率.

4. 高三(2)班有 43 名学生. 昨天上语文课时，张老师叫到了其中的 9 名同学回答问题，今天的语文课张老师又要叫 9 名同学回答问题. 如果今天每个人被叫到的可能性相同，计算昨天回答问题的学生中有 3 名又被叫到的概率.

数学文化

数学期望

惠更斯是一个名声和牛顿相当的大科学家．人们熟知他的贡献之一是物理中的单摆公式．他在概率论的早期发展历史上也占有重要的地位．他的主要著作《机遇的规律》在1657年出版．在这部著作中，他首先引进了“期望”这个术语，基于这个术语解决了一些当时感兴趣的博弈问题．他在这部著作中提出了14条命题，第一条命题是：

如果某人在赌博中以概率1/2赢a元，以概率1/2输b元，则他的期望是

$$\frac{a-b}{2}.$$

下面我们看期望是如何定义的．

8.2.6 离散型随机变量的数学期望

为说明离散型随机变量的数学期望，先看一个例子.

全年级有 $n=300$ 个学生，其中有 n_i 个同学的身高是 x_i cm：

身高 x_i(cm)	156	157	158	…	184	185
人数 n_i	n_1	n_2	n_3	…	n_{29}	n_{30}
比例 f_i	n_1/n	n_2/n	n_3/n	…	n_{29}/n	n_{30}/n

这里共有 $m=185-156+1=30$ 个不同的身高，$n=n_1+n_2+\cdots+n_m=300$. 全体同学的身高之和是 $x_1n_1+x_2n_2+\cdots+x_mn_m$.

从班中任选一个同学，用 X 表示这个同学的身高，则 X 有概率分布：

$$P(X=156)=\frac{n_1}{n},\ P(X=157)=\frac{n_2}{n},\ \cdots,\ P(X=185)=\frac{n_{30}}{n}.$$

将概率分布列表，得到

X	156	157	158	…	184	185
P	n_1/n	n_2/n	n_3/n	…	n_{29}/n	n_{30}/n

可以看出，随机变量 X 的分布就是全年级同学身高的分布：

$$p_i=P(X=x_i)=\frac{n_i}{n},\ i=1,\ 2,\ \cdots,\ m.$$

全年级同学的平均身高是

$$\mu=\frac{1}{n}\ (x_1n_1+x_2n_2+\cdots+x_mn_m).$$

由于 X 的分布是全年级同学身高的分布，我们把全年级的平均身高 μ 定义成 X 的均值，记作 $E(X)$. 于是，

$$\begin{aligned}E(X)&=x_1\frac{n_1}{n}+x_2\frac{n_2}{n}+\cdots+x_m\frac{n_m}{n}\\&=x_1p_1+x_2p_2+\cdots+x_mp_m.\end{aligned}$$

再看一个例子，设离散型随机变量 X 有概率分布

X	1	100
P	0.01	0.99

作为 X 的可能值的平均数，$\frac{1}{2}(1+100)=50.5$ 并不能真正体现

X 的取值的平均. 因为 X 取值 100 的概率比取值 1 的概率大得多，所以应当用 $1\times0.01+100\times0.99=99.01$ 表示 X 的平均取值.

定义 当离散型随机变量 X 有概率分布

$$p_j=P(X=x_j),\ j=0,\ 1,\ \cdots,\ n,$$

就称

$$E(X)=x_1p_1+x_2p_2+\cdots+x_np_n$$

为 X 的**数学期望**(mathematical expectation)或**均值**(mean).

从前面的例子知道，如果 X 是从某个总体中随机抽取的个体，X 的数学期望 $E(X)$ 就是总体均值 μ.

为什么称为数学期望呢？设想总体中只有三个个体，一个 2，一个 6，一个 12. 从中任取一个，取到几得几分. 用 X 表示取到的分数，你对 X 的期望是多少呢？你对 X 的期望就是总体平均，也就是 $E(X)$. 于是，$E(X)$ 是你对得分的期望.

例 1 甲击中目标的概率是 $\frac{1}{2}$，如果击中，赢 10 分，否则输 11 分. 用 X 表示他的得分，计算 X 的概率分布和数学期望.

解 $\{X=10\}$ 的充分必要条件是击中目标，所以

$$P(X=10)=\frac{1}{2}=0.5.$$

$\{X=-11\}$ 是 $\{X=10\}$ 的对立事件，所以

$$P(X=-11)=1-P(X=10)=0.5.$$

X 只取值 10 和 -11，于是

$$\begin{aligned}E(X)&=10\times P(X=10)+(-11)\times P(X=-11)\\&=10\times0.5-11\times0.5=-0.5.\end{aligned}$$

于是甲平均输 0.5 分. 我们也说，甲只能期望赢 -0.5 分.

例 2 在只需回答“是”与“不是”的知识竞赛中，每个选手回答两个不同的问题，都回答失败，输 1 分，否则赢 0.3 分. 用 X 表示甲的得分，如果甲随机猜测“是”与“不是”，计算 X 的概率分布和数学期望.

解 $\{X=-1\}$ 的充分必要条件是两次猜错. 所以

$$P(X=-1)=\frac{1}{4}=0.25.$$

$\{X=0.3\}$是$\{X=-1\}$的对立事件，所以

$$P(X=0.3)=1-P(X=-1)=0.75.$$

X 只取值-1 和 0.3，于是

$$\begin{aligned}E(X)&=(-1)\times P(X=-1)+0.3\times P(X=0.3)\\&=-0.25+0.3\times 0.75=-0.025.\end{aligned}$$

我们也说甲平均输 0.025 分，或说甲只能期望赢-0.025 分.

定理 关于离散型随机变量 X 的数学期望，有以下公式：

(1) 若 $Y=aX+b$，a，b 为常数，则 $E(aX+b)=aE(X)+b$；

(2) 当 X 服从两点分布 $B(1, p)$时，$E(X)=p$；

(3) 当 X 服从二项分布 $B(n, p)$时，$E(X)=np$；

(4) 当 X 服从超几何分布 $H(N, M, n)$ 时，$E(X)=n\dfrac{M}{N}$.

证 我们只证明公式(2)和(3).

(2) $E(X)=1\times p+0\times(1-p)=p$.

(3) 当 $j=0$，1，2，…，n 时，

$$jC_n^j=j\frac{n!}{j!\ (n-j)!}=nC_{n-1}^{j-1}.$$

设 $q=1-p$，$m=n-1$，有

$$\begin{aligned}E(X)&=0+C_n^1p^1q^{n-1}+2C_n^2p^2q^{n-2}+\cdots+nC_n^np^nq^0\\&=npC_{n-1}^0p^0q^{n-1}+npC_{n-1}^1p^1q^{n-2}+\cdots+npC_{n-1}^{n-1}p^{n-1}q^0\\&=np[C_m^0p^0q^m+C_m^1p^1q^{m-1}+\cdots+C_m^mp^mq^0]\\&=np(p+q)^m\\&=np.\end{aligned}$$

设一试验成功的概率为 p. 将该试验独立重复 n 次，用 X 表示成功的次数，则 X 有数学期望 np. $E(X)=np$ 是 n 次独立重复试验中平均成功的次数. $E(X)=np$ 说明平均成功的次数和 n 成正比，也和 p 成正比. 单次试验成功的概率越大，n 次独立重复试验中成功的平均次数就越多.

例 3 甲、乙比赛时，甲每局赢的概率是 $p=0.51$，乙每局赢的

概率是 $q=0.49$. 甲、乙一共进行了 10 局比赛. 当各局比赛的结果是相互独立的时，计算甲平均赢多少局，乙平均赢多少局.

解 用 X 表示 10 局中甲赢的局数，则 X 服从二项分布 $B(10, 0.51)$，因此，$E(X)=10\times0.51=5.1$，即甲平均赢 5.1 局.

用 Y 表示 10 局中乙赢的局数，则 Y 服从二项分布 $B(10, 0.49)$. 因此，$E(Y)=10\times0.49=4.9$，于是乙平均赢 4.9 局.

例 4 袋中有 3 个红球，7 个白球. 从中无放回地任取 5 个，取到几个红球就得几分. 问平均得几分？

解 用 X 表示得分数，则 X 也是取到的红球个数. X 服从超几何分布 $H(10, 3, 5)$，于是，$E(X)=n\times\dfrac{M}{N}=5\times\dfrac{3}{10}=1.5$，即平均得到 1.5 分.

练习

1. 投掷 6 枚骰子，用 Z 表示 6 朝上的骰子个数，求 $E(Z)$.
2. 一次单元测验由 20 道选择题构成，每道选择题有 4 个选项，其中仅有一个选项正确. 每道题选对得 5 分，不选或选错不得分，满分 100 分. 学生甲选对任意一题的概率为 0.9，学生乙则在测验中对每道题都从各选项中随机地选择一个. 分别求学生甲和学生乙在这次测验中的成绩的均值.

习题 7

学而时习之

1. 一条大鱼有 k 条后代小鱼的概率是 $p_k=C_n^k p^k(1-p)^{n-k}$，$k=0, 1, \cdots, n$，这条大鱼平均有多少条后代小鱼？
2. 投掷一枚骰子，掷出点数 j 时就得 j 分，期望得到几分？

8.2.7 离散型随机变量的方差

在上节引入介绍数学期望时，我们已经知道，全年级 300 个同学的身高是一个总体，总体均值是

$$\mu=\frac{1}{n}(x_1n_1+x_2n_2+\cdots+x_mn_m)，m=30.$$

总体方差是 $\sigma^2=\frac{1}{n}[(x_1-\mu)^2n_1+(x_2-\mu)^2n_2+\cdots+(x_m-\mu)^2n_m]$

$$=(x_1-\mu)^2p_1+(x_2-\mu)^2p_2+\cdots+(x_m-\mu)^2p_m.$$

从班中任选一个同学，用 X 表示这个同学的身高，则

$$p_i=P(X=x_i)=\frac{n_i}{n}，i=1，2，\cdots，m.$$

我们已经把全年级的平均身高定义成 X 的数学期望 $E(X)$，我们再把全年级身高的方差 σ^2 定义成随机变量 X 的方差，用 $D(X)$表示. 即有

$$D(X)=(x_1-\mu)^2p_1+(x_2-\mu)^2p_2+\cdots+(x_m-\mu)^2p_m.$$

于是引出离散型随机变量的方差的定义.

定义 当离散型随机变量 X 有概率分布

$$p_j=P(X=x_j)，j=0，1，\cdots，n$$

和数学期望 $\mu=E(X)$时，就称

$$D(X)=(x_1-\mu)^2p_1+(x_2-\mu)^2p_2+\cdots+(x_n-\mu)^2p_n$$

为 X 的**方差**(variance)，称$\sqrt{D(X)}$为 X 的**标准差**(standard deviation). 通常还用 σ^2 表示方差 $D(X)$，用 σ 表示标准差$\sqrt{D(X)}$.

X 的方差描述了随机变量 X 向它的数学期望集中的程度，方差越小，X 向数学期望 μ 集中得越好.

如果 X 是从某个总体中通过随机抽样得到的个体，X 的方差 $D(X)$就是总体方差 σ^2，X 的数学期望 $E(X)$就是总体均值 μ.

