经全国中小学教材审定委员会2004年初审通过

普通高中课程标准实验教科书

数学 3（必修）

SHUXUE

主　　编　严士健　王尚志
副 主 编　张饴慈　李延林　张思明
本册主编　严士健　吕建生　李亚玲
编写人员　（按 姓 氏 笔 画 排 序）
　　　　　王尚志　王建波　吕建生
　　　　　严士健　李亚玲　周长春

北京师范大学出版社
·北 京·

前　　言

你们将进入更加丰富多彩的数学世界.

你们将学到更多重要和有趣的数学知识、技能及应用.

你们将更多地感受到深刻的数学思想和方法.

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

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

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

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

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

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

研究性学习是我们特别提倡的. 在教材中强调了问题提出，抽象概括，分析理

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

“课题学习”引导同学们递进地思考问题，充分动手实践，是需要完成的部分.

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

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

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

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

我们的联系方式是：北京师范大学出版社基础教育分社（100875），010-58802811，58802790.

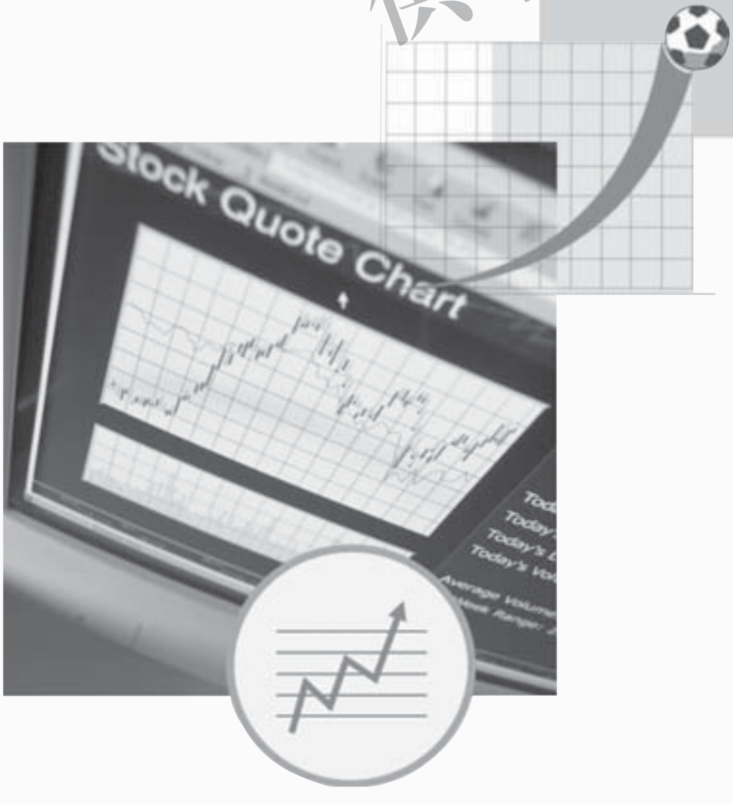
目　录

第一章 统 计

统计是研究如何合理收集、整理、分析数据的学科，它可以帮助我们从数据中提取有用的信息，并为制定决策提供依据.

在日常生活中，我们可以随处看到各种各样的数据，如股市的涨跌、天气的变化、人口的分布、人才的供求……而面对这些纷繁复杂的数据，从中获取所需要的信息是非常重要的. 并且，在更多的时候，我们常常需要根据不同的问题要求，采取有效的途径收集数据，再根据所获得的数据，提取有价值的信息来解决问题. 这些都是统计学所要讨论的问题.

在某种意义上，经典数学有助于提高演绎能力，统计学有助于提高归纳能力. 本章将通过对具体问题的分析，帮助我们体会用统计思想解决问题的过程.



在抽象的意义下，
　　一切科学都是数学；
在理性的世界里，
　　所有的判断都是统计学.

——著名统计学家
C. R. Rao

§1　从普查到抽样

问题提出

下面呈现的是 2000 年我国第五次人口普查关于人口分布情况的一部分统计数据和一些新闻.

- 人口普查显示我国男女婴出生比未超过国际标准（2001 年 4 月 28 日《北京青年报》）
- 计划生育 30 年全国少生 3 亿(新华网北京 2001 年 4 月 23 日电(记者沈路涛))
- 人口普查数据显示:我国东西部人口密度之比为 9∶1（2001 年 4 月 18 日《北京青年报》）
- 人口普查登记质量抽查表明漏登率为 1.81%（中新网北京 2001 年 3 月 28 日消息）
- 我国男女性别比为 106.74∶100（新华网北京 2001 年 3 月 28 日电）
- 第五次全国普查结果:我国总人口达到 12.95 亿（新华网北京 2001 年 3 月 28 日电）
- 武汉一人口普查员劳累过度以身殉职（2000 年 11 月 23 日《长江日报》）

参考下面的阅读材料,针对上述统计数据和新闻回答问题:

（1）人口普查对一个国家的发展有什么作用？依据上面所提供的信息,你能举例说明吗？

（2）根据上面的有关信息,我国第五次人口普查中漏登的人数大约是多少？你对人口普查中的漏登率是如何认识的？

(3) 你对上面“武汉一人口普查员劳累过度以身殉职”的报道有何看法?

小资料

普查是指一个国家或一个地区专门组织的一次性大规模的全面调查,目的是为了详细地了解某项重要的国情、国力.

普查主要有两个特点:

(1) 所取得的资料更加全面、系统;

(2) 主要调查在特定时段的社会经济现象总体的数量.

目前,我国所进行的普查主要有:人口普查、农业普查、工业普查、第三产业普查、基本单位普查等.

人口普查是一个规模宏大的政府工程.我国第五次人口普查从 2000 年 11 月 1 日开始.普查之前,国家动用大量人力、物力做普查前的准备工作,先后有 600 多万人参与这次普查,到 2001 年 3 月 28 日国家统计局发布第一号公报,历时 5 个月,此后还有大量的数据处理工作.这次人口普查要对我国 2000 年的人口总量、人口分布、民族人口、城乡人口、受教育程度、迁移流动、就业状况、人口住房等多方面情况进行统计分析,为国家的宏观决策提供可靠的依据.

普查是一项非常艰巨的工作,它要对所有的对象进行调查.人口普查是我国的一个重要统计活动,每隔一定的年限要进行一次.

当普查的对象很少时,普查无疑是一项非常好的调查方式.当普查的对象很多时,普查的工作量就很大,要耗费大量的人力、物力与财力,并且组织工作繁重、时间长.更值得注意的是,在很多情况下,普查工作难以实现.

例 医生是如何检验人的血液中血脂的含量是否偏高的?你觉得这样做的合理性是什么?

解 医生在检验人的血液中血脂含量是否偏高时,通常是抽取少量的血样进行检验,然后由此作出推断,认为这个人的血液状况基本如此.

医生在检验时是不可能将一个人的血液都抽出来进行普查的.

思考交流

我们来看下面几个问题,并与同学进行交流:

(1) 某工厂要检查一个批次(10 万个)螺钉的质量,请你给检验

员提供一些检验方法上的建议，并说明你的理由；

(2) 某灯管厂要对一个批次灯管的寿命(使用时间)进行检验，你认为应当怎样进行检验？说明你的理由.

抽象概括

从对上述问题的讨论中可以知道，由于检验对象的量很大，或检验对检验对象具有破坏性，所以采用普查的方法有时是行不通的. 通常情况下，从调查对象中按照一定的方法抽取一部分，进行调查或观测，获取数据，并以此对调查对象的某项指标作出推断，这就是**抽样调查**，其中，调查对象的全体称为**总体**，被抽取的一部分称为**样本**.

抽样调查与普查相比有很多的优点，最突出的有两点：

(1) 迅速、及时

要调查一个国家就业状况，如果采用普查，需要很长的时间去收集与处理数据，等统计数据出来之后，这个国家的就业状况又发生了一定的变化；而抽样调查则能及时地得到统计数据，对一个国家的宏观调控起到一定的指导作用.

(2) 节约人力、物力和财力

抽样调查面对的调查对象少，会节省更多的人力、物力与财力.

由于调查的对象少，因此可以对每个被调查个体的信息了解得更为详细，从而使获取的数据更加科学、可靠.

在抽样中，如果每个样本被抽到的概率相同，这种抽样称为随机抽样. 在“思考交流”提出的两个问题中，如果采用随机抽样，就可以使样本的螺钉质量与这一批次的 10 万个螺钉质量近似相同，使样本的灯泡的使用寿命与同批次的灯泡使用寿命情况近似相同，从而保证样本反映的状况和总体的状况近似相同. 我们在后续的章节将逐一学习简单随机抽样、系统抽样、分层抽样等基本的抽样方法.

抽象概括

在处理问题时，人们对随机性的把握是非常困难的，因为每个人在做选择的时候，常常会受到各种各样的主观因素的影响. 因此，在统计抽样时，为了做到随机性，人们常常会寻找一些方法来避免人为主观因素的影响.

在统计活动中，尤其是大型的统计活动，人们常常需要对统计方案进行仔细的设计，以避免一些外界因素的干扰. 通常需要确定调查

的对象、调查的方法与策略(如果是问卷调查,需要精心设计问卷),需要精心设计前期的准备工作和收集数据的方法,然后对数据进行分析(包括统计数据的汇总与呈现),得出统计推断.

思考交流

(1) 若要调查你所在的城市有多少人有酗酒或吸毒的历史,请你思考一下如何进行抽样,在抽样的过程中应当注意什么问题,并与同学交流自己的想法.

(2) 若要调查你所居住的小区在周末上班的人数,可在周末时打电话到某些人家里进行调查.你认为将会得到什么样的答案?在这个题目中,总体与样本分别是什么?将思考结果与同学进行交流.

练 习

1. 在抽样调查中,应当注意什么问题?
2. 什么样的样本才具有代表性?用具体的例子加以说明.
3. 如果现在有一项调查,调查你们学校学生的家庭平均月收入情况,那么你会怎样做?将你的想法写成调查方案,并与同学交流你的调查方案与想法,看看是否有需要改进的地方.

习 题 1—1

1. 要调查某种药物的疗效,请你思考如何进行抽样,在抽样的过程中应当注意什么问题.
2. 某学校要做一项社会调查,调查我国国民的收入状况.某同学调查了他所在班级同学父母亲的收入情况,然后作了一个分析报告.你认为这样的分析报告有什么问题?在这个题目中,总体与样本分别是什么?
3. 调查我国人口普查的历史与有关数据,体会收集数据和资料的过程,写成调查报告,并与同学进行交流.(可参阅《光明日报》2000-10-23)

阅读材料

选 举 的 预 测

历史上美国有一家有名的刊物《文学文摘》，它在预测 1936 年美国总统选举结果时发生了重大的失误.当年有两位候选人，一位是民主党的罗斯福，一位是共和党的兰登.当时，大多数民意测验、新闻机构和政治观察家都预测罗斯福会获胜，但《文学文摘》与众不同，它预言兰登会以 57%∶43%的优势战胜罗斯福，这在当时产生了很大的反响.而实际情况是：罗斯福以 62%∶38%的压倒性优势战胜兰登当选总统.正是由于这个重大失误，这家杂志不久即宣告破产.

当时，《文学文摘》做出这个预测，并非一种主观臆断，而是做了大样本(240 万份调查问卷)的民意测验做出的.为何根据这么大的样本却没有得到满意的结果呢？问题不是因为采用了抽样的方法而没有采用普查的方法造成的，而是出在样本的挑选上.

在《文学文摘》调查之后，盖洛普公司做过多次关于总统大选结果的民意测验，不仅与实际结果接近(当选预测无误，得票率估计略有误差)，而且调查的人数也不过几千人，但预测结果却相当成功.盖洛普公司测验的具体情况如表 1-1.

表 1 1

年份	样本量	当选者	盖洛普预测得票率	实际得票率
1952	5 385	艾森豪威尔	51.0%	55.4%
1956	8 144	艾森豪威尔	59.5%	57.5%
1960	9 015	肯尼迪	51.0%	50.1%
1964	6 625	约翰逊	64.0%	61.3%
1968	4 414	尼克松	43.0%	43.5%
1972	3 689	尼克松	62.0%	61.8%
1976	3 439	卡特	49.5%	51.1%

当时，《文学文摘》的访问对象是从电话号码簿和俱乐部会员名册上选取的.但在 1936 年，美国家庭的电话尚未普及，只有1 100万部左右，尤其是有条件参加俱乐部的人，大多是经济上较富有、政治上保守、倾向于共和党的选民，这就造成了显著的系统误差.当时，正值 1929 年至 1933 年美国经济大萧条过去不久，较贫困的阶层人数不少，失业人数多达 900 万.与兰登相比，罗斯福要推行的新政较多地考虑了这些人的利益，由于这些选民的意见没有在样本中得到体现，以致《文学文摘》的预测产生如此大的偏差.

另外，原计划《文学文摘》要访问的对象是1 000万人，这样的样本可能会好些，而实际回收的调查问卷只有 240 万份.较富有的人，对当时现实抱比较满意态度，做出回答的可能性要大些，这个倾向有利于共和党，这是另外一个系统误差.事实证明：《文学文摘》向芝加哥地区三分之一的登记选民发了问题单，有 20%的人做了回答，其中半数以上有利于兰登.但实际结果是：在芝加哥是以 2∶1 的优势有利于罗斯福.

说明：本阅读材料根据陈希孺的《机会的数学》中的有关内容进行改编.

§2 抽样方法

抽样的方法非常多，每个抽样方法都有各自的优越性与局限性，针对不同的问题应当选择适当的抽样方法. 下面我们就来讨论三种比较典型与常用的抽样方法：简单随机抽样、分层抽样与系统抽样.

2.1 简单随机抽样

问题 1 若要调查你所在的学校学生最喜欢的体育活动情况，应当怎样抽样？

> **注 意**
> 简单随机抽样的总体是有限的.

对于这个问题，可以从你所在学校的所有学生中，随机地抽取一些学生，然后对抽取的对象进行调查. 在抽取的过程中，要保证每个学生被抽到的概率相同. 这样的抽样方法叫作**简单随机抽样**. 这是抽样中一个最基本的方法. 简单随机抽样的具体实施方法是这样的：在总体的 N 个个体中机会均等地抽取第一个，然后在剩下的 $(N-1)$ 个个体中机会均等地抽取第二个……最后在剩余的 $[N-(n-1)]$ 个个体中机会均等地抽取第 n 个. 用这种抽样方法，每一个被抽到的概率是相同的.

我们知道，要能做到绝对随机地抽取样本是非常困难的，因此，在抽样的过程中，要尽可能地避免人为因素的影响，通常采用抽签法和随机数法（利用工具产生随机数）.

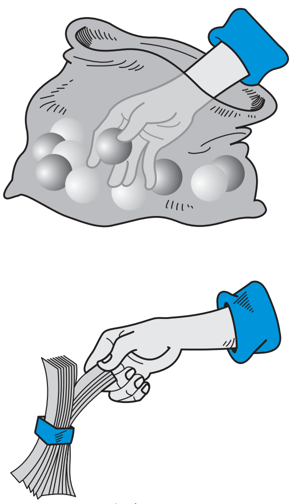


图 1-1

1. 抽签法

先把总体中的 N 个个体编号，并把编号写在形状、大小相同的签上（签可以是纸条、卡片或小球等，如图 1-1），然后将这些号签放在同一个箱子里均匀搅拌. 每次随机地从中抽取一个，然后将号签均匀搅拌，再进行下一次抽取. 如此下去，直至抽到预先设定的样本数.

抽签法的实施步骤：

(1) 给调查对象群体中的每个对象编号；

(2) 准备“抽签”的工具，实施“抽签”；

(3) 对样本中每一个个体进行测量或调查.