例 1 根据以往经验，一辆从北京开往天津的长途汽车在无雨天赢利 230 元，小雨天赢利 163 元，中雨天赢利 90 元. 根据天气预报，明天无雨的概率是 0.2，有小雨的概率是 0.3，有中雨的概率是 0.5. 问明天发一辆长途车期望赢利多少元？方差和标准差各是多少？

解 用 X 表示明天发一辆车的赢利. $\{X=230\}$发生的充分必要条件是明天无雨，$\{X=163\}$发生的充分必要条件是明天有小雨，$\{X=90\}$发生的充分必要条件是明天有中雨，于是，

$$P(X=230)=0.2,\ P(X=163)=0.3,\ P(X=90)=0.5.$$

$$E(X)=230\times0.2+163\times0.3+90\times0.5=139.9(\text{元}).$$

于是期望赢利 139.9 元，或说发一辆车平均赢利 139.9 元.

$$\begin{aligned}\text{方差 } D(X)&=(230-139.9)^2\times0.2+(163-139.9)^2\times\\&\quad 0.3+(90-139.9)^2\times0.5\\&=3\ 028.69.\end{aligned}$$

标准差 $\sigma=\sqrt{D(X)}=\sqrt{3\ 028.69}\approx55.$

用计算器计算得出.

定理 对于离散型随机变量 X，有以下的方差计算公式：

(1) 若 $Y=aX+b$，a，b 为常数，即 $D(aX+b)=a^2D(X)$；

(2) 当 X 服从两点分布 $B(1, p)$时，$D(X)=p(1-p)$；

(3) 当 X 服从二项分布 $B(n, p)$时，$D(X)=np(1-p)$；

(4) 当 X 服从超几何分布 $H(N, M, n)$时，

$$D(X)=\frac{nM}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}.$$

我们只给出两点分布情况下的证明.

证 因为$E(X)=p$,所以

$$\begin{aligned}D(X)&=(1-p)^2p+(0-p)^2(1-p)\\&=(1-p)(p-p^2+p^2)=p(1-p).\end{aligned}$$

例 2 某厂一批产品的合格率是 98%，检验单位从中有放回地随机抽取 10 件，计算：

(1) 抽出的 10 件产品中平均有多少件正品；

(2) 计算抽出的 10 件产品中正品数的方差和标准差.

解 用 X 表示抽得的正品数，由于是有放回地随机抽样，所以 X 服从二项分布 $B(10, 0.98)$.

(1) 利用二项分布的期望公式得到 $E(X)=10\times0.98=9.8$，因此平均有 9.8 件正品.

(2) X 的方差 $D(X)=10\times0.98\times0.02=0.196$，标准差 $\sigma=$

$\sqrt{D(X)}\approx 0.44$.

例 3 100 箱苹果中有 5 箱不合格，现在从中随机抽取 5 箱检查，计算：

(1) 抽出的 5 箱中平均有多少箱合格；

(2) 计算抽出的 5 箱中合格箱数的方差和标准差.

解 用 X 表示抽到的 5 箱中的合格箱数，则 X 服从超几何分布 $H(N, M, n)$，其中 $N=100$，$M=95$，$n=5$.

(1) $E(X)=5\times\frac{95}{100}=4.75$，因此平均有 4.75 箱合格.

(2) 利用超几何分布的方差计算公式得到

$$D(X)=5\times\frac{95}{100}\times\left(1-\frac{95}{100}\right)\times\frac{100-5}{100-1}\approx 0.228,$$

$$\sigma=\sqrt{D(X)}\approx 0.48.$$

例 4 甲、乙两名射手在同一条件下射击，所得环数 X_1，X_2 的分布列分别是

X_1	6	7	8	9	10
P_1	0.16	0.14	0.42	0.1	0.18

X_2	6	7	8	9	10
P_2	0.19	0.24	0.12	0.28	0.17

请根据环数的期望和方差比较这两名射手的射击水平.

解 $E(X_1)=6\times0.16+7\times0.14+8\times0.42+9\times0.1+10\times0.18$

$=8$,

$D(X_1)=(6-8)^2\times0.16+(7-8)^2\times0.14+(8-8)^2\times0.42+(9-8)^2\times0.1+(10-8)^2\times0.18$

$=1.6$;

$E(X_2)=6\times0.19+7\times0.24+8\times0.12+9\times0.28+10\times0.17=8$,

$D(X_2)=(6-8)^2\times0.19+(7-8)^2\times0.24+(8-8)^2\times0.12+(9-8)^2\times0.28+(10-8)^2\times0.17$

$=1.96$.

$\because$ $E(X_1)=E(X_2)$，$D(X_1)<D(X_2)$，

$\therefore$ 射手甲与射手乙的平均射击水平没有差异，但射手甲的射击

水平稳定性较好，射手乙的射击水平稳定性较差.

练习

1. 随机抛掷一枚质地均匀的骰子，求向上一面的点数 X 的均值、方差和标准差.
2. 甲每次投资获利的概率为 0.8，用 Z 表示甲在 10 次相互独立的投资中获利的次数，计算 $E(Z)$和 $D(Z)$.

习题 8

学而时习之

1. 已知随机变量 X 的分布列如右表，求 $E(X)$，$E(2X+5)$，$D(X)$，$D(2X+5)$.

X	-2	1	3
P	0.16	0.44	0.40

2. 对一个新产品的开发需要投资 10 万元，开发成功可以获利 1 000 万元. 如果开发成功的概率是 0.7，计算投资的平均收益和标准差.
3. 10 万张体育彩票中只有 2 个大奖，每个大奖 3 万元. 每张彩票 1 元钱，购买 1 张，平均赢利多少元？标准差是多少元？
4. 在某公司的一次投标工作中，中标可以获利 10 万元，没有中标损失成本费 0.05 万元. 如果中标的概率是 0.4，计算：

(1) 公司的平均赢利 μ；(2) 公司赢利的方差 $D(X)$；(3) 公司赢利的标准差.

5. 有某地的甲、乙两个单位都愿意聘用你，而你能获得如下信息：

甲单位不同职位月工资 X_1(元)	1 200	1 400	1 600	1 800
获得相应职位的概率 P_1	0.4	0.3	0.2	0.1

乙单位不同职位月工资 X_2(元)	1 000	1 400	1 800	2 000
获得相应职位的概率 P_2	0.4	0.3	0.2	0.1

根据工资待遇的差异情况，你愿意选择哪家单位？

数学文化

高斯与正态分布

伟大的天文学家伽利略(Galileo，1564—1642)可能是第一个提出随机误差概念的人．他在1632年出版的《关于托勒密和哥白尼两个世界系统的对话》中提到了观测误差，并谈到了观测误差的以下性质：

(1) 所有的观测都可以有误差，其来源可能归因于观测者、观测仪器和观测条件；

(2) 观测误差对称地分布在0的两侧；

(3) 小误差比大误差出现得更频繁．

这里的观测误差实际上是现在我们所说的随机误差．

1809年，高斯(Gauss，1777—1855)发表了天体力学的名著《绕日天体运动理论》，在这部著作的末尾，他写了一节有关数据组合的问题，实际上涉及的就是随机误差分布的问题．高斯在以后的研究工作中发现了正态分布．这一发现意义重大，也使正态分布有了高斯分布的名字．

高斯是一个伟大的数学家，一生中的重要贡献不胜枚举．德国在加入欧元区之前，流通使用的10马克纸币上印有高斯的头像和正态分布密度曲线，这就传达了一个信息：在高斯的科学贡献中，对人类文明影响最大者，正态分布也．

8.3 正态分布曲线

很早以前，人们不知道圆周率(我们也假设 π 是未知的)，为了研究圆的直径和周长的关系，需要对圆的周长进行测量. 在测量前，我们用 X 表示测量值，则 X 是随机变量.

由于各种随机因素的存在，例如测量本身的随机误差等，使得对直径相同的圆的测量会得到不同的结果. 于是，进行大量的独立重复测量后就得到了大量的测量值. 理论和试验证明这些测量值的样本方差会稳定在一个固定的数值 σ^2 附近.

我们把 $\sigma=\sqrt{D(X)}$ 叫作随机变量 X 的标准差，或测量标准差. 在一些实际问题中，标准差 σ 是已知的. 当标准差未知时，可以用多次测量值的样本标准差近似.

根据标准差的性质知道，σ 越小表示测量的精度越高，σ 越大表示测量精度越差，所以 σ 表示的是测量的精度.

例如，对直径为 1 cm 的圆的周长进行测量，由于多种偶然因素的影响，测量出的数据是有差异的. 若记 X 为测量出的数据，则 X 是一个随机变量. 实际问题中需要关心 X 取值的概率分布. 为了确定 X 的概率分布，我们记录了 90 次测量数据(样本数)，把它们进行分组整理后得如下分组数据表：

组号	组　　限	组频数	组频率	组频率/组距
1	[3.115，3.120)	4	4/90	8.888
2	[3.120，3.125)	6	6/90	13.332
3	[3.125，3.130)	8	8/90	17.776
4	[3.130，3.135)	11	11/90	24.444
5	[3.135，3.140)	14	14/90	31.11
6	[3.140，3.145)	15	15/90	33.332
7	[3.145，3.150)	12	12/90	26.666
8	[3.150，3.155)	10	10/90	22.222
9	[3.155，3.160)	6	6/90	13.332
10	[3.160，3.165)	4	4/90	8.888

以测量出的数据为横坐标，以组频数/组距为纵坐标，就可以得到频率直方图（如图 8－1）.

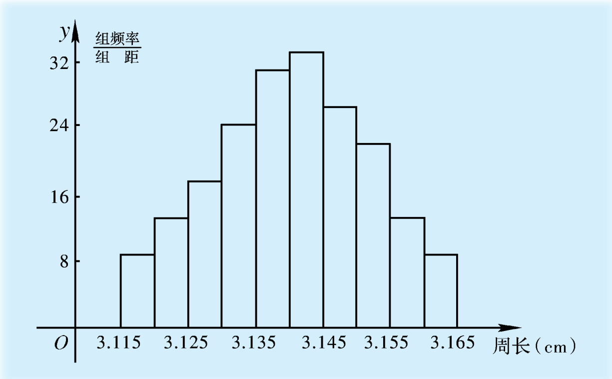


图 8－1

由图 8－1 可以看出，上述数据的分布有“中间高，两边低，左、右大致对称”的特点. 当样本数 n 越来越大，分组数越来越多(即组距无限缩小)时，频率直方图的顶边会无限缩小乃至形成一条光滑的曲线，且可用如下函数

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\quad(-\infty<x<+\infty)$$

的图象(如图 8－2)来近似表示. 其中 μ 和 σ 为常数，且 $\sigma>0$，$\mathrm{e}=2.718\ 28\cdots$，$\pi=3.141\ 59\cdots$. 此时，我们称 $p(x)$ 的图象为**正态分布密度曲线**，简称**正态曲线**.

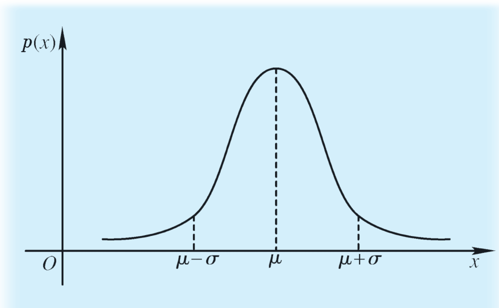


图 8－2

不同的 μ 和 σ 对应着不同的正态分布密度曲线(图 8－3).

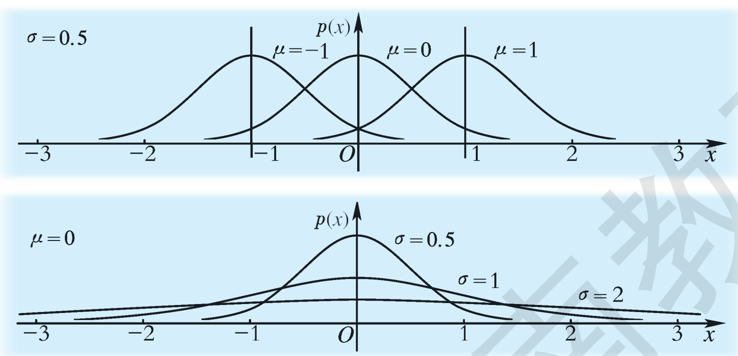


图 8－3

从图 8－2 与图 8－3 可以看出正态分布密度曲线具有如下特点：

1. 曲线位于 x 轴上方，与 x 轴不相交；
2. 曲线是单峰的，它关于直线 $x=\mu$ 对称；
3. $p(x)$在 $x=\mu$ 处达到最大值$\frac{1}{\sqrt{2\pi}\sigma}$；
4. 当 σ 一定时，曲线随着 μ 的变化而沿 x 轴平移；
5. 当 μ 一定时，σ 越大，正态曲线越扁平；σ 越小，正态曲线越尖陡；
6. 曲线与 x 轴之间所夹的面积等于 1.

概率的性质.

随机变量 X 落在区间 $(x_1, x_2]$（其中 $x_1<x_2$）中的概率可以通过函数 $p(x)$来描述，即 $P(x_1<X\leqslant x_2)$恰好是由 $p(x)$对应曲线，过点$(x_1, 0)$，点$(x_2, 0)$的两条 x 轴的垂线及 x 轴所围成的曲边梯形的面积（图 8－4），即 $P(x_1<X\leqslant x_2)=\int_{x_1}^{x_2}p(x)\mathrm{d}x=F(x_2)-F(x_1)$.

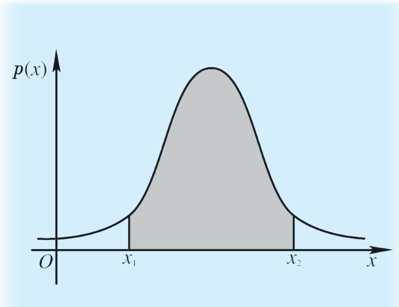


图 8－4

此时，我们称随机变量 X 为服从参数为 μ 和 σ^2 的**正态分布**(normal distribution)，简记为 $X\sim N(\mu, \sigma^2)$.

特别当 $\mu=0$，$\sigma^2=1$ 时称为**标准正态分布**（standardized normal distribution)，其密度函数记为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\quad(-\infty<x<+\infty),$$

其图象如图 8－5 所示，简记为 $X\sim N(0, 1)$，其分布函数记为 $\Phi(x)$.

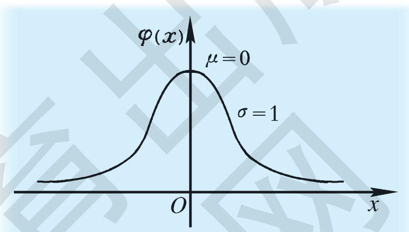


图 8－5

现实世界中的很多随机变量遵循正态分布．如反复测量长度时，其测量误差通常被认为服从正态分布；某一地区同性别同年龄组儿童的身高、体重等被近似地认为服从正态分布；某地每年某月份的平均气温、平均湿度等也被近似地认为服从正态分布．所以，正态分布广泛存在于自然现象、生产和生活实际之中，正态分布在概率和统计中占有重要

的地位.

若 $X \sim N(\mu, \sigma^2)$，则随机变量 X 在 μ 的附近取值的概率较大，在离 μ 很远处取值的概率较小.

具体地说随机变量 X 取值

落在区间 $(\mu-\sigma, \mu+\sigma)$ 内的概率约为 68.3%，

落在区间 $(\mu-2\sigma, \mu+2\sigma)$ 内的概率约为 95.4%，

落在区间 $(\mu-3\sigma, \mu+3\sigma)$ 内的概率约为 99.7%.

上述结果可用图 8－6 表示.

由图 8－6 可以看出，正态总体几乎总取值于区间 $(\mu-3\sigma, \mu+3\sigma)$ 之内. 而在此区间以外取值的概率只有 0.003，通常认为这种情况在一次试验中几乎不可能发生.

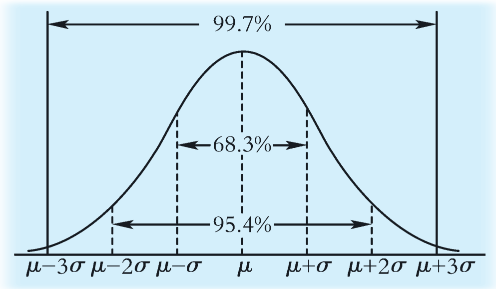


图 8－6

在实际应用中，通常认为服从于正态分布 $N(\mu, \sigma^2)$ 的随机变量 X 只取 $(\mu-3\sigma, \mu+3\sigma)$ 之间的值，并简称之为 3σ 原则.

事实上，μ 就是随机变量 X 的均值(期望)，σ^2 就是随机变量 X 的方差，它们分别反映 X 取值的平均大小和稳定程度.

通过查标准正态分布表(见附录 1)可以确定服从标准正态分布的随机变量的有关概率.

例 1 若随机变量 $X \sim N(0, 1)$，查标准正态分布表，求：

(1) $P(X \leqslant 1.52)$； (2) $P(X > 1.52)$；

(3) $P(0.57 < X \leqslant 2.3)$； (4) $P(X \leqslant -1.49)$.

解 (1) $P(X \leqslant 1.52) = \Phi(1.52) = 0.93574$.

(2) $P(X > 1.52) = 1 - P(X \leqslant 1.52) = 1 - 0.93574 = 0.06426$.

(3) $P(0.57 < X \leqslant 2.3) = P(X \leqslant 2.3) - P(X \leqslant 0.57) = \Phi(2.3) - \Phi(0.57) = 0.27358$.

(4) $P(X \leqslant -1.49) = P(X \geqslant 1.49) = 1 - P(X \leqslant 1.49) = 1 - \Phi(1.49) = 1 - 0.93189 = 0.06811$.

设 $X \sim N(\mu, \sigma^2)$，则有关 X 取值的概率计算主要依据如下关系式：

(1) $F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$； (2) $\Phi(-x)=1-\Phi(x)$，$x>0$.

其中，$\Phi(a)$的值可以通过查标准正态分布表(见附录 1)得到.

例 2 当 X 服从正态分布，且数学期望为 4 和标准差为 2 时，计算$P(6<X\leqslant 8)$.

解 ∵ $X\sim N(4,2^2)$，∴ $\mu=4$，$\sigma=2$.

∴ $P(6<X\leqslant 8)=F(8)-F(6)=\Phi\left(\frac{8-4}{2}\right)-\Phi\left(\frac{6-4}{2}\right)=\Phi(2)-\Phi(1)=0.977\ 25-0.841\ 3=0.135\ 95.$

练习

1. 设 $X\sim N(0,1)$，查表求(精确到 0.000 1)：

   (1) $P(0<X<1.90)$； (2) $P(-1.83<X<0)$.

2. 查表计算(精确到 0.000 1)：$\Phi(1)+\Phi(2)$.

习题 9

学而时习之

1. 设 $X\sim N(0,1)$，查表求(精确到 0.000 1)：

   (1) $P(X\leqslant 2.75)$； (2) $P(X<0.5)$； (3) $P(-2<X\leqslant 2.9)$.

2. 内科医生对某病人进行了血压的测量，用 X 表示测量的收缩压(单位：mmHg). 设 X 服从正态分布. 如果病人当时的真实收缩压是 μ，

   (1) 当血压计的测量标准差是 1，计算 $P(|X-\mu|\leqslant 1.96)$；

   (2) 当血压计的测量标准差是 1.5，计算 $P(|X-\mu|\leqslant 2.94)$.

收缩压就是我们常说的高压.

3. 设 $X\sim N(2.3,4)$，查表求 $P(1\leqslant X<3)$(精确到 0.000 1).

4. 用 X 表示包装机包装出的袋装食盐的重量. 已知 X 是服从正态分布的随机变量，有数学期望 0.5 kg，标准差 0.01 kg. 计算一袋食盐的重量少于 0.49 kg 的概率.

8.4 列联表独立性分析案例

在许多实际问题中，我们需要考察两种因素的关系．例如，患肺癌与吸烟是否有关？儿童语言能力是否与他们的性别有关？

为了分析这些问题，我们需要用问卷调查或现场记录等方式获取一批数据．例如，为了了解患肺癌是否与吸烟有关，就需要调查其他条件都基本相同的 n 个人，然后将调查结果列成表 8.2 的形式(表中 X 表示“是否吸烟”，Y 表示“是否患肺癌”)．

表 8.2

X \ Y	患肺癌(B)	不患肺癌($\overline{B}$)	总计
吸烟(A)	a	b	$a+b$
不吸烟($\overline{A}$)	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

我们称类似 8.2 的表格为列联表；称 X，Y 为两个因素；称“吸烟”和“不吸烟”为 X 的两个水平；称“患肺癌”和“不患肺癌”为 Y 的两个水平．

由于所涉及的两个因素 X，Y 均有两个水平，所以称表 8.2 为 2×2列联表．

列联表 8.2 的独立性分析就是根据表中的数据分析因素 X，Y 是否相互独立．

下面我们通过对具体案例的分析学习列联表的独立性分析方法．

案例 患肺癌与吸烟是否有关？

为研究患肺癌是否与吸烟有关，从一批在年龄、生活和工作环境等方面相仿的男性中分别随机抽取了 60 名肺癌患者和 40 名非肺癌患者，调查他们是否吸烟．调查结果列入表 8.3．

值得指出的是：这里要求被调查对象在年龄、生活和工作环境等因素方面尽量相同是为了避免这些因素对“是否患肺癌”的影响，因

为不同的年龄段或者不同的生活、工作环境等因素可能也会导致人们易患肺癌. 如果调查时不考虑这些因素，即使我们分析的结果是患肺癌与吸烟有关，也不清楚这种关系反映的是患肺癌与吸烟之间的关系，还是由其他因素引起的关系. 因此，只有尽量控制调查对象在其他方面尽可能一致，才能根据调查数据有效地分析患肺癌与吸烟的相关性.

将上述调查数据列成 2×2 列联表，并计算出各行各列的和，得到表 8.3.

表 8.3　　肺癌与吸烟的调查数据

X \ Y	患肺癌(B)	未患肺癌($\bar{B}$)	总计
吸烟(A)	39	15	54
不吸烟($\bar{A}$)	21	25	46
总计	60	40	100

从表 8.3 可以得出，在 54 个吸烟的人中有 39 人患肺癌，患者占 $39/54\approx72.22\%$；在不吸烟的 46 人中，有 21 人患肺癌，患者占 $21/46\approx45.65\%$. 吸烟者中患肺癌的比例比不吸烟者中患肺癌的比例高出

$$72.22\%-45.65\%=26.57\%.$$

这种差异似乎已经说明吸烟与患肺癌有很大关系. 但仔细想想，由于这 100 人是随机选取的，会不会由于随机抽样的误差使得所抽取的 60 名肺癌患者中碰到了较多的吸烟者，而在 40 名非肺癌患者中碰到了较多的不吸烟者？这样也可能导致吸烟者中肺癌患者的比例比不吸烟者中肺癌患者的比例高.