抽签的方法我们是熟悉的，在初中阶段我们做过很多关于摸球等方面的游戏. 但是，当总体容量很大时，操作起来比较麻烦.

思考交流

对于问题 1，请用抽签法设计一个调查方案，并与同学进行交流.

2. 随机数法

把总体中的 N 个个体依次编上 $0,1,\cdots,N-1$ 的号码，然后利用工具（转盘或摸球、随机数表、科学计算器或计算机）产生 $0,1,\cdots,N-1$ 中的随机数，产生的随机数是几，就选几号个体，直至抽到预先规定的样本数.

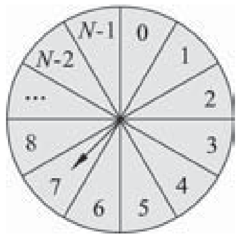


图 1-2

利用转盘产生随机数是比较简单的，就是将转盘分成 N 等份（如图 1-2），分别标上整数 $0,1,\cdots,N-1$，转动转盘，指针指向的数字是几就取几号个体.

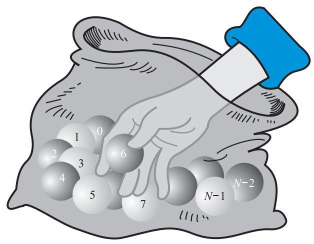


图 1-3

利用摸球产生随机数也是一样，就是将 N 个形状、大小、质地完全相同的球分别标上整数 $0,1,\cdots,N-1$，放入一个不透明的容器中进行摸球（如图 1-3），摸到几号球，就抽取相应标号的个体，然后将摸出的球放回，充分搅匀，准备下一次摸球.

这种产生随机数的方法，大家都比较熟悉，我们在初中时做过很多这方面的游戏. 这种方法简便易行，尤其是当总体容量不大时. 这种方法的缺点是，当总体容量非常大时，制作转盘和进行摸球就比较困难了.

在上面的摸球试验中，取 $N=10$，进行摸球试验，每摸出一个球，就将球的号码按行、列的方式依次写在一个空白表中，这样就形成了一个**随机数表**. 历史上，第一个随机数表《随机抽样数》是由英国统计学家梯培特（Tippett）制作，并于 1927 年出版，本书摘录了一部分（转录自陈希孺《机会的数学》）.

信息技术建议

利用计算机和科学计算器可以方便地产生随机数，具体的操作步骤参见本节的“信息技术应用”栏目.

表 1 2

7816	6572	0802	6314	0702	4369	9728	0198
3204	9243	4935	8200	3623	4869	6938	7481
2976	3413	2841	4241	2424	1985	9313	2322
8303	9822	5888	2410	1158	2729	6443	2943
5556	8526	6166	8231	2438	8455	4618	4445
2635	7900	3370	9160	1620	3882	7757	4950
3211	4919	7306	4916	7677	8733	9974	6732
2748	6198	7164	4148	7086	2888	8519	1620
7477	0111	1630	2404	2979	7991	9683	5125
5379	7076	2694	2927	4399	5519	8106	8501

续表

9264	4607	2021	3920	7766	3817	3256	1640
5858	7766	3170	0500	2593	0545	5370	7814
2889	6628	6757	8231	1589	0062	0047	3815
5131	8186	3709	4521	6665	5325	5383	2702
9055	7196	2172	3207	1114	1384	4359	4488
7900	5870	2602	8813	5509	4324	0030	4750
3693	9212	0557	7369	7162	9568	1312	9438
0380	3338	0138	4560	4230	6496	3806	0347
0246	4469	9719	8316	1285	0357	2389	2390
7266	0081	6897	2851	4666	0620	4596	3400
9312	4779	5737	8918	4550	3994	5573	9229
6111	6098	0965	7352	6847	3034	9977	3770
2310	4476	9148	0679	2662	2062	0522	9234
9826	8857	8675	6642	5471	8820	4308	2105
6703	8248	6064	6962	0053	8188	6494	4509
1110	9486	6533	3954	1944	1516	1682	3404
9651	1456	5613	0357	4244	3341	9605	3567
8350	5728	4338	0824	7899	1307	5814	8688
6982	5126	7736	3383	6215	3441	8578	2277
6490	7644	7085	8361	5662	4141	9877	3747
8570	2150	8140	4355	5321	2548	0208	7543
9169	0408	4353	6122	8913	9930	4169	6032
2127	0162	6176	4969	8185	9312	8748	8575
8090	9872	1968	0263	0081	2662	6831	3106
2959	9011	1448	4346	7019	8148	1557	8400

由于随机数中的每个数字都是随机产生的，因此我们可以利用随机数表来产生随机数. 如果总体的编号超过一位数，比如是两位数，那么，我们可以一次选取其中的两列（或两行），或选取两个数字，组成一个两位数.

例 1 总体由 80 个个体组成，利用随机数表随机地选取 10 个样本.

解 具体做法如下：

第一步 将总体中的每个个体进行编号：00，01，…，79；

第二步 由于总体是一个两位数的编号，每次要从随机数表中选取两列组成两位数.

从随机数表中任意一个位置，比如，从表 1-2 中第 6 列和第 7 列这两列的第 4 行开始选数，由上至下分别是：

82，52，90，91，19，11，07，60，76，62，18，19，87，21，33，46，08，…

其中 19 重复出现，82，90，91，87 超过 79，不能选取. 这样选取的 10 个样本的编号分别为：

52，19，11，07，60，76，62，18，21，33.

思考交流

在上面选数的过程中，我们是从表 1-2 中第 6 列和第 7 列这两列的第 4 行开始由上至下顺序进行选数的. 你还有哪些选法？与同学进行交流.

利用计算机、科学计算器或图形计算器产生随机数，对于总体容量很大时尤其显得有用. 一些计算器或计算机内部都有一种产生随机数的工具，我们只需做一些简单的处理就可以得到比较理想的结果.

信息技术应用

利用信息技术产生随机数

利用计算机和科学计算器都可以帮助我们很方便地得到随机数，下面我们分别来介绍它们是如何产生随机数的.

一般地，在计算机内部(如 Excel)都有一个随机函数 Rand( )产生 0～1 的随机数. 每次只要运行一次这个函数就产生一个随机数. 我们要做的工作，就是要将产生的 0～1 的实数变换成 0，1，2，…，$N-1$这 N 个整数，如果产生的随机数是 0，就认为取第 N 号个体，否则产生的随机数就是个体所对应的编号.

例如，一个 $N=100$ 的总体，当 Rand( )$=0.321\ 454\cdots$时，我们只要做一个变换 Int(100 * Rand( ))(相当于把所取的随机数乘 100，然后取其整数部分)，就把所产生的随机数变换成为 32. 一般地，我们只要对随机函数和取整函数做一个复合就可以了，也就是取 Int(N * Rand( )).

利用科学计算器也可以帮助我们很方便地产生随机数. 一般科学计算器内部都有一个随机函数 RAND，它可以产生 0～1 的随机数. 有些科学计算器还提供了随机函数 RANDI，它可以产生任意两个整数之间的随机整数. 其一般步骤如下(用不同型号的计算器产生随机数，步骤略有不同)：按键$\boxed{\text{PRB}}$，选 RANDI，再输入整数 a 和 b($a<b$)，即可产生 a 和 b 之间的随机整数. 另外，利用图形计算器也同样可以产生随机数. 产生随机数的方法是多种多样的，同学们可以根据自己的需要，选择不同的方式来得到随机数.

练 习

班级正准备组织某项课外活动，同学们都积极地要求参加. 班长想从班级中任意抽取 10 名同学参加，请用随机数表产生随机数的方法设计抽样方案，并与同学进行交流.

2.2 分层抽样与系统抽样

1. 分层抽样

问题 2 某市有大型、中型与小型的商店共 1 500 家，它们的家数之比为 1∶5∶9. 要调查商店的每日零售额情况，要求抽取其中的 30 家商店进行调查，应当采用怎样的抽样方法?

在这个问题中，商店有大型、中型和小型之分，商店的每日零售额直接受到商店规模的影响，题目中数据可以看出，大型商店占所有商店的比例为$\frac{1}{1+5+9}=\frac{1}{15}$，总数为 $1\ 500\times\frac{1}{15}=100$. 为保证样本的代表性，在所抽取的 30 家商店中，大型商店所占的比例也应为$\frac{1}{15}$，数量为 $30\times\frac{1}{15}=2$，所以应从 100 家大型商店中抽出 2 个代表. 同理，从 500 家中型商店中抽出 10 个代表，从 900 家小型商店中抽出 18 个代表. 因此，我们要对每个类型的商店分别进行抽样.

将总体按其属性特征分成若干类型(有时称作层)，然后在每个类型中按照所占比例随机抽取一定的样本. 这种抽样方法通常叫作**分层抽样**，有时也称为类型抽样.

> **注 意**
> 在每个层中进行抽样时，大多数情况下是采用简单随机抽样，有时也会用到其他的抽样方法，这要根据问题的需要来决定.

例 2 某地农田分布在山地、丘陵、平原、洼地不同的地形上，要对这个地区的农作物产量进行调查，应当采用什么抽样方法?

解 由于不同类型的农田之间的产量有较大差异，应当采用分层抽样的方法，对不同类型的农田按其占总数的比例来抽取样本.

例 3 某公司有 1 000 名员工，其中：高层管理人员为 50 名，属于高收入者；中层管理人员为 150 名，属于中等收入者；一般员工为 800 名，属于低收入者. 要对这个公司员工的收入情况进行调查，欲抽取 100 名员工，应当怎样进行抽样?

解 我们可以采用分层抽样的方法，按照收入水平分成三层：高收入者、中等收入者、低收入者. 从题中数据可以看出，高收入者为 50

名，占所有员工的比例为$\frac{50}{1\ 000}=5\%$，为保证样本的代表性，在所抽取的 100 名员工中，高收入者所占的比例也应为 5%，数量为 $100\times5\%=5$，所以应抽取 5 名高层管理人员. 同理，抽取 15 名中层管理人员、80 名一般员工，再对收入状况分别进行调查.

抽象概括

如果知道某一类个体在总体中占有的百分比，那么按照这个比例抽取这类个体，这样的抽样会提高对总体推断的精度.

2. 系统抽样

上面我们讨论了两类抽样方法，它们是基本的抽样方法，在社会生活与生产中应用非常广泛. 但当总体容量和样本容量都很大时，无论是采用分层抽样或简单随机抽样，都是非常麻烦的. 系统抽样就是为了解决这个问题.

系统抽样是将总体中的个体进行编号，等距分组，在第一组中按照简单随机抽样抽取第一个样本，然后按分组的间隔（称为**抽样距**）抽取其他样本. 这种抽样方法有时也叫**等距抽样**或**机械抽样**.

例 4　某工厂平均每天生产某种机器零件大约 10 000 件，要求产品检验员每天抽取 50 件零件，检查其质量状况. 假设一天的生产时间中生产机器零件的件数是均匀的，请你设计一个调查方案.

解　我们可以采用系统抽样，按照下面的步骤设计方案.

第一步　按生产时间将一天分为 50 个时间段，也就是说，每个时间段大约生产$\frac{10\ 000}{50}=200$件产品. 这时，抽样距就是 200.

第二步　将一天中生产出的机器零件按生产时间进行顺序编号. 比如，第一个生产出的零件就是 0 号，第二个生产出的零件就是 1 号等.

第三步　从第一个时间段中按照简单随机抽样的方法，抽取一件产品，比如是 k 号零件.

第四步　顺序地抽取编号分别为下面数字的零件：$k+200$，$k+400$，$k+600$，…，$k+9\ 800$，这样就抽取了容量为 50 的一个样本.

例 5　某装订厂平均每小时大约装订图书 362 册，要求检验员每小时抽取 40 册图书，检查其质量状况. 请你设计一个调查方案.

解　我们可以采用系统抽样，按照下面的步骤设计方案.

第一步　把这些图书分成 40 个组，由于$\frac{362}{40}$的商是 9，余数是 2，所以每个组有 9 册书，还剩 2 册书. 这时，抽样距就是 9.

第二步　先用简单随机抽样的方法从这些书中抽取 2 册书，不进行检验.

第三步　将剩下的书进行编号，编号分别为 0，1，…，359.

第四步　从第一组（编号分别为 0，1，…，8）的书中按照简单随机抽样的方法，抽取 1 册书，比如说，其编号为 k.

第五步　顺序地抽取编号分别为下面数字的书：$k+9$，$k+18$，$k+27$，…，$k+39\times 9$，这样就抽取了容量为 40 的一个样本.

分析理解

调查某年级学生的身高情况，利用系统抽样的方法，样本容量为 50. 这个年级共分 50 个组，每个组都是 8 名学生，他们的座次是按照身高自矮到高进行编排的. 李立是这样做的，抽样距是 8，按照每个小组的座次进行顺序编号. 你觉得这样抽取的样本具有代表性吗？

分析　假设这个年级的学生是这样编号（这个编号也代表他们的身高）的：

第一组　$a_1<a_2<a_3<a_4<a_5<a_6<a_7<a_8$；

第二组　$b_1<b_2<b_3<b_4<b_5<b_6<b_7<b_8$；

……

第五十组　$m_1<m_2<m_3<m_4<m_5<m_6<m_7<m_8$.

如果按照李立的抽样方法，比如在第一组抽到了 8 号，也就是 a_8，那么所抽取的样本分别为 a_8，b_8，…，m_8. 显然，这样的样本不具有代表性，它们代表的身高偏高.

思考交流

（1）在上述调查中，梁强是这样做的，抽样距是 8，将全年级的学生按照身高自矮到高的顺序进行编排，然后进行抽样. 你觉得这样抽取的样本具有代表性吗？说明你的理由，与同学进行交流.

（2）为了调查某路口一个月的交通流量情况，王俊采用系统抽样的方法，抽样距为 7，从每周中随机抽取一天，他正好抽到的是星期一，这样他每个星期一对这个路口的交通流量进行了统计，最后做出调查报告. 你认为王俊这样的抽样方法有什么问题？应当怎样改进？与同学进行交流.

抽象概括

在抽样时，如果总体的排列存在明显的周期性或者事先是排好序的，那么利用系统抽样的方法进行抽样时将会产生明显的偏差，因为这样抽取的样本不具有代表性.

习　题　1—2

1. 简述分层抽样和系统抽样各自的优缺点.
2. 某超市为了做好售货员的安排工作，要调查一下附近居民一周平均到超市购物的次数. 试着用三种抽样方法分别设计一个调查方案并进行调查，同时用初中学过的统计图表将调查的结果表示出来，为超市管理人员提供一个合理的建议.
3. 请你设计一个利用抽样方法进行人口调查的方案.
4. 某试验田中有试验苗 30 000 株，植保站为了调查此试验田防虫情况，要抽查试验田中每株试验苗的虫卵数量. 假设虫卵在试验田中的分布是均匀的，请设计一个调查方案.

§3 统 计 图 表

问题提出

前面我们已经介绍了收集数据的一些方法. 一旦数据被收集后，我们总希望从中找出所需要的信息. 但通过收集得到的数据一般比较多，我们无法直接将其包含的全部信息统统理解并加以表达，这样就需要对这些数据进行适当的整理、分析，将其转化为可以直接利用的形式，并从中获取相应的信息，以便帮助我们作出恰当的决策.

统计图表就是表达和分析数据的重要工具，它不仅可以帮助我们从数据中获取有用的信息，还可以帮助我们直观、准确地理解相应的结果. 我们在初中阶段已经学习过条形统计图、扇形统计图和折线统计图，在这里，我们将结合一些案例进一步对统计图表的特点和选用加以具体分析.

问题 1 我们对 50 人的智商情况进行了调查，如果按照区间 $[80,85)$，$[85,90)$，…，$[115,120)$ 进行分组，得到的分布情况如图1-4 所示.

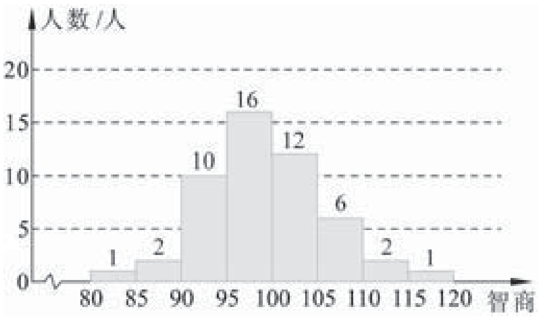


图 1-4

(1) 有多少人的智商在 90～105[1]？

(2) 有多少人的智商低于 100？

(3) 有多少人的智商不低于 100？

你还能从图中获得其他的信息吗？

[1] 90～105 是指区间$[90,105)$. 如无特殊说明，本章中类似的情形均指左闭右开区间.

显然，在 50 人中，有 38 人的智商在 90～105，29 人的智商低于 100，21 人的智商不低于 100.

问题 2　下面是关于某个总体包含的所有学生的身高分布的几种表述，其中哪一种表述反映的总体信息较多？

（1）身高在 160 cm 以下的学生数占 50%，不低于 160 cm 的学生数占 50%（如图 1-5(a)）.

（2）身高在 150 cm 以下、150 ～160 cm、不低于 160 cm 的学生数分别占 10%，40%，50%（如图 1-5(b)）.

（3）身高在 150 cm 以下、150 ～160 cm、160～170 cm、不低于 170 cm 的学生数分别占 10%，40%，40%，10%（如图1-5(c)）.

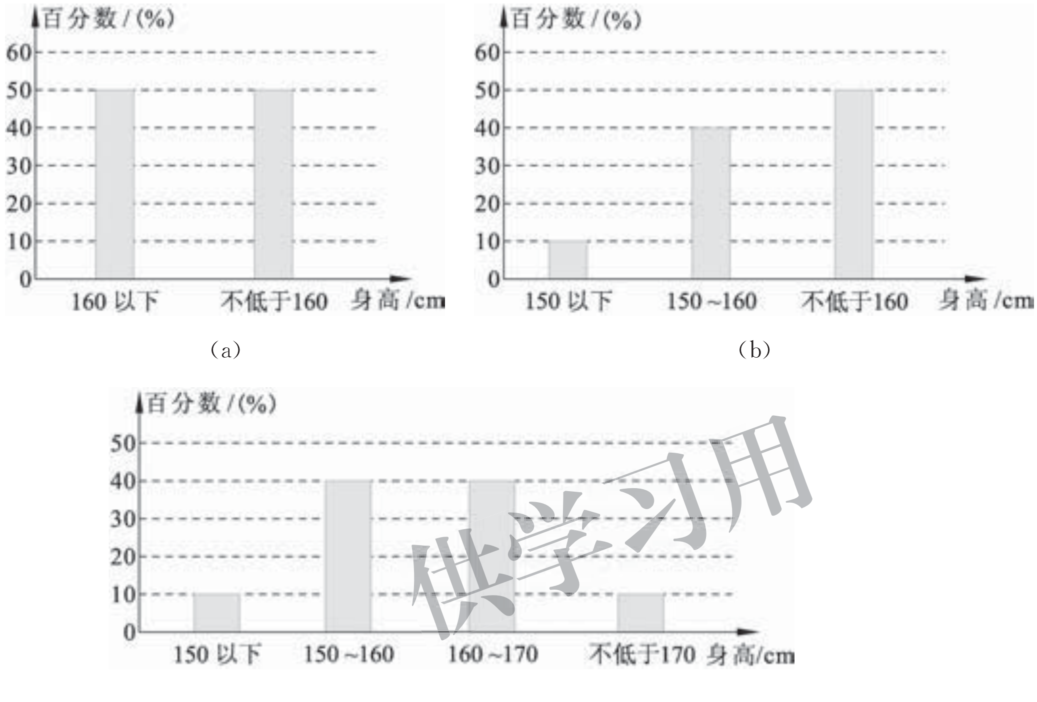


图 1-5

从该总体包含的所有学生的身高分布的几种表述（包括文字和统计图）来看，不难发现：(1)～(3)，反映的总体信息依次增多.

在实际问题中，我们常常根据问题的需要来选择不同的表达方式，以获得对数据适当的了解.

分析理解

王华所在的班级共有 50 位同学，图 1-6 是全班学生按照一定的高矮顺序在操场上拍的照片.

图 1-7 是王华按照图 1-6 所画的，观察此图，你会发现：右边同学的身高总是比左边的高，并且在每一列中后面同学的身高总是比前面的高.

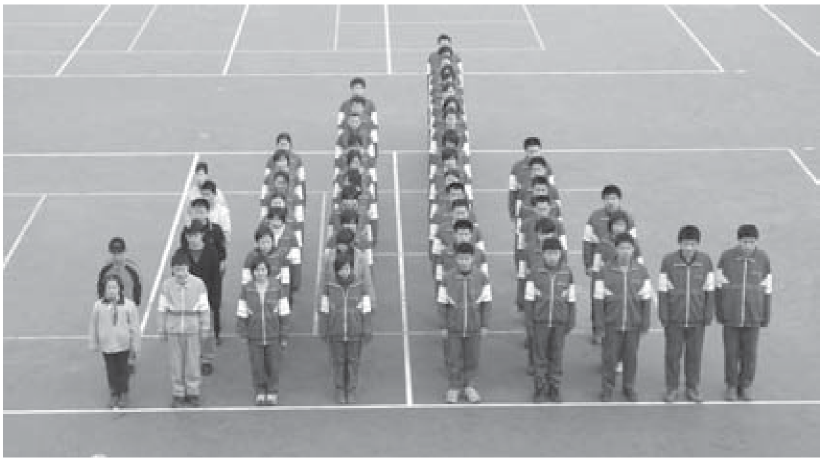


图 1-6

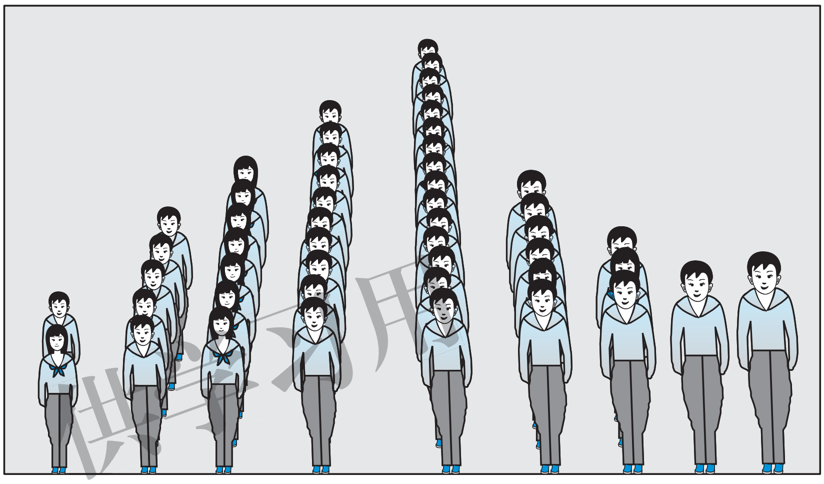


图 1-7

王华的同学徐良在拍照之前，已经根据所统计的班上同学的身高，绘制了**条形统计图**(如图 1-8 所示).

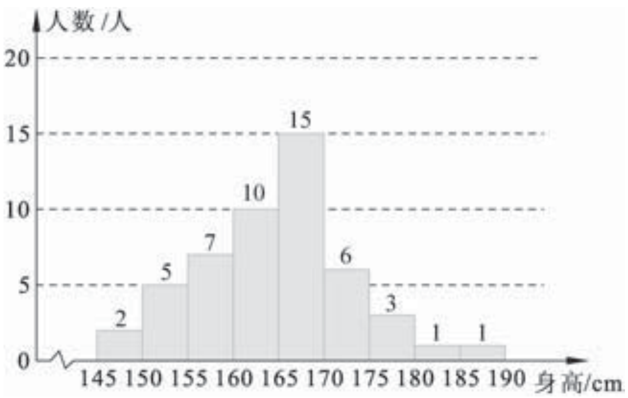


图 1-8

图 1-7 的本质虽然与图 1-8 是一致的，但它却更形象、更直观，所要表达的信息也一目了然.

思考交流

2001 年上海市居民支出构成情况如表 1-3 所示：

表 1 3

食品	衣着	家庭设备用品及服务	医疗保健	交通和通信	教育文化娱乐服务	居住	杂项商品和服务
39.4%	5.9%	6.2%	7.0%	10.7%	15.9%	11.4%	3.5%

有两位同学分别用**折线统计图**（图 1-9）和**扇形统计图**（图 1-10）表示了上面的数据.

2001 年上海市居民支出构成情况折线统计图

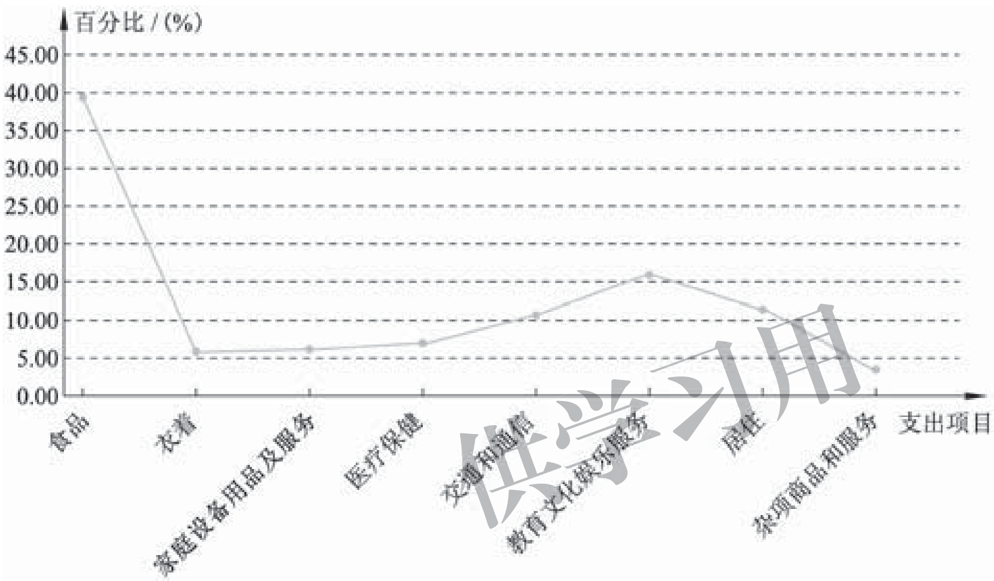


图 1-9

2001 年上海市居民支出构成情况扇形统计图

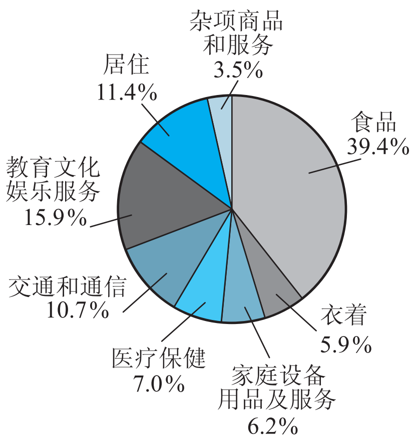


图 1-10

观察并比较这两种统计图：

（1）它们分别有什么特点？你觉得哪种统计图更合适？

（2）你还有其他表示 2001 年上海市居民支出构成情况的方法吗？与同学进行交流.

练 习 1

下面是 1990～2002 年我国每 10 万人口中各级学校平均在校生的人数. 请用不同的统计图将其表示出来，你能从统计图中获取哪些信息？

年份	高等学校[1]	高中阶段[2]	初中阶段[3]	小学	幼儿园
1990	326	1 337	3 426	10 707	1 725
1991	304	1 355	3 465	10 502	1 907
1992	313	1 365	3 518	10 413	2 072
1993	376	1 448	3 599	10 656	2 190
1994	433	1 293	3 681	10 819	2 219
1995	457	1 610	3 945	11 010	2 262
1996	470	1 780	4 180	11 273	2 208
1997	482	1 905	4 289	11 435	2 058
1998	504	1 978	4 408	11 287	1 944
1999	594	2 032	4 656	10 855	1 864
2000	723	2 018	4 969	10 335	1 782
2001	931	2 021	5 161	9 937	1 602
2002	1 146	2 283	5 240	9 525	1 595

❶高等学校包括：普通高等学校和成人高等学校.

❷高中阶段包括：普通高中、职业高中、普通中专、技工学校、成人中专和成人高中.

❸初中阶段包括：普通初中和职业初中.

数据来源：http://www.stats.gov.cn

我们已经知道，不同的统计图都有各自的特点和用途. 在面对实际问题时，我们常常会根据不同的需要，选择合适的统计图表来进行表示.

例 有关部门从甲、乙两个城市所有的自动售货机中分别随机抽取了 16 台，记录下上午 8:00～11:00 各自的销售情况（单位：元）：

甲：18，8，10，43，5，30，10，22，6，27，25，58，14，18，30，41；

乙：22，31，32，42，20，27，48，23，38，43，12，34，18，10，34，23.

你能用不同的方式分别表示上面的数据吗？

解 从上面的数据不易直接看出各自的分布情况，为此，我们可以先将以上数据按照不同的方式进行表示.

上述的数据可以用如图 1-11 所示的图形来表示，横线下面的数

字表示销售额的十位数，上面的数字分别表示各自销售额的个位数.

甲

8					
8					
8	4	7			
6	0	5	0	3	
5	0	2	0	1	8
0	1	2	3	4	5

乙

		7	8		
		3	4		
	8	3	4	8	
	2	2	2	3	
	0	0	1	2	
0	1	2	3	4	5

图 1-11

也可以用条形统计图（图 1-12）将上图进行简化：

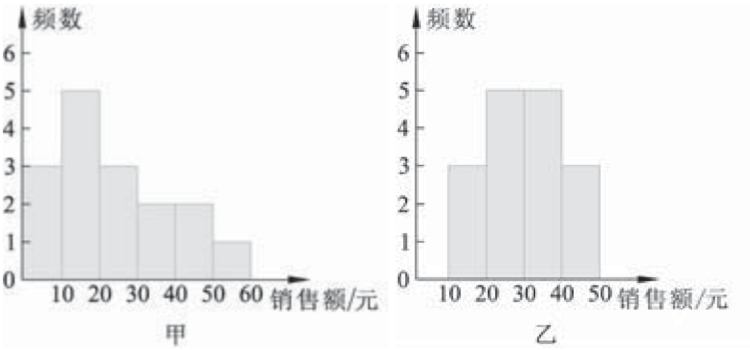


图 1-12

上述数据中乙的销售情况还可以用图 1-13 来表示，其中，竖线左边的数字分别表示各自销售额的十位数，右边的数字表示销售额的个位数.

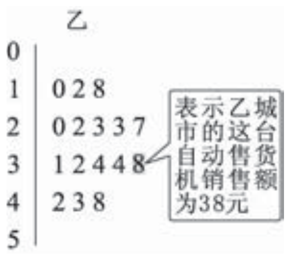


图 1-13

用同样的方式也可以表示甲的销售情况. 为了方便比较，我们仍用图 1-13 中竖线左边的数字表示甲销售额的十位数，在其左边再画一条竖线，用竖线左边的数字分别表示甲销售额的个位数（如图1-14）.