于是，我们还需进一步用统计方法说明单凭随机抽样的误差还不足以造成如此大的差异.

在本例中，$n=a+b+c+d=100$，

$$a=39,\ b=15,\ c=21,\ d=25;$$

$$a+b=54,\ c+d=46,\ a+c=60,\ b+d=40.$$

为分析 X，Y 是否独立(相关)，先提出假设 H_0：X，Y 独立(无

关），也就是假设“吸烟(A)”与“患肺癌(B)”独立(无关)．这时 A 与 B 独立，$\overline{A}$ 与 B 独立，A 与 $\overline{B}$ 独立，$\overline{A}$ 与 $\overline{B}$ 独立．

于是 $P(A\cap B)=P(A)P(B)$， $P(\overline{A}\cap B)=P(\overline{A})P(B)$，

$P(A\cap\overline{B})=P(A)P(\overline{B})$， $P(\overline{A}\cap\overline{B})=P(\overline{A})P(\overline{B})$．

根据概率与频率的关系，知道 $P(A\cap B)$的估计值为 $p_{AB}=\frac{a}{n}=0.39$，$P(\overline{A}\cap B)$的估计值为 $p_{\overline{A}B}=\frac{c}{n}=0.21$，$P(A\cap\overline{B})$的估计值为 $p_{A\overline{B}}=\frac{b}{n}=0.15$，$P(\overline{A}\cap\overline{B})$的估计值为 $p_{\overline{A}\,\overline{B}}=\frac{d}{n}=0.25$.

又 $P(A)$的估计值为 $p_A=\frac{a+b}{n}=0.54$，$P(\overline{A})$的估计值为 $p_{\overline{A}}=\frac{c+d}{n}=0.46$，$P(B)$的估计值为 $p_B=\frac{a+c}{n}=0.6$，$P(\overline{B})$的估计值为 $p_{\overline{B}}=\frac{b+d}{n}=0.4$.

因为假设 X，Y 独立，所以 $\mu_{AB}=|p_{AB}-p_Ap_B|$，$\mu_{\overline{A}B}=|p_{\overline{A}B}-p_{\overline{A}}p_B|$，$\mu_{A\overline{B}}=|p_{A\overline{B}}-p_Ap_{\overline{B}}|$，$\mu_{\overline{A}\,\overline{B}}=|p_{\overline{A}\,\overline{B}}-p_{\overline{A}}p_{\overline{B}}|$都相应比较小，我们用

$$\chi^2=\frac{n\mu_{AB}^2}{p_Ap_B}+\frac{n\mu_{\overline{A}B}^2}{p_{\overline{A}}p_B}+\frac{n\mu_{A\overline{B}}^2}{p_Ap_{\overline{B}}}+\frac{n\mu_{\overline{A}\,\overline{B}}^2}{p_{\overline{A}}p_{\overline{B}}}$$

$$=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (1)$$

表示 μ_{AB}，$\mu_{\overline{A}B}$，$\mu_{A\overline{B}}$，$\mu_{\overline{A}\,\overline{B}}$的总体大小．当 H_0 成立时，χ^2取值应该比较小，当 χ^2取值较大时，表示假设 H_0 不成立．

这是 1900 年皮尔逊(K·Pearson)引进的统计量 χ^2．有时为书写方便，χ^2也记作 K^2.

在本案例中，经过计算得到 χ^2的观测值为

$$\chi^2=\frac{100(39\times25-15\times21)^2}{54\times46\times60\times40}\approx7.31.$$

那么，$\chi^2=7.31$ 是否太大呢?

统计学家已经有明确的结论：如果 2×2 列联表中的两个因素 X，Y 是独立的，即在 H_0 成立的情况下，且当随机调查的数据 a，b，c，d 都不小于 5 时，随机事件“$\chi^2\geqslant6.635$”发生的概率约为 0.01，即

$$P(\chi^2\geqslant6.635)\approx0.01. \qquad (2)$$

当观测数据 a，b，c，d 中有小于 5 的数时，需采用很复杂的精确的检验方法．

也就是说，在 H_0 成立的情况下，χ^2的值大于 6.635 的概率非常

小，近似于 0.01. 即在 H_0 成立的情况下，对随机变量 χ^2 进行多次观测，观测值超过 6.635 的频率为 0.01.

在本案例中，由调查数据所得到的 $\chi^2=7.31>6.635$，这件事发生的概率 $P(\chi^2\geqslant 7.31)\leqslant P(\chi^2\geqslant 6.635)\approx 0.01$. 这是一个小概率事件，因此我们有 99% 的把握认为 H_0 不成立，即在本案例中发生的原因很可能是由于假定 H_0 “患肺癌与吸烟独立”不对. 于是否定 H_0，从而认为患肺癌与吸烟有关系.

这种思想类似于反证法.

值得指出的是，我们在作出上述判断时也有可能犯错误，因为患肺癌与吸烟无关系时，χ^2 的值仍有可能超过 6.635，但是这件事发生的概率不超过 0.01. 也就是说我们犯错误的概率不会超过 0.01.

上面这种利用随机变量 χ^2 来确定在多大程度上可以认为“两个分类变量有关系”的方法称为两个分类变量的**独立性检验**.

利用独立性检验来考察两个分类变量 X 与 Y 是否有关的具体做法是：(1) 提出假设 H_0：X 与 Y 无关；

(2) 根据 2×2 列联表与公式(1)计算 χ^2 的值；

(3) 查对临界值(表 8.4)，作出判断.

表 8.4

$P(\chi^2\geqslant x_0)$	0.50	0.40	0.25	0.15	0.10	0.05	0.025	0.010	0.005	0.001
x_0	0.455	0.708	1.323	2.072	2.706	3.841	5.024	6.635	7.879	10.828

例如：(1) 如果 $\chi^2>10.828$，就有 99.9% 的把握认为“X 与 Y 有关”；

(2) 如果 $\chi^2>6.635$，就有 99% 的把握认为“X 与 Y 有关”；

(3) 如果 $\chi^2>3.841$，就有 95% 的把握认为“X 与 Y 有关”；

如果 $\chi^2\leqslant 3.841$，就认为没有充分的证据显示“X 与 Y 有关”.

例 用两种检验方法对某食品做沙门氏菌检验，结果如下. 试比较两种方法和阴性结果是否有关系.

	阳性	阴性	合计
荧光抗体法	160	5	165
常规培养法	26	48	74
合计	186	53	239

解 提出假设 H_0：两种方法与阴性没有关系.

由题意可知，$a=160$，$b=5$，$c=26$，$d=48$，$a+b=165$，$c+d=74$，$a+c=186$，$b+d=53$，$n=a+b+c+d=239$，将它们分别代入公式得

$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}\approx 113.185.$$

查表 8.4 可知，当 H_0 成立时，$P(\chi^2\geqslant 10.828)\approx 0.001$，即当 H_0 成立时，$\chi^2\geqslant 10.828$ 的概率约为 0.001(或 0.1%)，而这里的$\chi^2\approx 113.185$ 远大于 10.828.

因此，我们有 99.9%的把握认为它们之间有关系.

练 习

某项试验，在 100 次试验中，成功率只有 10%，进行技术改造后，又进行了 100 次试验，试问：若要有 97.5%以上的把握认为“技术改造有明显效果”，试验的成功率最少应是多少？(设 $P(\chi^2\geqslant 5)=0.025$)

习题 10

学而时习之

为了考察某种新药预防疾病的作用，进行动物试验得到如下观测数据.

	患病	未患病	合　计
服用药	15	35	50
没服用药	4	46	50
合　计	19	81	100

请问能有多大把握认为药物有效？

8.5 一元线性回归案例

案例 海牛是一种体型较大的水生哺乳动物，体重可达到700 kg，以水草为食. 美洲海牛生活在美国的佛罗里达州，在船舶运输繁忙季节，经常被船的螺旋桨击伤致死. 下面是佛罗里达州记录的1977—1990年机动船只数目 x 和被船只撞死的海牛数 y 的数据.

年份	1977	1978	1979	1980	1981	1982	1983
船只数量 x	447	460	481	498	513	512	526
被撞死的海牛数 y	13	21	24	16	24	20	15
年份	1984	1985	1986	1987	1988	1989	1990
船只数量 x	559	585	614	645	675	711	719
被撞死的海牛数 y	34	33	33	39	43	50	47

现在问：

(1) 随着机动船的数量的增加，被撞死的海牛数是否会增加?

(2) 当机动船增加到750只，被撞死的海牛会是多少?

要解决上面的问题先为数据画出散点图，横坐标是 x，纵坐标是 y，见图 8－7.

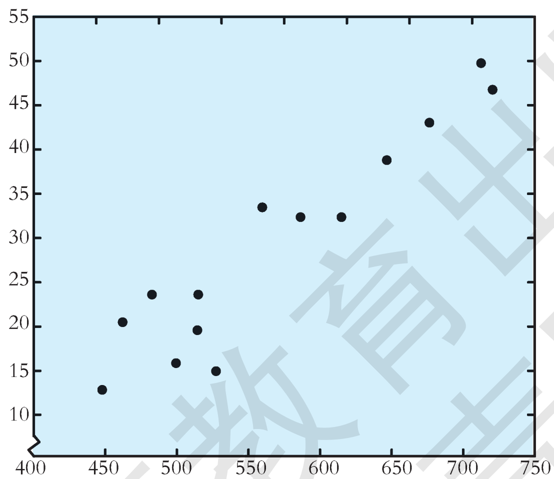


图 8－7 船只数量和被撞死的海牛数的散点图

从数据散点图上看到 y_i 有随着 x_i 的增加而沿某一直线增加的趋势．直线确定了，问题(1)也就解决了，但这条直线应当如何确定呢？

无论是从抽样调查中得到的成对数据，还是从科学试验、工农业生产中得到的成对数据，在统计学中都被称为观测数据或样本，数据的个数被称为样本容量．

上面案例中的 14 对观测数据称为样本，由图 8－7 的 14 个点表示．

样本容量是 n 的成对观测数据，用 $(x_1,\ y_1)$，$(x_2,\ y_2)$，…，$(x_n,\ y_n)$ 表示．这里，对固定的 i，x_i 和 y_i 来自相同的个体或是同一次试验的观测数据．对 $i\neq j$，$(x_i,\ y_i)$ 和 $(x_j,\ y_j)$ 来自不同的个体或是不同试验的观测数据．

对于上述观测数据，我们用 $\{x_i\}$ 表示数据 x_1，x_2，…，x_n，用 $\{y_i\}$ 表示数据 y_1，y_2，…，y_n．用 $\bar{x}$ 和 $\bar{y}$ 分别表示 $\{x_i\}$ 和 $\{y_i\}$ 的均值．用 s_x 表示 $\{x_i\}$ 的标准差，用 s_y 表示 $\{y_i\}$ 的标准差．

再引入

$$s_{xy}=\frac{x_1y_1+x_2y_2+\cdots+x_ny_n}{n}-\bar{x}\bar{y}.$$

定义 (1) 当 $s_xs_y\neq 0$，我们称

$$r_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$=\frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)\left(\sum_{i=1}^{n}y_i^2-n\bar{y}^2\right)}}=\frac{s_{xy}}{s_xs_y}$$

为 $\{x_i\}$ 和 $\{y_i\}$ 的**相关系数**；

(2) 当 $r_{xy}>0$ 时，我们称 $\{x_i\}$ 和 $\{y_i\}$ **正相关**；

(3) 当 $r_{xy}<0$ 时，我们称 $\{x_i\}$ 和 $\{y_i\}$ **负相关**；

(4) 当 $r_{xy}=0$ 时，我们称 $\{x_i\}$ 和 $\{y_i\}$ **不相关**．

理论上可以证明相关系数 r_{xy} 具有以下性质：

1. r_{xy} 总是在区间 $[-1,1]$ 中取值；

2. 当 r_{xy} 越接近于 1 时，x，y 的线性相关程度越强，且 x 增加，y 也倾向于增加，这时数据 $(x_1,\ y_1)$，$(x_2,\ y_2)$，…，$(x_n,\ y_n)$ 分散在

一条上升的直线附近；

3. 当 r_{xy} 越接近于 -1 时，x，y 的线性相关程度越强，且 x 增加，y 倾向于减少，这时数据 (x_1, y_1)，(x_2, y_2)，…，(x_n, y_n) 分散在一条下降的直线附近.