甲		乙
8 6 5	0	
4 0 8 0 8	1	0 2 8
7 5 2	2	0 7 3 2 3
0 0	3	8 1 2 4 4
1 3	4	2 8 3
8	5	

图 1-14

思考交流

在上例中：

（1）从哪一种统计图中能看出甲的销售额中有 25 元这一数据？哪一种统计图反映了收集到的全部数据信息？哪一种统计图损失了部分数据信息？

（2）如果收集到的数据很多，例如有 100 个，你认为用哪一种统计图更能直观地反映这些数据分布的大致情况？

（3）你还有其他表示这些数据的方法吗？

抽象概括

我们通常把图 1-13 这样的统计图叫作茎叶图. 用茎叶图表示数据有个突出的优点，茎叶图不但可以随时记录数据，还可以在记录的过程中随时观察到数据的一些特征，从而及时对数据进行分析.

在对数据进行分析和整理时要根据实际需要恰当地选用统计图. 当数据量很大时一般选用条形图，它能更直观地反映数据分布的大致情况，并能清晰地表示出各个区间的具体数目，但是条形图会损失数据的部分信息. 折线图能够表现出数据的变化趋势，但不能直观反映数据的分布情况. 扇形统计图可以直观地反映出各种情况所占的比例，但是看不出具体数据的多少. 茎叶图可以动态地表现数据的分布特征，但不适合数据量比较大的情况.

信息技术应用

利用计算机画统计图

本节讲了统计表的制作，利用计算机电子表格软件（如 Excel）可以帮助我们很方便地制作统计图，你不妨试一试.

打开电子表格，把要处理的数据输入表中，屏幕上会出现如图 1-15 所示的表；然后选中你输入的所有数据（格子高亮显示），此时应再按工具栏中的“图表向导”按钮，屏幕会提醒你选择需要制作

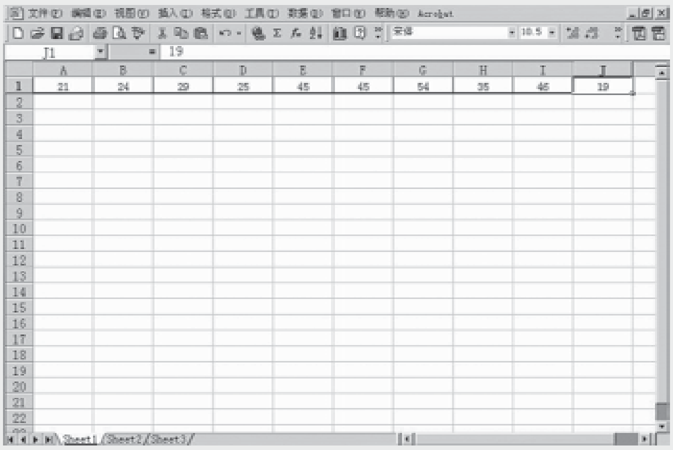


图 1-15

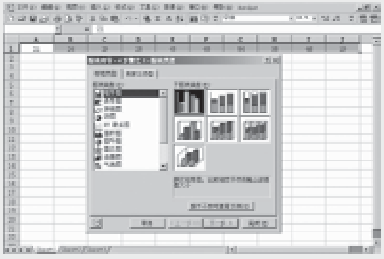


图 1-16

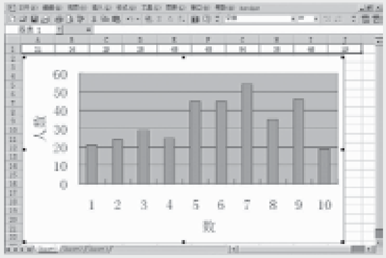


图 1-17

的统计图类型(如图 1-16),选择好类型后按“完成”,就会出现制作好的统计图.

例如,我们要将图 1-15 所示的表制成条形统计图,只要选择“条形图”,就会出现图 1-17;同样,我们也可以方便地得到折线统计图和扇形统计图.

如果遇到困难,别忘了点击“帮助”按钮.

动手实践

统计你所在的班级全体同学的右手一拃长(如图 1-18),用你喜欢的方式把统计得到的信息表示出来,并与同学进行交流.

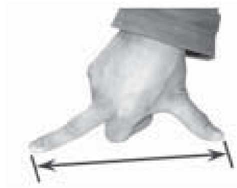


图 1-18

练 习 2

为了了解各自受欢迎的程度,甲、乙两个网站分别随机选取了 14 天,记录下上午 8:00～10:00 各自的点击量:

甲:73,24,58,72,64,38,66,70,20,41,55,67,8,25;

乙:12,37,21,5,54,42,61,45,19,6,19,36,42,14.

你能用茎叶图表示上面的数据吗? 你认为甲、乙两个网站哪个更受欢迎?

习 题 1—3

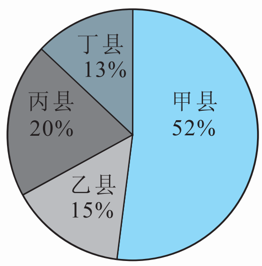
1. 如图为某个人口为 300 000 人的城市的人口分布:

(1) 甲县有多少人?

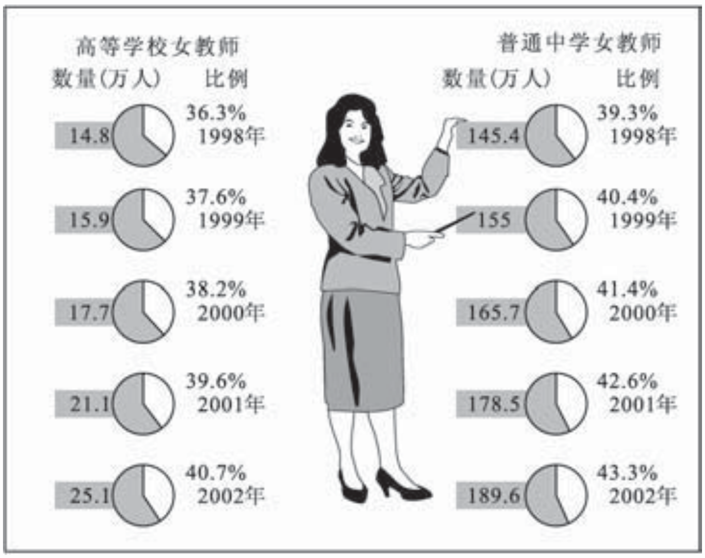
(2) 乙县和丁县共有多少人?

(3) 甲县和丙县相差多少人?

2. 下面给出了 1998 年至 2002 年我国高等学校和普通中学女教师的人数及其相应的比例,请你用自己的方式表示其中的数据.



(第 1 题)



(第 2 题)

3. 1994 年美国家庭收入(单位:美元)的百分比分布如下表:

15 000 以下	15 000～25 000	25 000～35 000	35 000～50 000	50 000～75 000	不少于 75 000
22.7%	16.7%	14.2%	16.3%	16.5%	13.6%

请将上面的数据用统计图表示出来,你觉得哪种统计图更合适?

4. 在波士顿马拉松比赛中,前 30 名男运动员的成绩(单位:分)的排列如下:129,130,130,133,134,135,136,136,138,138,138,141,141,141,142,142,142,142,143,143,143,143,143,144,144,145,145,145,145,145.

请用适当的方式把上面的信息表示出来.

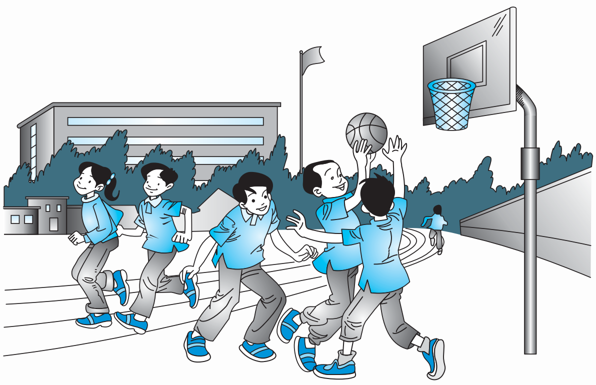
5. 某赛季甲、乙两名篮球运动员每场比赛得分(单位:分)情况如下:

甲的得分:12,15,24,25,31,31,36,36,37,39,44,49,50;

乙的得分:8,13,14,16,23,26,28,33,38,39,51.

你能用不同的方式分别表示此赛季甲、乙得分的情况吗?

6. 调查你们班同学每周进行体育锻炼的时间,并用你喜欢的方式把统计得到的信息表示出来.



§4　数据的数字特征

数据的信息除了通过前面介绍的各种统计图表来加以整理和表达之外，还可以通过一些统计量来表述，也就是将多个数据“加工”为一个数值，使这个数值能够反映这组数据的某些重要的整体特征.

4.1　平均数、中位数、众数、极差、方差

我们在初中阶段已经学习过平均数、中位数、众数、极差、方差等，它们都是一些统计量，反映了数据的集中趋势或离散程度. 我们常常根据问题的需要而选择不同的统计量来表达数据的信息. 在这一节中，我们将结合一些案例进一步对数字特征的特点加以具体分析.

例 1　某公司员工的月工资情况如表 1-4 所示.

表 1-4

月工资/元	8 000	5 000	4 000	2 000	1 000	800	700	600	500
员工/人	1	2	4	6	12	8	20	5	2

(1) 分别计算该公司员工月工资的平均数、中位数和众数.

(2) 公司经理会选取上面哪个数来代表该公司员工的月工资情况？税务官呢？工会领导呢？

解　(1) 该公司员工的月工资平均数为

$$\frac{8\,000\times1+5\,000\times2+4\,000\times4+2\,000\times6+1\,000\times12+800\times8+700\times20+600\times5+500\times2}{1+2+4+6+12+8+20+5+2}$$

$$\approx1\,373,$$

即月工资平均数为 1 373 元；把所有员工的月工资数按从大到小的顺序依次排列，中间的两个数均为 800，因此月工资中位数为 800 元；在所有员工的月工资数中，700 出现了 20 次，出现的次数最多，因此月工资众数为 700 元.

(2) 公司经理为了显示本公司员工的收入高，采用月工资平均数 1 373元作为月工资的代表；而税务官希望取月工资中位数 800 元，以便知道目前的所得税率对该公司的多数员工是否有利；工会领导则主张用月工资众数 700 元作为代表，因为每月拿 700 元的员工数最多.

信息技术建议

利用计算机电子表格软件（如 Excel）和科学计算器可以方便地计算一组数据的数字特征，具体的操作步骤参见本节的“信息技术应用”栏目.

如果你要应聘该公司，你会如何看待公司员工的收入情况？

例 2　在上一节中，从甲、乙两个城市随机抽取的 16 台自动售货机的销售额可以用茎叶图表示，如图 1-19 所示：

（1）甲、乙两组数据的中位数、众数、极差分别是多少？

（2）你能从图中分别比较甲、乙两组数据平均数和方差的大小吗？

甲		乙
8 6 5	0	
8 8 4 0 0	1	0 2 8
7 5 2	2	0 2 3 3 7
0 0	3	1 2 4 4 8
3 1	4	2 3 8
8	5	

图 1-19

解　(1)观察茎叶图，我们不难看出：甲城市销售额的中位数为 20，众数为 10，18，30，极差为 53；乙城市销售额的中位数为 29，众数为 23，34，极差为 38.

(2)从茎叶图中我们可以看出：甲城市的销售额分布主要在茎叶图的上方且相对较散，而乙城市的销售额分布则相对集中在茎叶图的中部. 由此，我们可以估计：甲城市销售额的平均数比乙城市的小，而方差比乙城市的大.

抽象概括

平均数、中位数和众数刻画了一组数据的集中趋势，极差、方差刻画了一组数据的离散程度. 它们作为一组数据的代表各有优缺点，也各有各的用处，从不同的角度出发，不同的人会选取不同的统计量来表达同一组数据的信息.

4.2　标准差

甲、乙两台机床同时生产直径是 40 mm 的零件. 为了检验产品质量，从两台机床生产的产品中各抽取 10 件进行测量，结果如表 1-5 所示 .

表 1 5

甲机床生产的零件直径/mm	40.0	39.8	40.1	40.2	39.9	40.0	40.2	39.8	40.2	39.8
乙机床生产的零件直径/mm	40.0	40.0	39.9	40.0	39.9	40.1	40.1	40.1	40.0	39.9

经过简单计算可以得出：甲、乙两台机床生产的这 10 件产品直径的平均数都是 40 mm，但从表 1-5 中的数据不难发现，甲生产的产品尺寸波动幅度比乙大，我们用折线统计图（图 1-20）可以直观地表示出这两组数据的离散情况：

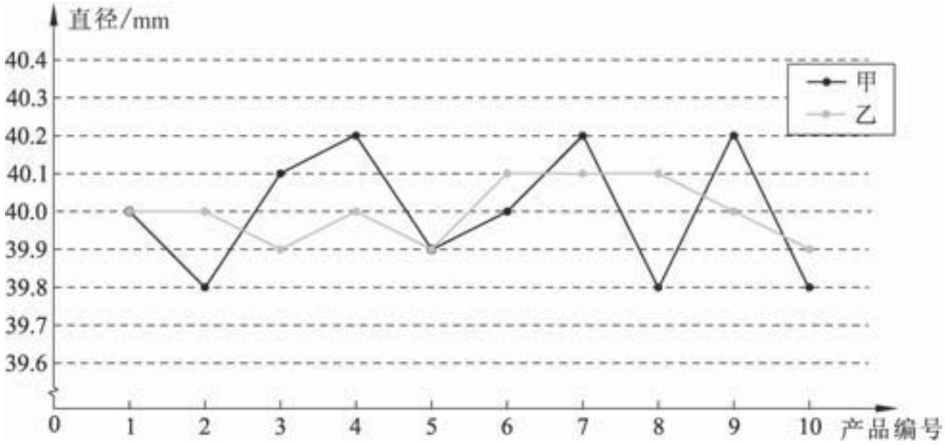


图 1-20

你能选择适当的数分别表示这两组数据的离散程度吗?

下面给出了几种不同的方法：

方法 1(极差)

甲：$40.2-39.8=0.4$(mm)，乙：$40.1-39.9=0.2$(mm)；

方法 2(方差)

甲：$s_{甲}^2=\frac{1}{10}[(40-40)^2+(39.8-40)^2+\cdots+(39.8-40)^2]$

$=0.026(\text{mm}^2)$，

乙：$s_{乙}^2=\frac{1}{10}[(40-40)^2+(40-40)^2+\cdots+(39.9-40)^2]$

$=0.006(\text{mm}^2)$；

方法 3

甲：$\frac{1}{10}(|40-40|+|39.8-40|+\cdots+|39.8-40|)$

$=0.14$(mm)，

乙：$\frac{1}{10}(|40-40|+|40-40|+\cdots+|39.9-40|)$

$=0.06$(mm)；

方法 4

甲：$\frac{1}{10}(|40-40|^3+|39.8-40|^3+\cdots+|39.8-40|^3)$

$=0.005(\text{mm}^3)$，

乙：$\frac{1}{10}(|40-40|^3+|40-40|^3+\cdots+|39.9-40|^3)$

$=0.000\ 6(\text{mm}^3)$.

信息技术建议

在以后的数据处理中，尽量使用科学计算器、计算机软件和图形计算器等工具帮助计算.

抽象概括

在以前，我们已经学习了刻画数据离散程度的统计量，如极差与方差等. 刻画数据离散程度的度量，其理想形式应满足以下三条原则：

(1) 应充分利用所得到的数据，以便提供更确切的信息；

(2) 仅用一个数值来刻画数据的离散程度；

(3) 对于不同的数据集，当离散程度大时，该数值亦大.

极差显然不满足上面的第一条原则，它只是利用了数据中最大和最小的两个值，而且对极值过于敏感. 但由于只涉及两个数据，便于得到，所以极差在实际中也经常应用.