4. 当 r_{xy} 越接近于 0 时，x，y 的线性相关程度越弱.

图 8-8 至图 8-9 分别是 $\{x_i\}$ 和 $\{y_i\}$ 之间正相关和负相关的举例，其中样本容量都是 50.

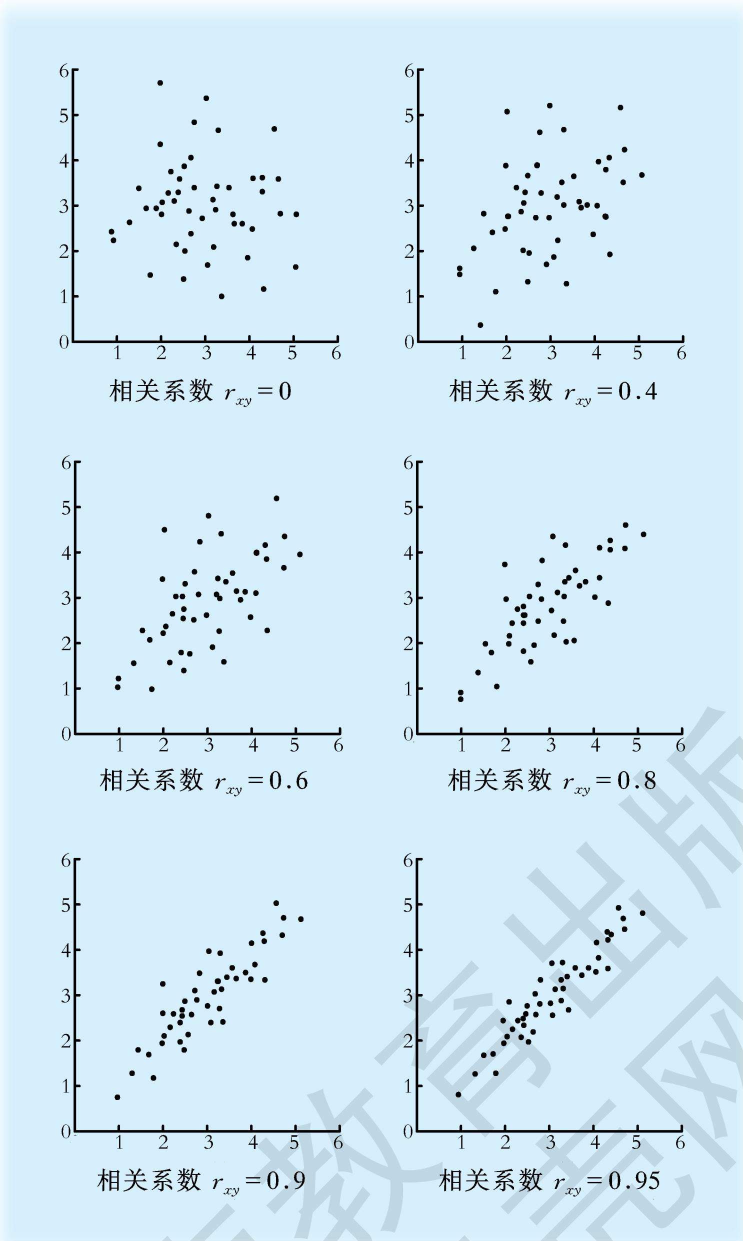


图 8-8　$r_{xy}\geqslant 0$

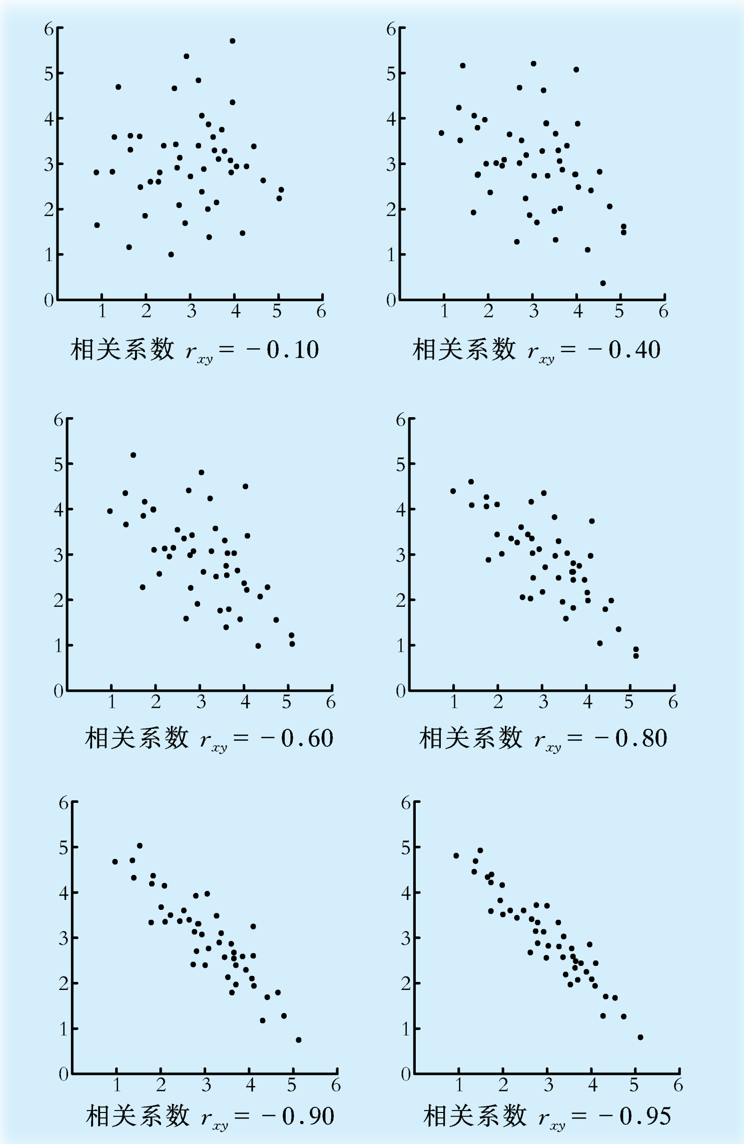


图 8-9 $r_{xy}<0$

从图中看出当 $r_{xy}>0.8$ 时，y 有随着 x 的增加而增加的趋势，这时我们认为 $\{x_i\}$ 和 $\{y_i\}$ 是高度正相关的．当 $r_{xy}<-0.8$ 时，y 有随着 x 的增加而减少的趋势，这时我们称 $\{x_i\}$ 和 $\{y_i\}$ 是高度负相关的．

通常在解决具体问题时，若 $|r_{xy}|$ 大于 0.8，则我们就认为两个变量高度相关．

现在我们来解决本节开始的案例中的两个问题．

为解决问题(1)，先计算机动船数 $\{x_i\}$ 和被撞死的海牛数 $\{y_i\}$ 的相关系数 r_{xy}．

本案例中，经过计算得到

$$\overline{x}=567.5,\quad \overline{y}=29.43,$$

$$s_x=88.56,\quad s_y=11.75,\quad s_{xy}=979.36,$$

于是相关系数

$$r_{xy}=\frac{979.36}{88.56\times 11.75}\approx 0.941\ 2.$$

这说明被撞死的海牛数 y 和机动船数 x 高度正相关．因此，只要机动船数增加，被撞死的海牛数就会增加．

为了解决问题(2)，需要为数据建立回归直线．设回归直线是

$$l:\ \hat{y}=bx+a.$$

我们可以认为 y_i 和 x_i 满足以下的关系：

$$y_i=bx_i+a+e_i,\ i=1,\ 2,\ \cdots,\ n.$$

其中的 e_1，e_2，…，e_n 表示随机误差．我们称上述的模型为一元**线性回归模型**(linear regression model)．

解决一元线性回归模型的方法是求出直线 l，这里的直线 l 就是以前学习的回归直线．

利用最小二乘法得到的 b，a 的最小二乘估计值是

$$b=\frac{s_{xy}}{s_x^2}=\frac{979.36}{88.56^2}\approx 0.125,$$

$$a=\bar{y}-b\bar{x}=29.43-0.125\times 567.5\approx -41.5.$$

回归直线是

$$l:\ \hat{y}=0.125x-41.5.$$

此回归直线的图形如图 8－10 所示．

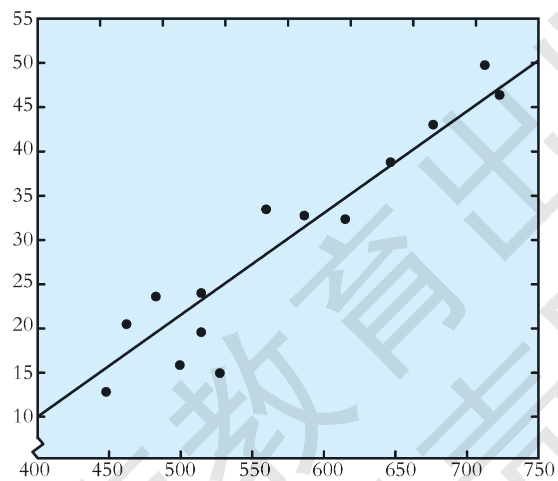


图8－10　船只数量和被撞死的海牛数的回归直线

当机动船数增加到 750 条时，被撞死的海牛数的预测值是

$$\hat{y}=0.125\times750-41.5\approx52(\text{只}).$$

例 1 下面是我国 1990—2000 年出口贸易额 x(百亿美元)和我国 GDP(国内生产总值)y(百亿元人民币)的数据.

年　份	1990	1991	1992	1993	1994	1995
x	6.21	7.19	8.49	9.17	12.10	14.88
y	185.48	216.18	266.38	346.34	467.59	584.78
年　份	1996	1997	1998	1999	2000	
x	15.11	18.29	18.37	19.49	24.92	
y	678.85	744.63	783.45	820.68	894.42	

试建立 y 与 x 之间的回归方程.

解 根据收集的数据作散点图（图 8－11).

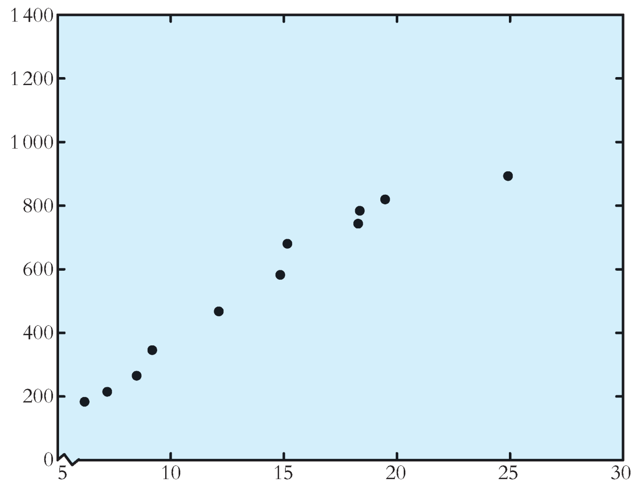


图 8－11　出口贸易额和 GDP 数据散点图

从图 8－11 可以看出(x_i, y_i)（$i=1, 2, \cdots, 11$）集中在一条直线附近，因此可以用线性回归模型 $y_i=bx_i+a+e_i(i=1, 2, \cdots, 11)$ 来描述 x，y 之间的关系. e_i 描述了随机误差造成的影响，其中也包括了观测误差. 经计算，

$$\bar{x}=14.02,\ \bar{y}=544.43,\ s_x=5.67,\ s_y=247.67,\ s_{xy}=1\,372.09.$$

相关系数

$$r_{xy}=\frac{1\,372.09}{5.67\times247.67}\approx0.977.$$

因此，x 和 y 是高度正相关的，这说明外贸出口带动了 GDP 的增长．通过计算可得出 a，b 的最小二乘估计值分别为：

$$b=\frac{s_{xy}}{s_x^2}=\frac{1\ 372.09}{5.67^2}\approx 42.68,$$

$$a=\bar{y}-b\bar{x}=544.43-42.68\times 14.01\approx -53.94.$$

于是，回归直线为 $\hat{y}=42.68x-53.94$（见图 8－12）．

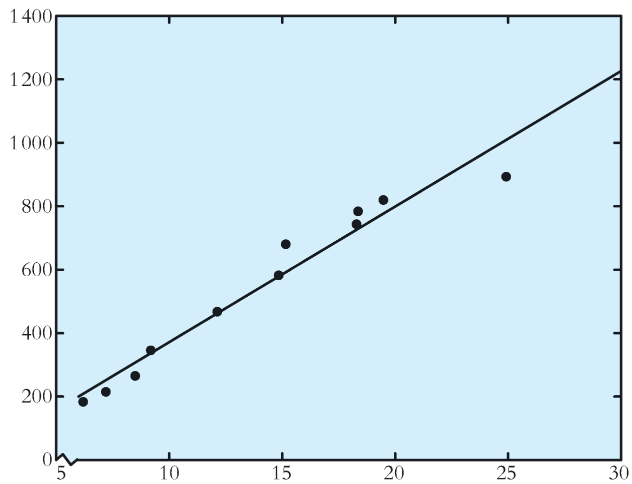


图 8－12　出口贸易额和 GDP 的回归直线

对于 2001 年的出口贸易额 $x=26.61$，可以用回归直线作出 2001 年 GDP 的预测值为

$$\hat{y}=42.68\times 26.61-53.94=1\ 081.77.$$

例 2　某地区对本地企业的人均资本 x(万元)与人均产出 y(万元)进行了一次抽样调查，下表是这次抽查中所得到的各企业的数据：

人均资本 x(万元)	3	4	5.5	6.5	7	8	9	10.5	11.5	14
人均产出 y(万元)	4.12	4.67	8.68	11.01	13.04	14.43	17.50	25.46	26.66	45.20

(1) 若 y 与 x 之间具有近似关系 $y\approx ax^b$（a，b 为常数），试根据表中数据估计 a 和 b 的值；

(2) 估计当企业人均资本为 16 万元时的人均产出(精确到 0.01)．

分析　根据 x，y 所具有的关系可知，此问题不是线性回归问题，不能直接用线性回归方程处理．但由对数性质可知，只要对 $y\approx ax^b$ 的两边取自然对数，就能将其转化为线性关系．

在实际问题中，有时两个变量之间的关系并不是线性关系，如例 2．这需要我们根据专业知识或散点图，选择适当的曲线方程，然后通过适当的变量代换，把非线性方程化为线性回归方程，从而确定未知参数．下面列举出一些常用的曲线方程，并给出相应的化为线性回归方程的换元公式．

解 (1) 在 $y\approx ax^{b}$ 的两边取自然对数，可得 $\ln y\approx\ln a+b\ln x$，设

$$\ln y=y',\ \ln a=a',\ \ln x=x',$$

则

$$y'\approx a'+bx'.$$

根据数据计算如下表所示.

人均资本 x(万元)	3	4	5.5	6.5	7
人均产出 y(万元)	4.12	4.67	8.68	11.01	13.04
$x'=\ln x$	1.098 61	1.386 29	1.704 75	1.871 80	1.945 91
$y'=\ln y$	1.415 85	1.541 16	2.161 02	2.398 80	2.568 02
人均资本 x(万元)	8	9	10.5	11.5	14
人均产出 y(万元)	14.43	17.5	25.46	26.66	45.20
$x'=\ln x$	2.079 44	2.197 22	2.351 38	2.442 35	2.639 06
$y'=\ln y$	2.669 31	2.862 20	3.237 11	3.283 16	3.811 10

仿照例 1，可得 a'，b 的最小二乘估计值分别为 $a'=-0.496\ 27$，$b=1.567\ 7$，由 $a'=\ln a=-0.496\ 27$ 可得，

$$a\approx0.608\ 8,$$

即 a，b 的值分别约为 0.608 8 和 1.567 7.

(2) 由(1)知

$$\hat{y}=0.608\ 8x^{1.567\ 7}.$$

样本数据及回归曲线的图形如图 8－13 所示.

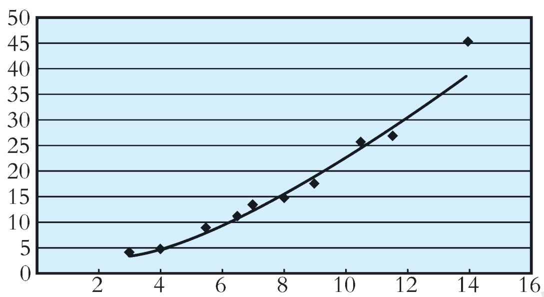


图 8－13

当 $x=16$ 时，$\hat{y}=0.608\ 8\times16^{1.567\ 7}\approx47.01$(万元)，故当企业人均资本为 16 万元时，人均产值约为 47.01 万元.

(1) $y=a+\dfrac{b}{x}$，令 $y'=y$，$x'=\dfrac{1}{x}$，则有 $y'=a+bx'$.

(2) $y=ax^{b}$，令 $y'=\ln y$，$x'=\ln x$，$a'=\ln a$，则有 $y'=a'+bx'$.

(3) $y=ae^{bx}$，令 $y'=\ln y$，$x'=x$，$a'=\ln a$，则有 $y'=a'+bx'$.

(4) $y=ae^{\frac{b}{x}}$，令 $y'=\ln y$，$x'=\dfrac{1}{x}$，$a'=\ln a$，则有 $y'=a'+bx'$.

(5) $y=a+b\ln x$，令 $y'=y$，$x'=\ln x$，则有 $y'=a+bx'$.

练习

下表是从某大学随机抽取的 8 名女大学生的身高 x(cm)和体重 y(kg)的数据：

编号	1	2	3	4	5	6	7	8
身高 x	165	165	157	170	175	165	155	170
体重 y	48	57	50	54	64	61	43	59

试求出 y 与 x 之间的回归方程，并预报一名身高为 172 cm 的女生的体重.

习题 11

学而时习之

1. 研究某灌溉渠道水的流速 y 与水深 x 之间的关系，测得一组数据如下：

水深 x/m	1.40	1.50	1.60	1.70	1.80	1.90	2.00	2.10
流速 y/($\mathrm{m\cdot s^{-1}}$)	1.70	1.79	1.88	1.95	2.03	2.10	2.16	2.21

(1) 画出散点图；

(2) 求 y 对 x 的回归直线方程；

(3) 预测水深为 1.95 m 时水的流速是多少？

2. 一只红铃虫的产卵数 y 与温度 x 有关. 现收集了 7 组观测数据列于下表中，

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

试建立 y 与 x 之间的回归方程.（要求分 $y=ae^{bx}$ 和 $y=cx^2+d$ 两种形式进行探讨.）

多知道一点

假设检验案例

案例 1 一条新建的交通干线全长 10 km，前半段 5 km，后半段 5 km. 在刚刚通车的一个月中，后半段就发生了 4 起交通事故，而前半段没有发生交通事故. 能否认为后半段发生交通事故的概率比前半段大？

解 同一起交通事故发生在后半段就不能发生在前半段，就像硬币掷出反面时就不会出现正面一样. 4 起交通事故的发生是相互独立的，它们之间没有联系.

如果前、后半段发生交通事故的概率相同，则每一起事故发生在后半段的概率都是 0.5，于是这 4 起交通事故都发生在后半段的概率是

$$0.5^4=0.0625.$$

这是一个很小的概率，一般不会发生. 所以我们认为后半段发生交通事故的概率比前半段大.

作出以上结论也是有可能犯错误的，犯错误的概率正是0.062 5. 这是因为当前、后半段发生交通事故的概率相同，而 4 起交通事故又都出现在后半段时，我们才犯错误. 也就是说我们犯错误的概率等于前、后半段发生交通事故的概率相同的条件下，4 起交通事故都出现在后半段时的概率. 这一概率正是 0.062 5. 于是，我们判断正确的概率是 $1-0.0625=93.75\%$.

因此我们是以 93.75% 的把握保证后 5 km 比前 5 km 更容易发生交通事故.

前方是事故多发路段，请小心驾驶.

得到了上述结果后，交通管理部门很快在进入后半段的地点安放了警示牌：前方是事故多发路段，请小心驾驶.

案例 2 一服装店出售标价为 180 元的夹克. 售货员声称对前来问价的顾客以 180 元推销成功的概率是 $p=0.6$. 现在 1 小时内有 4

位顾客前来问价，服务员对这 4 位顾客都没有推销成功. 能否判定售货员的 $p=0.6$ 不对?

解 用 A_1，A_2，A_3，A_4 分别表示对第 1，第 2，…，第 4 位顾客没有推销成功，则 A_1，A_2，A_3，A_4 相互独立. $A=A_1\cap A_2\cap A_3\cap A_4$ 表示对这 4 位顾客都没有推销成功. 利用

$$P(A_j)=1-0.6=0.4,\ j=1,\ 2,\ 3,\ 4$$

得到

$$P(A)=P(A_1)P(A_2)P(A_3)P(A_4)=0.4^4=0.025\ 6.$$

这是一个小概率事件，其发生的原因很可能是 $p=0.6$ 不对. 所以应当判定售货员的 $p=0.6$ 不对. 作出这个结论也可能犯错误. 犯错误的概率是 2.56%. 于是判断正确的概率是 97.44%. 因此，我们以 97.44% 的概率保证 $p<0.6$.