方差虽然满足上面的三条原则，然而它有局限性：方差的单位是原始观测数据的单位的平方，而刻画离散程度的一种理想度量应当具有与原始数据相同的单位. 解决这个局限性的一种方法是取方差的正的平方根 $s=\sqrt{s^2}=\sqrt{\dfrac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}}$，称为**标准差**. 标准差的单位与原始测量单位相同，在统计中，我们通常用标准差来刻画数据的离散程度.

例 3 分别计算上面从甲、乙两台机床抽取的 10 件产品直径的标准差.

解 从数据很容易得到甲、乙两台机床生产的这 10 件产品直径的平均数

$$\bar{x}_{甲}=\bar{x}_{乙}=40(\text{mm}).$$

我们分别计算它们直径的标准差：

$$s_{甲}=\sqrt{[(40-40)^2+(39.8-40)^2+\cdots+(39.8-40)^2]/10}$$
$$=0.161(\text{mm}),$$

$$s_{乙}=\sqrt{[(40-40)^2+(40-40)^2+\cdots+(39.9-40)^2]/10}$$
$$=0.077(\text{mm}).$$

由上面的计算可以看出：甲、乙两台机床生产的产品直径的平均数相同，而甲机床生产的产品直径的标准差为 0.161 mm，比乙机床的标准差 0.077 mm 大，说明乙机床生产的零件要更标准些，即乙机床的生产过程更稳定一些.

动手实践

教师记时，从一个确定的时刻起叫“开始”，经过一段时间之后再叫“停止”，让全班同学估计这一段时间到底有多长.

(1) 请记录下你估计的时间.

(2) 汇总全班同学的估计结果，并用适当的统计图表将其表示出来.

(3) 计算这些数据的平均数和标准差，并与确切的时间值作比较，你发现了什么？

增加“开始”和“停止”之间的时间间隔，重复刚才的试验，试验结果与刚才有什么不同？

信息技术应用

利用信息技术计算数字特征

我们经常需要计算一组数据的数字特征(如平均数、标准差)，以便更好地分析数据的信息. 实际上，利用计算机和科学计算器都可以帮助我们很方便地得到数据的数字特征. 下面我们就以本节例 3 中的数据为例，分别来介绍它们是如何进行计算的.

利用计算机电子表格软件(如 Excel)计算一组数据的数字特征，其一般步骤如下：

打开电子表格，把要处理的数据输入表中，例如，我们将例 3 中甲机床生产产品的直径输入表格中；选中一个空白格，作为给出答案的位置；然后点击工具栏中的“＝”，在“＝”这一行的最前面会出现一个可下拉的菜单：点击这个菜单，选中“AVERAGE”，拖动鼠标，将刚才输入的数据全选中，此时，在 Number 1 这一格中会显示这列数据所在的范围(从 A_1 到 A_{10})，按一下“确定”，计算机就会在刚才选中给出答案的位置显示出“平均数”的值为 40；改选“STDEVP”，可得“标准差”的值为 0.161 245 155.（如图 1-21、图 1-22 所示）.

利用科学计算器计算一组数据的数字特征的一般步骤如下.(用不同型号的计算器进行计算，步骤略有不同)[1]

(1) 打开科学计算器，进入统计状态

按 $\boxed{\text{2nd}}$ $\boxed{\text{DATA}}$ 键(“STAT”)，计算器上显示三种选择：1—

[1] 以本章 §2“信息技术应用”中所示的科学计算器为例.

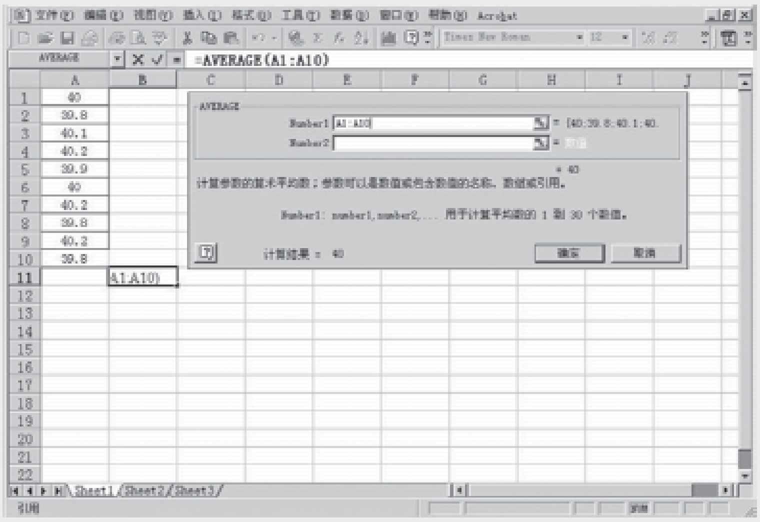


图 1-21

图 1-22

VAR(一个变量)、2—VAR(两个变量)、CLRDATA,利用$\boxed{\rightarrow}$键或$\boxed{\leftarrow}$键选择 1—VAR,并按$\boxed{\text{ENTER}}$键确认.

(2) 输入数据

按$\boxed{\text{DATA}}$键,输入第一个数据,再按$\boxed{\downarrow}$键并输入这个数据相应的次数,重复上述步骤,直至输入了所有的数据为止.

(3) 显示结果

按$\boxed{\text{STATVAR}}$键后,利用$\boxed{\rightarrow}$键或$\boxed{\leftarrow}$键选择 $\bar{x}$ 或 σ_x,屏幕上就会自动显示出这组数据的平均数或标准差,对于例 3 中乙机床生产产品的直径,通过计算可以很方便地得出:$\bar{x}=40$,
$\sigma_x=0.077\ 459\ 667$.

(4) 退出

运算结束后,可以按$\boxed{\text{2nd}}$ $\boxed{\text{DATA}}$键后选择 CLRDATA 来清除

所有数据但不退出统计状态；也可按 2nd STATVAR ENTER 键来清除所有数据并退出统计状态.

计算数据数字特征的方法是多种多样的，同学们可以根据自己的需要，选择不同的方式来得到数据的数字特征. 你不妨试一试，并与自己计算得到的结果作比较.

练 习

下表给出了小宇和志强在最近 8 场篮球比赛中的得分：

小宇	7	13	11	21	16	9	15	12
志强	12	9	13	10	26	6	10	16

他们在这 8 场比赛中的平均得分分别是多少？谁发挥得更稳定些？

习 题 1—4

1. 为了了解面包的销售情况，面包店随机选取了 24 个营业日，分别记录下每天销售的新鲜面包的数量(个)：

53 49 27 48 60 52 44 38 47 52 82 46

55 31 39 54 51 47 50 45 50 61 43 64

(1) 请用不同的方式分别表示上面的数据；

(2) 分别计算以上数据的平均数、中位数和众数；

(3) 根据以上结果，你认为该面包店每天应该生产多少新鲜面包？

2. 在 1976～1998 年的几届冬季奥运会中，男子、女子 1 500 米速滑的冠军成绩分别如下表所示：

年份	1976	1980	1984	1988	1992	1994	1998
男子	1′59.38″❶	1′55.44″	1′58.36″	1′52.06″	1′54.81″	1′51.29″	1′47.87″
女子	2′16.58″	2′10.95″	2′03.42″	2′00.68″	2′05.87″	2′02.19″	1′57.58″

❶1′59.38″表示 1 分 59.38 秒.

(1) 分别求出男子、女子 1 500 米速滑的冠军成绩的平均数和中位数；

(2) 分别求出男子、女子 1 500 米速滑的冠军成绩的标准差；

(3) 通过(1)(2)的计算，请用自己的语言描述这几届冬季奥运会男子、女子1 500米速滑的冠军成绩分别有什么特点.

§5 用样本估计总体

问题提出

从前面的分析可以知道，当研究一个对象时，如果能得到它们的全部数据（可以看作是总体），我们就可以直接从中分析总体的各种信息. 如人口普查得到的数据较为全面，从中可以很好地反映对象的重要信息. 但是，在实际问题中，总体的信息往往不能全部得到，因此我们需要进行抽样调查，从总体中抽取一部分作为样本，并用样本的各种信息来估计总体的情况，包括它的分布和基本数字特征.

5.1 估计总体的分布

如果把 18 岁男孩的身高作为总体，那么这个总体的分布是指不同身高范围在总体中占的比例（百分比）；如果把我国初生婴儿的性别作为总体，那么它的分布是指男女性别的比例. 一般地，总体分布是指总体中个体所占的比例.

统计的中心任务之一是推断总体的分布及分布特征. 为了达到此目的，就需要从总体中按照一定法则抽取若干个体进行观测或试验，以获得总体的信息. 当我们选择合理的抽样方法时，样本的分布和某些特征就与总体近似相同，我们可以用样本来估计总体.

如何通过样本来估计总体的分布情况呢？这就需要我们先将样本的分布情况表示出来，下面我们通过具体的例子来说明.

例 1895 年，在英国伦敦有 106 块男性头盖骨被挖掘出土. 经考证，这些头盖骨的主人死于 1665～1666 年的大瘟疫. 人类学家分别测量了这些头盖骨的宽度，数据如下所示（单位：mm）：

146	141	139	140	145	141	142	131	142	140	144	140
138	139	147	139	141	137	141	132	140	140	141	143
134	146	134	142	133	149	140	140	143	143	149	136
141	143	143	141	138	136	138	144	136	145	143	137
142	146	140	148	140	140	139	139	144	138	146	153

148　152　143　140　141　145　148　139　136　141　140　139
158　135　132　148　142　145　145　121　129　143　148　138
149　146　141　142　144　137　153　148　144　138　150　148
138　145　145　142　143　143　148　141　145　141

请你估计在 1665～1666 年，英国男性头盖骨宽度的分布情况.

解　这里，如果把总体看作是 1665～1666 年的英国男性头盖骨的宽度，那么我们就是要通过上面挖掘出土得到的样本信息，来估计总体的分布情况. 但从上面的数据很难直接估计出总体的分布情况，为此，我们可以先将以上数据按每个数据出现的频数和频率汇成表 1-6：

表 1 6

宽度/mm	频　数	频　率	宽度/mm	频　数	频　率
121	1	0.009	142	7	0.066
129	1	0.009	143	10	0.094
131	1	0.009	144	5	0.047
132	2	0.019	145	8	0.075
133	1	0.009	146	5	0.047
134	2	0.019	147	1	0.009
135	1	0.009	148	8	0.075
136	4	0.038	149	3	0.028
137	3	0.028	150	1	0.009
138	7	0.066	152	1	0.009
139	7	0.066	153	2	0.019
140	12	0.113	158	1	0.009
141	12	0.113			

从表格中，我们就能估计出总体大致的分布情况了，如在 1665～1666 年，英国男性头盖骨宽度主要在 136～149 mm，135 mm以下以及 150 mm 以上所占的比例相对较小等. 但是，这些关于分布情况的描述仍不够形象. 为了得到更为直观的信息，我们可以将表 1-6 中的数据按照下面的方式分组❶（如表 1-7），再画频数分布直方图（如图 1-23），用图中矩形的高度来反映频数.

表 1 7

宽 度 分 组(Δx_i)	频 数(n_i)	频 率(f_i)	$\dfrac{f_i}{\Delta x_i}$
120～125 mm	1	0.009	0.001 8
125～130 mm	1	0.009	0.001 8
130～135 mm	6	0.057	0.011 4
135～140 mm	22	0.208	0.041 6
140～145 mm	46	0.434	0.086 8
145～150 mm	25	0.236	0.047 2
150～155 mm	4	0.038	0.007 6
155～160 mm	1	0.009	0.001 8

❶ 当数据在 120 个以内时，通常按照数据的多少分成 5～12 组. 在实际操作中，一般要求各组的组距相等.

我们也可以用区间上矩形的面积来反映频率，得到图 1-24.

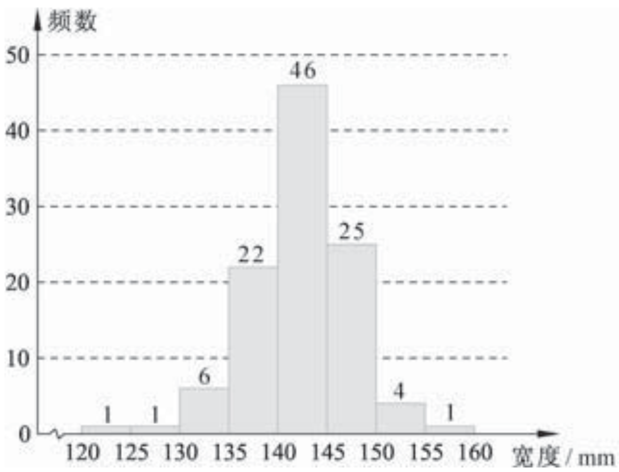


图 1-23

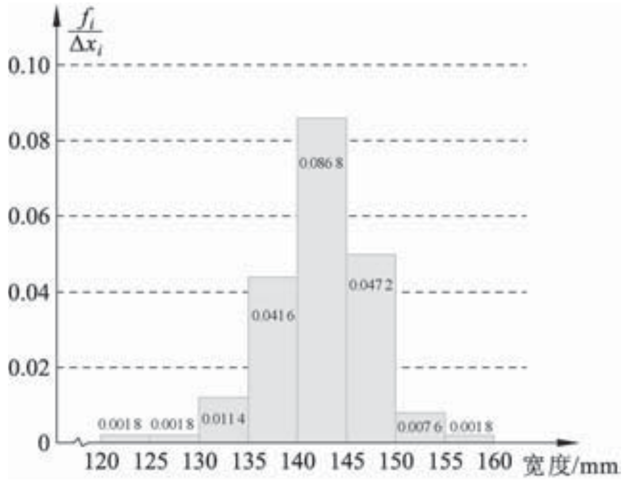


图 1-24

思考交流

观察图 1-24，你能知道：

(1) 头盖骨的宽度位于哪个区间的数据最多？

(2) 头盖骨的宽度在 140～145 mm 的频率约是多少？

(3) 头盖骨的宽度小于 140 mm 的频率约是多少？

(4) 头盖骨的宽度在 137～142 mm 的频率约是多少？

抽象概括

从以上的作图过程中我们知道：在图 1-24 中，每个小矩形的宽度为 Δx_i（分组的宽度），高为 $\frac{f_i}{\Delta x_i}$，小矩形的面积恰为相应的频率 f_i. 通常我们称这样的图形为**频率分布直方图**. 从图中可以得到，头盖骨的宽度落在各个宽度区间内的频率的大小（如：宽度在 140～145 mm 的头盖骨所占的频率为 43.4%，宽度在 137～142 mm 的头盖骨所占的频率为 29.8%等），这个频率的值就是该宽度区间所对应的频率直方图的面积. 图中所有小矩形的面积之和，也就是头盖骨的宽度落在各个宽度区间内的频率之和，等于 1.

另外，当样本量较大时，样本中落在每个区间内的样本数的频率会稳定于总体在相应区间内取值的概率. 因此，我们就可以用样本的频率分布直方图来估计总体在任意区间内取值的概率，也即总体的分布情况.

通常，在频率分布直方图中，按照分组原则，再在左边和右边各加

一个区间.从所加的左边区间的中点开始,用线段依次连接各个矩形的顶端中点,直至右边所加区间的中点,就可以得到一条折线(如图1-25),我们称之为**频率折线图**,有时也用它来估计总体的分布情况.

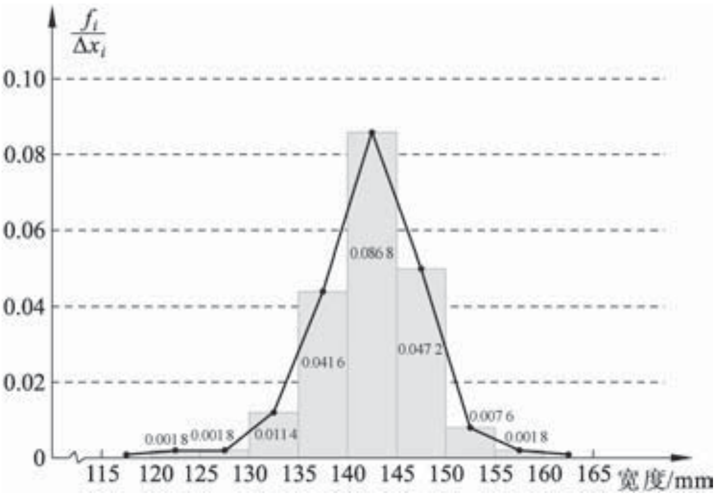


图 1-25

思考交流

(1) 在上面的例题中,如果只用前面的 50 个数据来估计 1665～1666 年英国男性头盖骨宽度的分布情况,你会得到怎样的结果?与例题中得到的估计有哪些不同?

(2) 假设在原挖掘出土地的附近,又新挖掘出了 100 块死于同时期的男性头盖骨.用两次得到的所有数据来估计 1665～1666 年,英国男性头盖骨宽度的分布情况,得到的结果又会有哪些不同?

抽象概括

在上面的例子中,虽然我们是用样本数据的频率分布来估计总体的分布,与真正的总体分布是有差别的,但是当样本量不断增大时,样本中落在每个区间内的样本数的频率会越来越稳定于总体在相应区间内取值的概率.也就是说,一般地,样本容量越大,用样本的频率分布去估计总体的分布就越精确.

另外,随着样本量的增大,所划分的区间数也可以随之增多,而每个区间的长度则会相应随之减小.如上面的例题中,若样本量增大,每个区间的长度可以由原来的 5 mm 减小为 4 mm 或 3 mm,相应的频率折线图就会越来越接近于一条光滑曲线.

当然,样本容量越大,工作量也就越大.所以,在实际问题当中,我们一般都要根据不同的情况选择适当的样本.

练　习

一位植物学家想要研究某类植物生长 1 年之后的高度.他随机抽取了 60 株此类植物,测得它们生长 1 年之后的高度如下所示(单位:cm):

73	84	91	68	72	83	75	58	87	41
48	61	65	72	92	68	73	43	57	78
80	59	84	42	67	69	64	73	51	65
63	82	90	54	63	76	61	68	66	78
55	81	94	79	45	67	70	98	76	72
72	91	86	75	76	50	69	69	56	74

(1) 完成下表:

高度分组(Δx_i)	频数(n_i)	频率(f_i)	$\frac{f_i}{\Delta x_i}$
40～50 cm			
50～60 cm			
60～70 cm			
70～80 cm			
80～90 cm			
90～100 cm			

(2) 根据上表画出相应的频率分布直方图和频率折线图,并描述此类植物生长 1 年之后的高度分布情况.

5.2　估计总体的数字特征

前面我们已经学习了如何用样本的频率分布来估计总体的分布.同样,假设通过随机抽样得到的样本为 $x_1, x_2, \cdots, x_n$,我们把

$$\bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}$$

和

$$s=\sqrt{s^2}=\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}}$$

分别称为样本平均数和样本标准差,用它们来分别估计总体的平均数和标准差.

分析理解

在 1996 年美国亚特兰大奥运会上,中国香港风帆选手李丽珊,以惊人的耐力和斗志,勇夺金牌,为香港体育史揭开了“突破零”的新

一页. 在风帆比赛中，成绩以低分为优胜. 比赛共 11 场，并以最佳的 9 场成绩计算最终的名次. 前 7 场比赛结束后，排名前 5 位的选手积分如表 1-8 所示：

表 1 8

排名	运动员	比赛场次											总分
		1	2	3	4	5	6	7	8	9	10	11	
1	李丽珊(中国香港)	3	2	2	2	4	2	7					22
2	简度(新西兰)	2	3	6	1	10	5	5					32
3	贺根(挪威)	7	8	4	4	3	1	8					35
4	威尔逊(英国)	5	5	14	5	5	6	4					44
5	李科(中国)	4	13	5	9	2	7	6					46

根据上面的比赛结果，我们如何比较各选手之间的成绩及稳定情况呢？如果此时让你预测谁将获得最后的胜利，你会怎么看？

由表 1-8，我们可以分别计算 5 位选手前 7 场比赛积分的平均数和标准差，分别作为量度各选手比赛的成绩及稳定情况的依据，结果如表 1-9 所示.

表 1 9

排名	运动员	平均积分($\overline{x}$)	积分标准差(s)
1	李丽珊(中国香港)	3.14	1.73
2	简度(新西兰)	4.57	2.77
3	贺根(挪威)	5.00	2.51
4	威尔逊(英国)	6.29	3.19
5	李科(中国)	6.57	3.33

从表 1-9 中可以看出：李丽珊的平均积分及积分标准差都比其他选手的小，也就是说，在前 7 场的比赛过程中，她的成绩最为优异，而且表现也最为稳定.

尽管此时还有 4 场比赛没有进行，但这里我们可以假定每位运动员在各自的 11 场比赛中发挥的水平大致相同(实际情况也确实如此)，因而可以把前 7 场比赛的成绩看作是总体的一个样本，并由此估计每位运动员最后比赛的成绩. 从已经结束的 7 场比赛的积分来看，李丽珊的成绩最为优异，而且表现最为稳定，因此，我们有足够的理由相信她在后面的 4 场比赛中会继续保持优异而稳定的成绩，获得最后的冠军.

当然，事实也进一步验证了我们的预测，李丽珊正是凭着自己优异而稳定的表现，成为香港首位奥运金牌得主的.

小资料

“二战”期间，盟军非常想知道德军总共制造了多少辆坦克. 德国人在制造坦克时是墨守成规的，他们把坦克从 1 开始进行了连续编号. 在战争进行过程中，盟军缴获了一些德军坦克，并记录了它们的生产编号. 怎样利用这些缴获坦克的编号来估计德军坦克的总数呢？

这里，我们假设总体参数是未知的德军坦克总数 N，而缴获坦克的编号则是样本，如果假定缴获的坦克代表了所有坦克的一个随机样本，问题就转化成如何通过样本来估计总体参数 N 了. 我们可以先求出被缴获坦克编号的算术平均数 $\bar{x}$（样本平均数），用其作为总体平均数 $a=\dfrac{1+2+\cdots+N}{N}=\dfrac{N+1}{2}$ 的估计. 这样，用样本平均数 $\bar{x}$ 的 2 倍减 1 即可得到德军坦克总数 N 的估计值了（当然，这样得到的估计值可能小于记录中的最大编号，在这种情况下可以进行适当的变形以得到更好的估计值）.

这种方法及其各种变形的确用于“二战”之中. 从战后发现的德军记录来看，盟军的估计值非常接近德军所生产的坦克的真实值，由此我们也不难看出统计方法在现实中应用的广泛性和有效性.

动手实践

请设计一个抽样方案，调查你们学校高中学生的身高分布情况，估计身高的平均数和标准差，并与同学进行交流.

思考交流

(1) 在上面的活动中，比较全班同学各自的估计结果，看看有哪些不同.

(2) 假设你的同桌抽取的样本量与你抽取的相同，他得到的样本频率分布直方图与你的一定相同吗？样本的平均数和标准差呢？

(3) 假如你用同样的方法先后从总体中抽取了两个大小相同的样本，两次得到的样本频率分布直方图一定相同吗？样本的平均数和标准差呢？为什么？

抽象概括

用样本估计总体时，如果抽样的方法比较合理，那么样本可以反映总体的信息，但从样本得到的信息会有偏差. 在上面的活动中，尽管所有的样本都来自同一个总体，从这些样本中所得到的有关总体的估计仍然可能互不相同，这一现象是由抽样的随机性引起的. 如果

抽样方案没有问题的话，那么这些结论之所以不同，其原因就在于样本的随机性. 在随机抽样中，这种偏差是不可避免的.

虽然我们从样本数据得到的分布、平均数和标准差(通常称之为样本分布、样本平均数和样本标准差)并不是总体真正的分布、平均数和标准差，而只是总体的一个估计，但这种估计是合理的，特别是当样本量很大时，它们确实反映了总体的信息.

练　习

某科研单位分别用一台 Au-CH-1 型自动测色仪及一台 302D 型自动测色仪对中国人的面部自然肤色，从新生婴儿直到老年人的色度特征用简单随机抽样方法进行了广泛取样测定，测定人数共1 668人，其中新生婴儿到 2 岁幼儿 508 人，3 岁到 17 岁儿童及青少年 548 人，18 岁到 78 岁的成年人及老年人 612 人，被测者的籍贯包括 26 个省市，除汉族外还包括蒙古族、藏族、维吾尔族、壮族等 17 个少数民族. 在此基础上研制成功具有中国人典型自然肤色的色样一套. 测量数据如下表：

反射率/(%)	人数分布		
	男	女	合计
12.0～14.0	4	0	4
14.0～16.0	14	2	16
16.0～18.0	23	0	23
18.0～20.0	75	22	97
20.0～22.0	90	50	140
22.0～24.0	117	119	236
24.0～26.0	143	139	282
26.0～28.0	113	141	254
28.0～30.0	78	135	213
30.0～32.0	91	102	193
32.0～34.0	48	71	119
34.0～36.0	20	33	53
36.0～38.0	16	15	31
38.0～40.0	4	3	7
总　　计	836	832	1 668

请分别估计中国男性和女性的肤色反射率的平均数及标准差.(为计算方便，反射率取每个区间的中点，如 12.0～14.0 取 13.0，其他类似计算)

习 题 1—5

1. 截至 2007 年，美国历届总统中，就任时年纪最小的是罗斯福，他于 1901 年就任，当时年仅 42 岁；就任时年纪最大的是里根，他于 1981 年就任，当时 69 岁. 下面按时间顺序（从 1789 年的华盛顿到 2001 年的布什，共 43 任）给出了历届美国总统就任时的年龄：

57 61 57 57 58 57 61 54 68 51 49 64 50 48 65
52 56 46 54 49 51 47 55 55 54 42 51 56 55 51
54 51 60 62 43 55 56 61 52 69 64 46 54

(1) 完成下表：

年龄分组(Δx_i)	频　数(n_i)	频率(f_i)	$\frac{f_i}{\Delta x_i}$
40～45 岁			
45～50 岁			
50～55 岁			
55～60 岁			
60～65 岁			
65～70 岁			

(2) 根据上表画出相应的频率分布直方图和频率折线图，并用自己的语言描述一下历届美国总统就任时年龄的分布情况.

2. 城市公交车的数量太多容易造成资源的浪费，太少又难以满足乘客的需求. 为此，某城市公交公司在某站台随机调查了 80 名乘客，他们的候车时间如下所示（单位：min）：

17 14 20 12 10 24 18 17 1 22 13 19 28 5 34 7
25 18 28 1 15 31 12 11 10 16 12 9 10 13 19 10
12 12 16 22 17 23 16 15 16 11 9 3 13 2 18 22
19 9 23 28 15 21 28 12 11 14 15 3 11 6 2 18
25 5 12 15 20 16 12 28 20 12 28 15 8 32 18 9

(1) 将数据进行适当的分组，并画出相应的频率分布直方图和频率折线图；
(2) 这 80 名乘客候车时间的平均数是多少？标准差呢？
(3) 你能为公交公司提出什么建议？

3. 在一项农业试验中，A，B 两种肥料分别被用于同类橘子树的生长. 为了了解这两种肥料的效果，试验人员分别从施用这两种肥料的橘子树中随机抽取了 12 棵，下面给出了每一棵橘子树的产量（单位：kg）：
肥料 A：63，54，19，20，72，92，8，10，22，11，24，5；
肥料 B：57，86，33，40，59，56，73，25，44，31，64，45.
(1) 请用茎叶图表示分别施用 A，B 两种肥料的橘子树的产量，并从图中比较各自平均数的大小；
(2) 从图中你认为哪一个数据集的标准差更小？
(3) 分别计算施用 A，B 两种肥料的橘子树产量的平均数和标准差，看看与你的估计结果是否一致；
(4) 你认为哪种肥料对橘子树的产量影响更大？为什么？

4. 请设计一个抽样方案，调查你们学校高中学生的体重分布情况，并估计体重的平均数和标准差.

阅读材料

标准差的用途

标准差除了可以用来刻画一组数据的离散程度外，还可以刻画每个数据偏离平均水平的程度. 图 1-26 是 27 个人每 30 秒心跳次数的直方图，平均心跳次数是 34.4，标准差是 6.9.

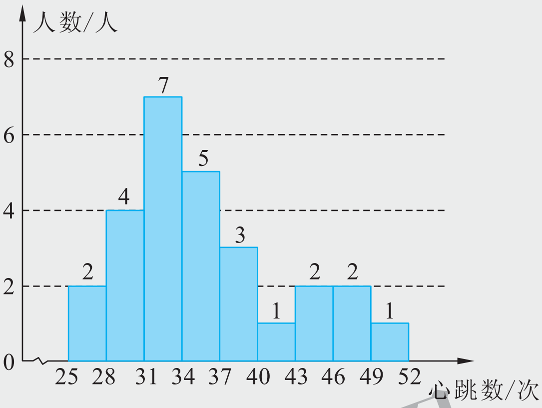


图 1-26

（资料来源：Statistical Thinking，Swarthmore College，Springer 1995）

平均数加标准差是 $34.4+6.9=41.3$，平均数减标准差是 $34.4-6.9=27.5$；平均数加两倍的标准差是 $34.4+2\times6.9=48.2$，平均数减两倍的标准差是 $34.4-2\times6.9=20.6$. 从平均数减两倍的标准差到平均数加两倍的标准差，即从 20.6 到 48.2，几乎包含了所有的观测值，27 个观测值中只有一个落在了这个区间之外. 平均数加减一个标准差，即从 27.5 到 41.3，包含了所有观测值的近 $\frac{3}{4}$.

在前面的风帆赛中，虽然李科在前 7 场比赛积分的算术平均数和标准差比威尔逊都略大一些，但威尔逊在第三场比赛的积分 14 分超过了其平均数加两倍的标准差（$6.29+2\times3.19=12.67$）. 有意思的是，在最后比赛结果的报道中，威尔逊的排名低于李科.

在上一节估计时间的活动中，分别计算平均数加减一个标准差和两个标准差，并看看有多少同学的估计值分别落在这两个区间之内.

§6　统计活动：结婚年龄的变化

问题提出

为了帮助同学们更好地理解统计的全过程，我们在这一节设计了一个统计活动.

在日常生活中，我们或许都有这样的感觉：人们初次结婚的年龄在随着时代的发展而逐渐增大. 那么，实际情况是否的确如此呢？请大家对这个问题设计一个调查方案并开展统计活动.

动手实践

我们可以按照如下的步骤来进行这个统计活动.

1. 确定调查对象

全班同学的父母辈和祖父母辈.

2. 收集数据

每位同学收集自己父母辈和祖父母辈的初次结婚年龄（例如，调查自己的父亲、母亲、祖父、祖母的初婚年龄），按照以下方式记录下来（如表 1-10）.

表 1 10

	父辈	母辈	祖父辈	祖母辈
初次结婚年龄/岁				

3. 整理数据

（1）先将本小组成员收集到的数据按表 1-11 汇总.

表 1 11　　第____小组

成员 \ 初次结婚年龄/岁	父辈	母辈	祖父辈	祖母辈
小组成员 1				
小组成员 2				
⋮	⋮	⋮	⋮	⋮
小组成员 n				

（2）再把班上所有同学的数据按照小组进行汇总，得到表 1-12.