案例 3 某地区的山羊患某种疾病的概率是 0.4，且每只山羊患病与否是相互独立的. 现在为了判断一种新的预防药是否有预防作用，随机选取了 6 只山羊做试验，这 6 只山羊用药后都没有得这种病. 问此新药是否有效?

解 初看起来，6 只山羊用药后都没有得病，应当判断新药有效.

但是细想一下，就会发现即使新药无效，6 只山羊也可以都不得病. 为了作出正确的判断，让我们假设新药无效，然后看看事实是否支持这个假设.

用 A_i 表示第 i 只山羊不得病，则 A_1，A_2，…，A_6 相互独立，$A=A_1\cap A_2\cap\cdots\cap A_6$ 表示 6 只山羊都不得病.

假设新药无效，则 $P(A_i)=1-0.4=0.6$，于是有

$$\begin{aligned}P(A)&=P(A_1\cap A_2\cap\cdots\cap A_6)\\&=P(A_1)P(A_2)\cdot\cdots\cdot P(A_6)\\&=0.6^6\approx0.046\ 7.\end{aligned}$$

这个概率很小，一般是不会发生的，它的发生说明我们的假设有问题. 于是我们否定原来的假设，认为新药是有效的.

否定新药无效也可能犯错误，但是犯错误的概率是 0.046 7. 因为只有在新药无效的条件下，6 只山羊都不得病，我们才犯错误.

案例 3 解决的问题是统计中的假设检验问题. 先作一个假设：新药有效，在这个假设下看看实际情况是否支持这个假设，如果在这个假设下小概率事件发生了，说明实际情况不支持这个假设，于是我们就否定这个假设.

实际问题中经常将小于或等于 0.05(或 0.1)的概率视为小概率，这时否定“假设”时，犯错误的概率不超过 0.05(或 0.1).

小结与复习

1. 随机对照试验： 随机选取试验组和对照组是安排试验的基本原则. 随机对照试验是指随机选取试验组和对照组的试验. 我们把对照组中的处理方法称为使用安慰剂.

2. 概率：

(1) 加法公式：如果 Ω 的事件 A_1，A_2，…，A_m 两两互斥，则 $P(A_1 \cup A_2 \cup \cdots \cup A_m)=P(A_1)+P(A_2)+\cdots+P(A_m)$.

(2) 条件概率公式：设 A，B 是事件，用 $P(B|A)$ 表示已知 A 发生的条件下，B 发生的条件概率. 如果 $P(A)>0$，则有

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

(3) 事件的独立性：如果事件 A_1，A_2，…，A_n 是相互独立的，则

$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)P(A_2)\cdot\cdots\cdot P(A_n).$$

(4) 如果随机变量 X 的取值是 x_1，x_2，…，x_n，则 $\{X=x_i\}$ 是事件，用 $p_i=P(X=x_i)$ 表示事件 $\{X=x_i\}$ 的概率，则称

$$p_i=P(X=x_i),\ i=1,\ 2,\ \cdots,\ n$$

是随机变量 X 的概率分布. X 的概率分布 $\{p_i\}$ 还可以用下面的表格表示.

X	x_1	x_2	x_3	…
P	p_1	p_2	p_3	…

(5) 两点分布 $B(1,p)$：对于任一个试验，引入随机变量

$$X=\begin{cases}1, & \text{当试验成功；}\\ 0, & \text{当试验不成功.}\end{cases}$$

则 X 服从两点分布 $B(1,p)$：

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

并且

$$E(X)=p,\ D(X)=p(1-p).$$

(6) 二项分布 $B(n,p)$：设一试验成功的概率为 p，$p\in(0,1)$. 将该试验独立重复 n 次，用 X 表示成功的次数，则 X 服从二项分布 $B(n,p)$，即

$$P(X=k)=\mathrm{C}_n^k p^k q^{n-k},\ k=0,1,\cdots,n，其中\ q=1-p,$$

这时，

$$E(X)=np,\ D(X)=np(1-p).$$

(7) 超几何分布 $H(N,M,n)$：N 件产品中有 M 件次品，从中随机抽取 n 件，用 X 表示这 n 件中的次品数，则 X 服从超几何分布 $H(N,M,n)$，即

$$P(X=k)=\frac{\mathrm{C}_M^k\mathrm{C}_{N-M}^{n-k}}{\mathrm{C}_N^n},\ k=0,1,\cdots,m,\ m=\min\{M,n\}.$$

这时，

$$E(X)=n\frac{M}{N},\ D(X)=\frac{nM}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}.$$

(8) 数学期望和方差：当随机变量 X 有概率分布

$$p_j=P(X=x_j),\ j=0,1,\cdots,n,$$

就称

$$E(X)=x_1p_1+x_2p_2+\cdots+x_np_n$$

为 X 的数学期望或均值.

用 $\mu=E(X)$ 表示 X 的数学期望时，称

$$D(X)=(x_1-\mu)^2p_1+(x_2-\mu)^2p_2+\cdots+(x_n-\mu)^2p_n$$

为 X 的方差，称 $\sqrt{D(X)}$ 为 X 的标准差. 我们还用 σ^2 表示方差 $D(X)$，用 σ 表示标准差 $\sqrt{D(X)}$.

3. 正态分布：如果用 X 表示测量误差，则 X 服从正态分布.

4. 列联表：在许多实际问题中，经常需要考察两种因素之间的关系. 列联表的独立性分析方法是检验所述的两个因素是否独立的有效方法.

5. 回归直线：当 $\{x_i\}$ 和 $\{y_i\}$ 的相关系数 r_{xy} 的绝对值 $|r_{xy}|$ 较大

时，可以用一条直线描述数据$\{x_i\}$和$\{y_i\}$的关系，这条直线就是回归直线．用

$$l：\hat{y}=bx+a$$

表示这条回归直线时，其中的 a，b 可以用下面的公式进行计算：

$$b=\frac{s_{xy}}{s_x^2}，\ a=\bar{y}-b\bar{x}.$$

用回归直线进行预测：得到了回归直线后，只要$\{x_i\}$与$\{y_i\}$高度相关，即只要相关系数 $|r_{xy}|>0.8$，对于新的 x，就可以用回归直线上的点 $\hat{y}=bx+a$ 作为 y 的预测值．事实证明：$|r_{xy}|$ 越接近于 1，预测就越准确；x 越接近 $\bar{x}$，预测也越好．

复习题八

学而时习之

1. 在对一种新的安眠药进行药效评估时，调查了 20 名开始使用这种药的人，结果有 16 人认为新药比常用药更有效．

(1) 能否作出新药比常用药更有效的结论？

(2) 如果不能作出上述结论，应当采用怎样的试验方案？

(3) 如何安排对照组和试验组？

(4) 本例中是否应当使用安慰剂？

2. 早在 70 多年前，美国通用电器公司成立了由社会学家和公司人事部成员组成的研究组．研究组的任务是考察照明程度对生产灯泡的工人的生产率有何影响．研究中发现，增加照明度后产量增加．但是奇怪的是降低照明度后，产量也增加．原因是（　　）

(A) 工人们对研究组的研究工作有了反应

(B) 增加照明度确能提高生产率

(C) 降低照明度确能提高生产率

3. 某人的手机收到的短信中有 1/6 是广告，已知他今天收到了 6 条短信，用 X 表示这 6 条短信中的广告数. 计算：

(1) $P(X=3)$； (2) $P(X\geqslant 3)$； (3) $P(X<3)$；

(4) $P(2\leqslant X\leqslant 5)$； (5) $E(X)$； (6) $D(X)$.

4. 设某人的手机在一天中收到的短信数 X 服从下面的分布.

k	0	1	2	3	4	5	6	7	8
p_k	0.01	0.06	0.16	0.25	0.25	0.17	0.07	0.02	0.01

(1) 计算 $E(X)$； (2) 计算 $D(X)$.

5. 某人每天打出 k 次电话的概率 p_k 如下：

k	0	1	2	3	4	5	6	7	8
p_k	0.01	0.02	0.07	0.17	0.25	0.25	0.16	0.06	0.01

如果每打一个电话的话费是 0.3 元，计算他每天平均花多少钱打电话，方差是多少.

6. 一批产品有 100 件，其中含有 4 件次品，从中随机抽取 9 件. 计算：

(1) 这 9 件产品都是次品的概率；

(2) 这 9 件产品都是正品的概率；

(3) 这 9 件产品中有 6 件正品、3 件次品的概率；

(4) 这 9 件产品中平均有多少件正品；

(5) 这 9 件产品中平均有多少件次品.

7. 在一副标准扑克的 52 张(去掉两张王牌后)中任取五张，计算概率：

(1) 五张都是草花； (2) 两张草花，三张黑桃；

(3) 两张草花，两张黑桃，一张红桃； (4) 五张中有两张红桃.

8. 在一副标准扑克的 52 张(去掉两张王牌后)中任取八张，计算概率：

(1) 得到两张 2； (2) 得到三张 2；

(3) 得到两张 2 和三张 A； (4) 八张牌是同花顺.

9. 设 $X\sim N(0, 1)$，查表计算：

(1) $P(X<2.2)$； (2) $P(-1.8<X\leqslant 2)$.

10. 下表是 1976—1977 年美国佛罗里达州 20 个地区的人命案中对被告的 326 个宣判结果. 对是否判死刑和被告是否为黑人进行独立性分析. (已知 $P(\chi^2\geqslant 2.71)=0.1$)

被告＼判刑	死刑	非死刑	合计
白人	19	141	160
黑人	17	149	166
合计	36	290	326

11. 新兴电脑公司有 8 名产品推销员，其工作年限与年推销金额数据如下表：

推销员编号	1	2	3	4	5	6	7	8
工作年限 x(年)	3	2	10	5	8	4	4	8
年推销金额 y(万元)	22	18	95	40	75	45	40	78

(1) 求年推销金额 y 关于工作年限 x 的相关系数；

(2) 求年推销金额 y 关于工作年限 x 的线性回归方程；

(3) 分别估计工作年限为 7 年和 11 年时的年推销金额.

温故而知新

12. 投掷两枚骰子和两枚硬币. 计算骰子的点数和是 8，两枚硬币正面都朝上的概率.

13. 20 个同学来自不同的地方.

(1) 计算他们的生日都在 7 月份的概率；

(2) 他们中有 5 个同学的生日在 7 月份的概率；

(3) 他们中平均有多少个同学的生日在 7 月份？

14. 对于事件 A，B，证明：$P(A\cap B)=P(A)P(B|A)$.

15. 如果 $P(A\cap B)=P(A)P(B)$，证明：$P(A\cap\bar{B})=P(A)P(\bar{B})$.

16. 6 个人同时向一目标射击，每个人击中目标的概率都是 0.7. 计算：

(1) 目标没被击中的概率；　　(2) 目标被击中 3 次的概率；

(3) 目标平均被击中几次；　　(4) 目标被击中次数的标准差.

17. 公共汽车一共要停靠 9 站，在每站停车的概率是 0.9. 平均需要停车多少次？

18. 甲、乙二人进行羽毛球比赛. 如果每局甲胜的概率是 0.6，计算在他们的 10 局比赛中，甲期望赢多少局？方差是多少？

19. 在投掷两枚硬币时，如果出现两个正面，甲得 1 分，否则输 0.25 分. 用 X 表示他的得分，计算 X 的概率分布和数学期望.

20. 鱼塘中只有 800 条鲤鱼和 200 条草鱼，每条鱼被打捞的可能性相同. 捞鱼者一天打捞上来 45 条鱼. 计算这 45 条鱼中平均有多少条鲤鱼，方差是多少.

21. 对一个新产品的科研和开发需要投资 50 万元，开发成功可以获利 5 000 万元. 如果开发成功的概率是 0.6，计算投资的平均收益和标准差.

22. 电视台在公布招聘多名节目主持人后，收到了 3 份符合条件的申请简历. 根据以往的经验，每个符合条件的人员在面试时能够被录用的概率是 0.6，设每个申请者能否被录用是相互独立的，面试这 3 个申请者时，计算：

(1) 3 个申请者都被录用的概率； (2) 只有 1 个申请者被录用的概率.

23. 一个随机抽取的样本包括 110 名女士和 90 名男士，女士中约有 9%是左利手，男士中约有 11%是左利手. 基于这些数据，你认为在样本所代表的总体中，左利手与性别有关吗？为什么？

24. 在彩色显像中，根据以往的经验，形成染料光学密度 y 与析出银的光学密度 x 之间存在关系式 $y=ae^{-\frac{b}{x}}(b>0)$. 现对 y 与 x 同时作 10 次观测，获得 10 对数据如下表，试根据表中数据，求出 a 与 b 的估计值.

编　号	1	2	3	4	5	6	7	8	9	10
x	0.05	0.06	0.07	0.10	0.14	0.20	0.25	0.31	0.38	0.43
y	0.10	0.14	0.23	0.37	0.59	0.79	1.00	1.12	1.19	1.25

上下而求索

25. 每门高炮击中飞机的概率是 0.8，要以 99%的把握击中飞机，需要几门高炮？

26. 某跳高运动员跳过 1.8 m 的概率是 $p=0.8$. 不计每次试跳消耗的体能，计算：

(1) 他首次试跳成功的概率； (2) 第 3 次试跳才首次成功的概率；

(3) 要以 99%的概率跳过 1.8 m，至少需要试跳多少次.

27. 某车间为了规定工时定额，需要确定加个某零件所花费的时间，为此作了次实验，得到的数据如下：

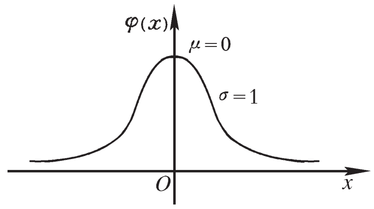
零件的个数 x(个)	2	3	4	5
加工的时间 y(h)	2.5	3	4	4.5

(1) 求出 y 关于 x 的线性回归方程；

(2) 试预测加工 10 个零件需要多少时间？

附录 1

标准正态分布表



a	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	a
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	0.0
0.1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753	0.1
0.2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141	0.2
0.3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517	0.3
0.4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879	0.4
0.5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224	0.5
0.6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549	0.6
0.7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852	0.7
0.8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133	0.8
0.9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389	0.9
1.0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621	1.0
1.1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830	1.1
1.2	8849	8869	8888	8907	8925	8944	8962	8980	8997	90147	1.2
1.3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774	1.3
1.4	91924	92073	92220	92364	92507	92647	92785	92922	93056	93189	1.4
1.5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408	1.5
1.6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449	1.6
1.7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327	1.7
1.8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062	1.8
1.9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670	1.9
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169	2.0
2.1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574	2.1
2.2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899	2.2
2.3	98928	98956	98983	$9^{2}0097$ *	$9^{2}0358$	$9^{2}0613$	$9^{2}0863$	$9^{2}1106$	$9^{2}1344$	$9^{2}1576$	2.3
2.4	$9^{2}1802$	$9^{2}2024$	$9^{2}2240$	$9^{2}2451$	$9^{2}2656$	$9^{2}2857$	$9^{2}3053$	$9^{2}3244$	$9^{2}3431$	$9^{2}3613$	2.4
2.5	$9^{2}3790$	$9^{2}3963$	$9^{2}4132$	$9^{2}4297$	$9^{2}4457$	$9^{2}4614$	$9^{2}4766$	$9^{2}4915$	$9^{2}5060$	$9^{2}5201$	2.5
2.6	$9^{2}5339$	$9^{2}5473$	$9^{2}5604$	$9^{2}5731$	$9^{2}5855$	$9^{2}5975$	$9^{2}6093$	$9^{2}6207$	$9^{2}6319$	$9^{2}6427$	2.6
2.7	$9^{2}6533$	$9^{2}6636$	$9^{2}6736$	$9^{2}6833$	$9^{2}6928$	$9^{2}7020$	$9^{2}7110$	$9^{2}7197$	$9^{2}7282$	$9^{2}7365$	2.7
2.8	$9^{2}7445$	$9^{2}7523$	$9^{2}7599$	$9^{2}7673$	$9^{2}7744$	$9^{2}7814$	$9^{2}7882$	$9^{2}7948$	$9^{2}8012$	$9^{2}8074$	2.8
2.9	$9^{2}8134$	$9^{2}8193$	$9^{2}8250$	$9^{2}8305$	$9^{2}8359$	$9^{2}8411$	$9^{2}8462$	$9^{2}8511$	$9^{2}8559$	$9^{2}8605$	2.9
3.0	$9^{2}8650$	$9^{2}8694$	$9^{2}8736$	$9^{2}8777$	$9^{2}8817$	$9^{2}8856$	$9^{2}8893$	$9^{2}8930$	$9^{2}8965$	$9^{2}8999$	3.0
3.1	$9^{3}0324$	$9^{3}0646$	$9^{3}0957$	$9^{3}1260$	$9^{3}1553$	$9^{3}1836$	$9^{3}2112$	$9^{3}2378$	$9^{3}2636$	$9^{3}2886$	3.1
3.2	$9^{3}3129$	$9^{3}3363$	$9^{3}3590$	$9^{3}3810$	$9^{3}4024$	$9^{3}4230$	$9^{3}4429$	$9^{3}4623$	$9^{3}4810$	$9^{3}4991$	3.2
3.3	$9^{3}5166$	$9^{3}5335$	$9^{3}5499$	$9^{3}5658$	$9^{3}5811$	$9^{3}5959$	$9^{3}6103$	$9^{3}6242$	$9^{3}6376$	$9^{3}6505$	3.3
3.4	$9^{3}6631$	$9^{3}6752$	$9^{3}6869$	$9^{3}6982$	$9^{3}7091$	$9^{3}7197$	$9^{3}7299$	$9^{3}7398$	$9^{3}7493$	$9^{3}7585$	3.4
3.5	$9^{3}7674$	$9^{3}7759$	$9^{3}7842$	$9^{3}7922$	$9^{3}7999$	$9^{3}8074$	$9^{3}8146$	$9^{3}8215$	$9^{3}8282$	$9^{3}8347$	3.5
3.6	$9^{3}8409$	$9^{3}8469$	$9^{3}8527$	$9^{3}8583$	$9^{3}8637$	$9^{3}8689$	$9^{3}8739$	$9^{3}8787$	$9^{3}8834$	$9^{3}8879$	3.6
3.7	$9^{3}8922$	$9^{3}8964$	$9^{4}0039$	$9^{4}0426$	$9^{4}0799$	$9^{4}1158$	$9^{4}1504$	$9^{4}1838$	$9^{4}2159$	$9^{4}2468$	3.7
3.8	$9^{4}2765$	$9^{4}3052$	$9^{4}3327$	$9^{4}3593$	$9^{4}3848$	$9^{4}4094$	$9^{4}4331$	$9^{4}4558$	$9^{4}4777$	$9^{4}4988$	3.8
3.9	$9^{4}5190$	$9^{4}5385$	$9^{4}5573$	$9^{4}5753$	$9^{4}5926$	$9^{4}6092$	$9^{4}6253$	$9^{4}6406$	$9^{4}6554$	$9^{4}6696$	3.9
4.0	$9^{4}6833$	$9^{4}6964$	$9^{4}7090$	$9^{4}7211$	$9^{4}7327$	$9^{4}7439$	$9^{4}7546$	$9^{4}7649$	$9^{4}7748$	$9^{4}7843$	4.0
4.1	$9^{4}7934$	$9^{4}8022$	$9^{4}8106$	$9^{4}8186$	$9^{4}8263$	$9^{4}8338$	$9^{4}8409$	$9^{4}8477$	$9^{4}8542$	$9^{4}8605$	4.1
4.2	$9^{4}8665$	$9^{4}8723$	$9^{4}8778$	$9^{4}8832$	$9^{4}8882$	$9^{4}8931$	$9^{4}8978$	$9^{5}0226$	$9^{5}0655$	$9^{5}1066$	4.2
4.3	$9^{5}1460$	$9^{5}1837$	$9^{5}2199$	$9^{5}2545$	$9^{5}2876$	$9^{5}3193$	$9^{5}3497$	$9^{5}3788$	$9^{5}4066$	$9^{5}4332$	4.3
4.4	$9^{5}4587$	$9^{5}4831$	$9^{5}5065$	$9^{5}5288$	$9^{5}5502$	$9^{5}5706$	$9^{5}5902$	$9^{5}6089$	$9^{5}6268$	$9^{5}6439$	4.4
4.5	$9^{5}6602$	$9^{5}6759$	$9^{5}6908$	$9^{5}7051$	$9^{5}7187$	$9^{5}7318$	$9^{5}7442$	$9^{5}7561$	$9^{5}7675$	$9^{5}7784$	4.5
4.6	$9^{5}7888$	$9^{5}7987$	$9^{5}8081$	$9^{5}8172$	$9^{5}8258$	$9^{5}8340$	$9^{5}8419$	$9^{5}8494$	$9^{5}8566$	$9^{5}8634$	4.6
4.7	$9^{5}8699$	$9^{5}8761$	$9^{5}8821$	$9^{5}8877$	$9^{5}8931$	$9^{5}8983$	$9^{6}0320$	$9^{6}0789$	$9^{6}1235$	$9^{6}1661$	4.7
4.8	$9^{6}2067$	$9^{6}2453$	$9^{6}2822$	$9^{6}3173$	$9^{6}3508$	$9^{6}3827$	$9^{6}4131$	$9^{6}4420$	$9^{6}4696$	$9^{6}4958$	4.8
4.9	$9^{6}5208$	$9^{6}5446$	$9^{6}5673$	$9^{6}5889$	$9^{6}6094$	$9^{6}6289$	$9^{6}6475$	$9^{6}6652$	$9^{6}6821$	$9^{6}6981$	4.9

* “$9^{2}0097$” 表示 “990097”，依次类推.

附录 2

数学词汇中英文对照表

（按词汇所在页码的先后排序）

中文名	英　文　名	页　码
排列	permutation	12
阶乘	factorial	12
组合	combination	19
德·摩尔根	De Morgan	36
试验组	experimental group	45
对照组	control group	45
费歇尔	Fisher	46
萨凯	Salk	47
样本点	sample outcome	49
样本空间	sample space	49
独立	independent	55
随机变量	random variable	60
离散型随机变量	discrete random variable	61
伯努利	Bernoulli	63
二项式	binomial	64
数学期望	mathematical expectation	70
均值	mean	70
方差	variance	73
标准差	standard deviation	73
伽利略	Galileo	77
高斯	Gauss	77
正态分布	normal distribution	80
标准正态分布	standardized normal distribution	80