表 1 12

成员 \ 初次结婚年龄/岁	父辈	母辈	祖父辈	祖母辈
第 1 小组				
第 2 小组				
⋮	⋮	⋮	⋮	⋮
第 m 小组				

4. 分析数据

(1) 将上面的数据用适当的统计图表示出来，并与同伴进行交流. 你觉得哪种统计图更合适？

(2) 分别估计父辈、母辈、祖父辈、祖母辈的初次结婚年龄的平均数与标准差，并进行比较.

5. 作出推断

从上面的数据，你能得到什么结论？它与你在从事这个统计活动之前的猜想一致吗？

练　习

请调查你身边在最近 5 年内初次结婚的人，记录他们初次结婚的年份和年龄.

思考交流

(1) 在上一课时统计活动的过程中，调查的问题和目的分别是什么？你经历了哪些主要的步骤？

(2) 根据调查的问题和目的，应当如何确定调查的对象？

(3) 你认为如何收集数据比较方便？

(4) 你将如何对收集到的数据快捷并且正确地进行整理？

(5) 从统计数据中，要想得到一个比较好的结论，应当对数据作哪些分析？

动手实践

请根据你们全班同学课前收集的数据，分析在最近的 5 年内，人们初次结婚的年龄是否随着时代的发展而逐渐增大.

实际上，网上也有很多与此相关的信息和统计数据. 表 1-13 给出了 1995～1999 年全国大陆地区女性平均初次结婚年龄的数据，从中我们不难看出，就全国大陆的平均趋势和大部分地区的情况而言，女性初次结婚的年龄确实随着时代的发展在逐渐增大.

表 1 13　1995～1999 年全国大陆地区女性平均初次结婚年龄　（单位：岁）

地　区	1995 年	1996 年	1997 年	1998 年	1999 年
全国大陆	22.93	23.20	23.39	23.57	23.62
北京市	24.71	25.16	25.25	25.30	25.88
天津市	23.31	23.64	23.88	23.90	24.13
河北省	23.07	23.42	23.35	23.76	23.50
山西省	22.03	22.26	22.92	22.60	22.83
内蒙古自治区	22.92	23.06	23.20	23.47	23.29
辽宁省	22.98	23.07	23.42	23.64	23.95
吉林省	22.97	23.31	23.42	23.79	23.78
黑龙江省	23.04	23.11	22.76	23.04	22.99
上海市	24.07	24.10	24.26	24.59	25.12
江苏省	23.04	23.18	23.54	23.26	23.50
浙江省	23.66	24.03	24.36	24.14	24.48
安徽省	23.07	23.42	23.68	23.54	23.49
福建省	22.62	23.11	23.24	23.27	23.61
江西省	22.24	22.58	22.81	22.81	23.30
山东省	24.46	24.58	24.76	24.83	24.74
河南省	23.16	23.29	23.33	23.70	23.40
湖北省	22.71	22.97	23.44	23.66	23.31
湖南省	23.27	23.32	23.23	23.56	23.87
广东省	23.57	24.17	24.54	24.70	24.71
广西壮族自治区	23.35	23.85	23.51	24.00	23.78
海南省	22.65	22.57	24.05	24.40	23.35
重庆市	—	—	22.68	23.17	23.28
四川省	22.13	22.15	22.63	23.16	23.13
贵州省	22.45	22.81	22.64	23.09	23.34
云南省	21.91	22.34	22.24	22.17	23.00
西藏自治区	23.42	21.50	21.43	22.94	21.89

续表

地　区	1995 年	1996 年	1997 年	1998 年	1999 年
陕西省	22.16	22.50	22.94	23.00	23.01
甘肃省	22.61	23.05	23.36	23.39	23.11
青海省	21.83	22.02	21.71	22.01	22.21
宁夏回族自治区	21.91	21.69	21.92	22.17	22.08
新疆维吾尔自治区	22.78	23.17	22.58	22.67	23.27

数据来源：国家统计局

习 题 1—6

如果要调查在全国范围内，人们初次结婚的年龄也是随着时代的发展而逐渐增大，你该如何应用统计方法来说明？

§7 相 关 性

问题提出

在现实生活中，有些量与量之间有着明确的函数关系. 例如：

正方形的边长 a 和面积 S，有着 $S=a^2$ 的关系；

真空中做自由落体运动的物体，其下落的距离 h 和下落的时间 t 有着 $h=\frac{1}{2}gt^2$ 的关系；

一辆行驶在公路上的汽车，每个时刻 t 都有一个确定的速度 v，它们之间也是函数关系，尽管我们无法知道这个函数的解析表达式，也画不出它的图像.

但是，在现实生活中还有一些量不满足函数关系，如人的身高与体重. 一般说来，人的身高越高，体重越重，二者确实有关系. 但是身高相同的人，体重却不一定相同，也就是说，给定身高 h 没有唯一的体重 m 与之对应. 在现实中，这样的例子还有很多，如人的年龄与血压、农作物的施肥量与产量等.

为了了解人的身高与体重的关系，我们随机地抽取 9 名 15 岁的男生，测得他们的身高、体重如表 1-14：

表 1 14

编号	1	2	3	4	5	6	7	8	9
身高/cm	165	157	155	175	168	157	178	160	163
体重/kg	52	44	45	55	54	47	62	50	53

从表 1-14 中，不难看出，同一身高 157 cm 对应着不同的体重 44 kg和 47 kg，体重不是身高的函数. 如果把身高看作横坐标、体重看作纵坐标，在坐标系中画出对应的点，就会发现，随着身高的增长，体重基本上是呈直线增加的趋势（如图 1-27）.

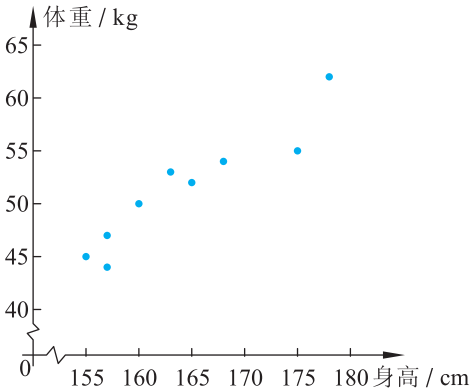


图 1-27

信息技术建议

利用计算机电子表格软件(如 Excel)可以方便地制作两个变量之间的散点图.

在考虑两个量的关系时,为了对变量之间的关系有一个大致的了解,人们通常将变量所对应的点描出来,这些点就组成了变量之间的一个图,通常称这种图为变量之间的**散点图**.

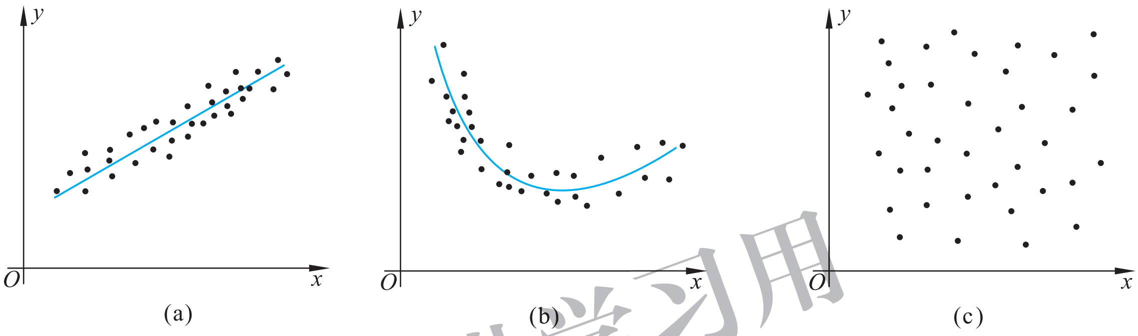


图 1-28

从散点图上可以看出,如果变量之间存在着某种关系,这些点会有一个集中的大致趋势,这种趋势通常可以用一条光滑的曲线来近似,这样近似的过程称为**曲线拟合**. 若两个变量 x 和 y 的散点图中,所有点看上去都在一条直线附近波动,则称变量间是**线性相关**的. 此时,我们可以用一条直线来近似(如图 1-28(a)).

若所有点看上去都在某条曲线(不是一条直线)附近波动,则称此相关为**非线性相关**的. 此时,可以用一条曲线来拟合(如图 1-28(b)). 如果所有的点在散点图中没有显示任何关系,则称变量间是**不相关**的(如图 1-28(c)).

例　一般说来,一个人的身高越高,他的手就越大,相应地,他的右手一拃长(如图 1-18)就越长,因此,人的身高与右手一拃长之间存在着一定的关系. 为了对这个问题进行调查,我们收集了北京市某中学 2003 年高三年级 96 名学生的身高与右手一拃长的数据如表1-15.

表 1 15

性别	身高/cm	右手一拃长/cm	性别	身高/cm	右手一拃长/cm
女	152	18.5	女	172	18.5
女	153	16.0	女	173	18.0
女	156	16.0	女	173	22.0
女	157	20.0	男	162	19.0
女	158	17.3	男	164	19.0
女	159	20.0	男	165	21.0
女	160	15.0	男	168	18.0
女	160	16.0	男	168	19.0
女	160	17.5	男	169	17.0
女	160	17.5	男	169	20.0
女	160	19.0	男	170	20.0
女	160	19.0	男	170	21.0
女	160	19.0	男	170	21.5
女	160	19.5	男	170	22.0
女	161	16.1	男	171	21.5
女	161	18.0	男	171	21.5
女	162	18.2	男	171	22.3
女	162	18.5	男	172	21.5
女	163	20.0	男	172	23.0
女	163	21.5	男	173	20.0
女	164	17.0	男	173	20.0
女	164	18.5	男	173	20.0
女	164	19.0	男	173	20.0
女	164	20.0	男	173	21.0
女	165	15.0	男	174	22.0
女	165	16.0	男	174	22.0
女	165	17.5	男	175	16.0
女	165	19.5	男	175	20.0
女	166	19.0	男	175	21.0
女	167	19.0	男	175	21.2
女	167	19.0	男	175	22.0
女	168	16.0	男	176	16.0
女	168	19.0	男	176	19.0
女	168	19.5	男	176	20.0
女	170	21.0	男	176	22.0
女	170	21.0	男	176	22.0
女	170	21.0	男	177	21.0
女	171	19.0	男	178	21.0
女	171	20.0	男	178	21.0
女	171	21.5	男	178	22.5

续表

性别	身高/cm	右手一拃长/cm	性别	身高/cm	右手一拃长/cm
男	178	24.0	男	182	18.5
男	179	21.5	男	182	21.5
男	179	21.5	男	182	24.0
男	179	23.0	男	183	21.2
男	180	22.5	男	185	25.0
男	181	21.1	男	186	22.0
男	181	21.5	男	191	21.0
男	181	23.0	男	191	23.0

(1) 根据表 1-15 中的数据，制成散点图. 你能从散点图中发现身高与右手一拃长之间的近似关系吗?

(2) 如果近似成线性关系，请画出一条直线来近似地表示这种线性关系.

(3) 如果一个学生的身高是 188 cm，你能估计他的右手一拃大概有多长吗?

思考交流

根据表 1-15 中的数据，制成的散点图如图 1-29.

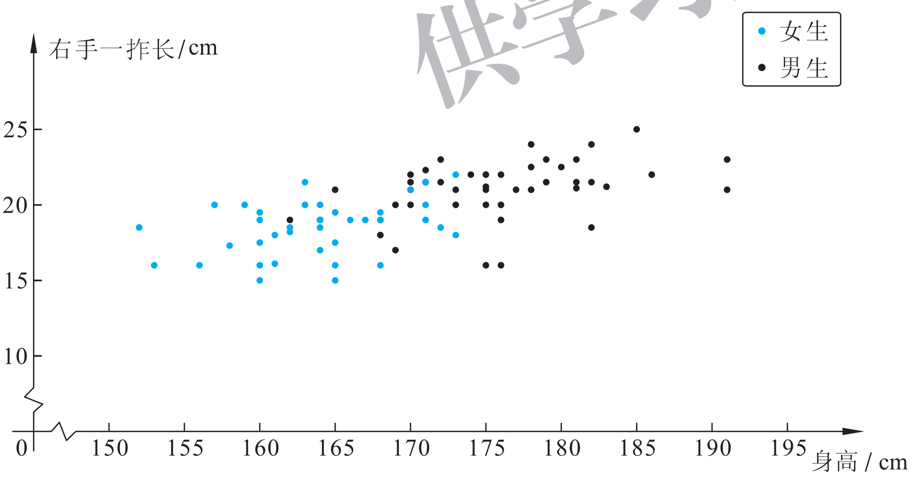


图 1-29

从散点图上可以发现，身高与右手一拃长之间的总体趋势是成一直线，也就是说，它们之间是线性相关的. 那么，怎样确定这条直线呢? 你是怎么想的? 与同学进行交流.

分析理解

同学甲说：我从左端点开始，取两条直线，如图 1-30，再取这两条

直线的"中间位置"作一条直线. 根据我的想法，一个身高 188 cm 的学生，他的右手一拃长大概为 21 cm.

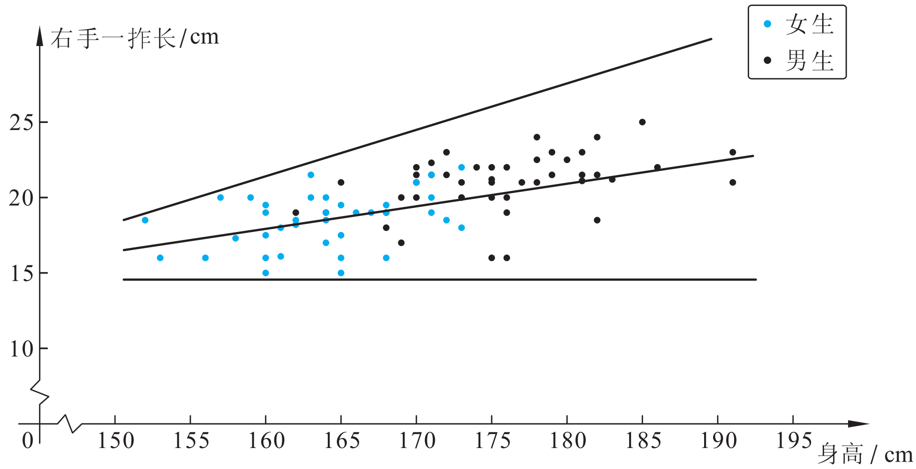


图 1-30

同学乙说：这样做不准确. 我先求出相同身高同学右手一拃长的平均数，画出散点图，如图 1-31，再画出近似的直线，使得在直线两侧的点数尽可能一样多. 根据我的想法，一个身高 188 cm 的学生，他的右手一拃长大概为 22 cm.

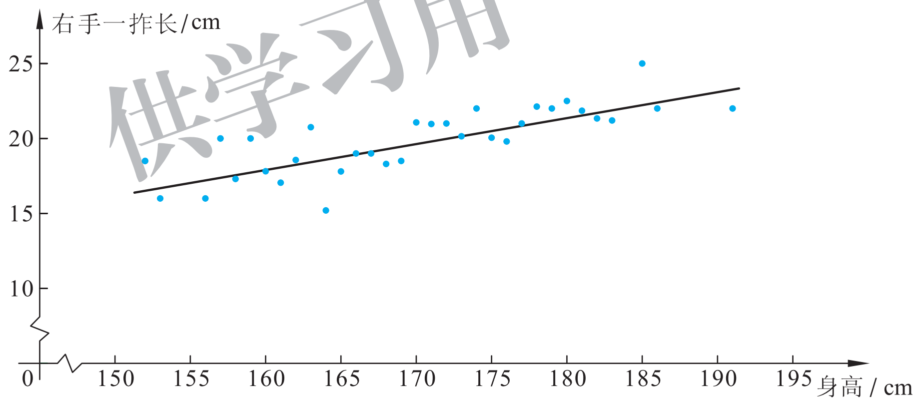


图 1-31

同学丙说：我先将所有的点分成两部分，一部分是身高在 170 cm 以下的，一部分是身高在 170 cm 以上的；然后，每部分的点求一个"平均点"——身高的平均数作为平均身高，右手一拃长的平均数作为平均右手一拃长，即(164，19)，(177，21)；最后，将这两点连接成一条直线.

设这条直线的方程是：$y=kx+b$，其中 $k=\dfrac{21-19}{177-164}=\dfrac{2}{13}\approx 0.154$. 代入一点的坐标求出 $b=-\dfrac{81}{13}\approx -6.231$，进而 $y=0.154x-6.231$ 即为所求的直线方程. 根据我的想法，一个身高 188 cm 的学生，他的右手一拃长大概为 22.7 cm.

同学丁说：我先将所有的点按横坐标从小到大的顺序进行排列，

尽可能地平均分成三等份；每部分的点按照同学丙的方法求一个“平均点”，“最小点”为(161.3，18.2)，“中间点”为(170.5，20.1)，“最大点”为(179.2，21.3)．求出这三个点的“平均点”为(170.3，19.9)．我再用直尺连接“最大点”与“最小点”，然后平行地推，画出过“平均点”(170.3，19.9)的直线(如图 1-32)．

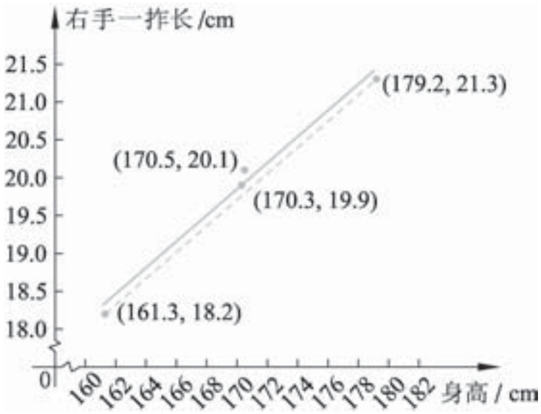


图 1-32

设这条直线的方程是 $y=kx+b$，其中 $k=\dfrac{18.2-21.3}{161.3-179.2}=\dfrac{31}{179}\approx 0.173$，代入点(170.3，19.9)的坐标求出 $b=-\dfrac{8\ 586}{895}\approx -9.593$，进而 $y=0.173x-9.593$ 即为所求的直线方程．根据我的想法，一个身高 188 cm 的学生，他的右手一拃长大概为 23.0 cm．

从上面的讨论看，这些同学的处理方法差别很大，那么我们应当选取一个什么样的方法来处理更好些呢？这将是我们下面一节中要讨论的．

在这里需要强调的是，身高和右手一拃长之间没有函数关系．我们得到的直线方程，只是对其变化趋势的一个近似描述．对一个给定身高的人，人们可以用这个方程来估计这个人的右手一拃长，这是十分有意义的．

练　习

下表是某小卖部 6 天卖出热茶的杯数与当天气温的对比表：

气温/℃	26	18	13	10	4	-1
杯数/杯	20	24	34	38	50	64

(1) 将上表中的数据制成散点图．

(2) 你能从散点图中发现气温与饮料杯数近似成什么关系吗？

(3) 如果近似成线性关系，请画出一条直线来近似地表示这种线性关系．

(4) 如果某天的气温是 -5 ℃，预测这天小卖部卖出热茶的杯数．

习 题 1—7

1. 采用抽样的方法，调查下面两个问题，并写出调查方案与分析报告.

   (1) 你们所在学校高一年级学生的身高与右手一拃长之间的关系；

   (2) 你们所在学校高一年级学生的左手一拃长与右手一拃长之间的关系.

2. 下面是世界上 10 名男网球选手的身高(x)与体重(y)情况.

姓名	身高(x)/cm	体重(y)/kg
Carlos Moya	190	82
Richard Fromberg	196	88
Marcelo Rios	175	63
Pat Rafter	185	79
Jason Stoltenberg	186	80
Andre Agassi	180	75
Todd Martin	198	96
Karol Kucera	188	75
Mark Philippoussis	194	92
Greg Rusedski	193	86

   (1) 将上表中的数据制成散点图.

   (2) 你能从散点图中发现身高与体重近似成什么关系吗？

   (3) 若近似成线性关系，请画出一条直线来近似地表示这种线性关系.

   (4) 若某名男网球运动员的身高是 172 cm，请预测他的体重.

3. 为了保证交通安全，高速公路上通常设置一些交通标志. 比如，前面××米是××出口. 美国宾夕法尼亚州的一个研究小组对 30 位志愿者进行了研究，这些志愿者的年龄从 18 岁到 82 岁，研究他们对交通标志的最大可识别距离与年龄之间是否有关系. 调查之前，研究人员重新设计了新的交通标志，调查数据如下表.

年龄/岁	最大可识别距离/英尺[1]	年龄/岁	最大可识别距离/英尺	年龄/岁	最大可识别距离/英尺
18	510	37	420	68	300
20	590	41	460	70	390
22	560	46	450	71	320
23	510	49	380	72	370
24	460	53	460	73	280
25	490	55	420	74	420
27	560	63	350	75	460
28	510	65	420	77	360
29	460	66	300	79	310
32	410	67	410	82	360

[1] 1 英尺＝0.304 8 米.

(1) 根据上表中的数据，制成散点图.你能从散点图中发现年龄与最大可识别距离之间的近似关系吗？

(2) 如果近似成线性关系，请画出一条直线来近似地表示这种线性关系.

(3) 如果一个美国司机的年龄是 50 岁，你能估计他的最大可识别距离大概有多少吗？

(4) 根据以上的调查数据，你能得出什么结论？由此，你能对老年司机给出哪些建设性的意见呢？

§8 最小二乘估计

问题提出

在上一节的讨论中，我们知道，人的身高与右手一拃长之间近似存在着线性关系，这种线性关系可以有多种方法来进行刻画. 那么用什么样的线性关系刻画会更好一些呢？

有一个非常直观的想法，一个好的线性关系要保证这条直线与所有点都近. 最小二乘法就是基于这种想法.

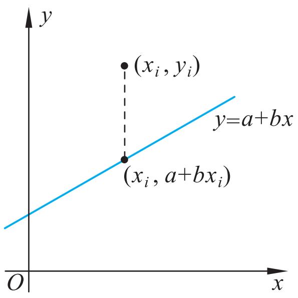


图 1-33

假设一条直线的方程为：$y=a+bx$，任意给定一个样本点(x_i,y_i)，我们用

$$[y_i-(a+bx_i)]^2$$

来刻画这个样本点与这条直线之间的“距离”，用它来表示二者之间的接近程度（如图 1-33）.

思考交流

（1）如果有 5 个样本点，其坐标分别为(x_1,y_1)，(x_2,y_2)，(x_3,y_3)，(x_4,y_4)，(x_5,y_5)，怎样刻画这些样本点与直线 $y=a+bx$ 的接近程度？

（2）如果有 10 个样本点，其坐标分别为(x_1,y_1)，(x_2,y_2)，…，(x_{10},y_{10})，怎样刻画这些样本点与直线 $y=a+bx$ 的接近程度？

（3）如果有 100 个样本点，其坐标分别为(x_1,y_1)，(x_2,y_2)，…，(x_{100},y_{100})，怎样刻画这些样本点与直线 $y=a+bx$ 的接近程度？

抽象概括

如果有 n 个点(x_1,y_1)，(x_2,y_2)，…，(x_n,y_n)，可以用下面的表达式来刻画这些点与直线 $y=a+bx$ 的接近程度：

$$[y_1-(a+bx_1)]^2+[y_2-(a+bx_2)]^2+\cdots+[y_n-(a+bx_n)]^2.$$

使得上式达到最小值的直线 $y=a+bx$ 就是我们所要求的直线，这种方法称为**最小二乘法**.

如果用 $\bar{x}$ 表示 $\frac{x_1+x_2+\cdots+x_n}{n}$，用 $\bar{y}$ 表示 $\frac{y_1+y_2+\cdots+y_n}{n}$，则可以求得

$$b=\frac{(x_1-\bar{x})(y_1-\bar{y})+(x_2-\bar{x})(y_2-\bar{y})+\cdots+(x_n-\bar{x})(y_n-\bar{y})}{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}$$

$$=\frac{x_1y_1+x_2y_2+\cdots+x_ny_n-n\bar{x}\,\bar{y}}{x_1^2+x_2^2+\cdots+x_n^2-n\bar{x}^2}.$$

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

（这两个公式的推导过程不在这里讨论，有兴趣的同学可以利用配方法试着进行推导.）

这样得到的直线方程称为**线性回归方程**，a，b 是线性回归方程的系数.

回　归

一种统计方法，它通过计算变量之间的相关系数进而估计它们之间的联系公式.

小资料

用最小二乘法推导 3 个点的线性回归方程

设有 3 个点 (x_1,y_1)，(x_2,y_2)，(x_3,y_3)，则由最小二乘法可知直线 $y=a+bx$ 与这 3 个点的接近程度由下面表达式刻画：

$$[y_1-(a+bx_1)]^2+[y_2-(a+bx_2)]^2+[y_3-(a+bx_3)]^2.$$

这个表达式可以整理成为关于 a 的一元二次函数 $f(a)$，如下所示：

$$f(a)=3a^2-2a[(y_1-bx_1)+(y_2-bx_2)+(y_3-bx_3)]+(y_1-bx_1)^2+(y_2-bx_2)^2+(y_3-bx_3)^2$$

$$=3[a^2-2a(\bar{y}-b\bar{x})]+(y_1-bx_1)^2+(y_2-bx_2)^2+(y_3-bx_3)^2.$$

利用配方法即得

$$f(a)=3[a-(\bar{y}-b\bar{x})]^2+(y_1-bx_1)^2+(y_2-bx_2)^2+(y_3-bx_3)^2-3(\bar{y}-b\bar{x})^2.$$

从而当 $a=\bar{y}-b\bar{x}$ 时，使得函数 $f(a)$ 达到最小值.

将 a 代入第一个表达式，整理成为关于 b 的一元二次函数 $g(b)$，如下所示：

$$g(b)=[(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2]b^2-2b[(y_1-\bar{y})(x_1-\bar{x})+(y_2-\bar{y})(x_2-\bar{x})+(y_3-\bar{y})(x_3-\bar{x})]+[(y_1-\bar{y})^2+(y_2-\bar{y})^2+(y_3-\bar{y})^2].$$

同样使用配方法可以得到，当

$$b=\frac{(y_1-\bar{y})(x_1-\bar{x})+(y_2-\bar{y})(x_2-\bar{x})+(y_3-\bar{y})(x_3-\bar{x})}{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2}$$

$$=\frac{x_1y_1+x_2y_2+x_3y_3-3\bar{x}\,\bar{y}}{x_1^2+x_2^2+x_3^2-3\bar{x}^2}$$

时，使得函数 $g(b)$ 达到最小值.

从而可以得到 3 个点 (x_1,y_1)，(x_2,y_2)，(x_3,y_3) 的线性回归方程：

$$y=\bar{y}-\frac{(x_1y_1+x_2y_2+x_3y_3-3\bar{x}\,\bar{y})\bar{x}}{x_1^2+x_2^2+x_3^2-3\bar{x}^2}+\frac{x_1y_1+x_2y_2+x_3y_3-3\bar{x}\,\bar{y}}{x_1^2+x_2^2+x_3^2-3\bar{x}^2}x.$$

思考交流

如果样本点只有两个，那么上面用最小二乘法估计得到的直线与用两点式求出的直线方程一致吗？试给出证明.

例 1　在上一节练习中，从散点图可以看出，某小卖部 6 天卖出热茶的杯数(y)与当天气温(x)之间是线性相关的. 数据如表 1-16：

表 1 16

气温(x_i)/℃	26	18	13	10	4	-1
杯数(y_i)/杯	20	24	34	38	50	64

(1) 试用最小二乘法求出线性回归方程.

(2) 如果某天的气温是 -3 ℃，请预测这天可能会卖出热茶多少杯.

解　(1) 从散点图 1-34 中可以看出，表 1-16 中的两个变量是线性相关的.

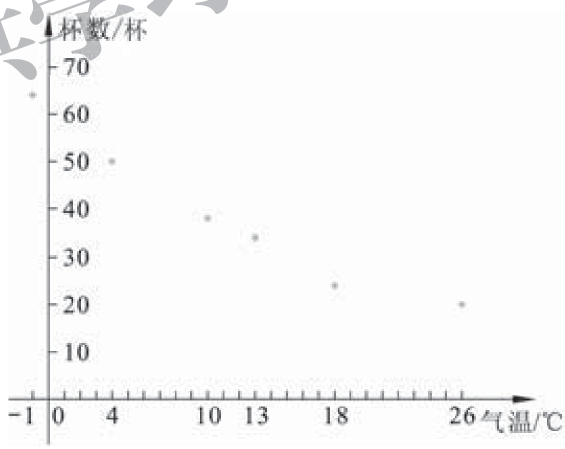


图 1-34

先列表求出 $\bar{x}=\frac{35}{3}$，$\bar{y}=\frac{115}{3}$，其他数据❶如表 1-17.

❶科学计算器具有计算平方和等功能，能帮助计算. 也可以利用计算机电子表格软件（如 Excel）中的计算功能，求出回归方程的系数. 具体的操作步骤参见本节的“信息技术应用”栏目.

表 1 17

i	x_i	y_i	x_i^2	x_iy_i
1	26	20	676	520
2	18	24	324	432
3	13	34	169	442
4	10	38	100	380
5	4	50	16	200
6	−1	64	1	−64
合计	70	230	1 286	1 910

进而,可以求得

$$b=\frac{1\,910-6\times\frac{35}{3}\times\frac{115}{3}}{1\,286-6\times\frac{35}{3}\times\frac{35}{3}}\approx-1.648,$$

$$a\approx57.557.$$

于是,线性回归方程为

$$y=57.557-1.648x.$$

(2) 由上面的最小二乘估计得出的线性回归方程知,当某天的气温是-3 ℃时,卖出热茶的杯数估计为:

$$57.557-1.648\times(-3)=62.501\approx63.$$

练　习

参见§7中北京市某中学学生关于身高与右手一拃长之间的数据.

(1) 作出女生数据的散点图,看是否成线性关系.如果是,用最小二乘法估计二者之间的线性回归方程.

(2) 作出男生数据的散点图,看是否成线性关系.如果是,用最小二乘法估计二者之间的线性回归方程.

(3) 比较这两个方程,说说你的看法.

动手实践

参见§7中北京市某中学学生关于身高与右手一拃长之间的数据,用最小二乘法估计二者之间的线性回归方程.

(1) 任意选取其中的5个数据;

(2) 任意选取15个数据;

(3) 利用所有数据;

（4）比较这三个方程，说说你的看法.

抽象概括

利用实验数据进行拟合时，所用数据的多少直接影响拟合的结果. 从理论上来说，数据越多，拟合的效果越好. 在以后的选修课中，还将研究数据之间的相关程度. 相关程度越高，估计的结果越好.

另外，需要注意的是，即使我们选取相同的样本数，得到的直线方程也可能是不相同的，这是由样本的随机性造成的. 样本量越大，所估计的直线方程越能更好地反映变量之间的关系.

信息技术建议

尝试使用数学软件或图形计算器，利用最小二乘法求出例 2 中两个变量之间的线性回归方程.

例 2 下面是两个变量的一组数据：

表 1 18

x	1	2	3	4	5	6	7	8
y	1	4	9	16	25	36	49	64

请用最小二乘法求出这两个变量之间的线性回归方程.

解 根据表 1-18 的数据，可以计算出：$\overline{x}=4.5$，$\overline{y}=25.5$，其他数据如表 1-19.

表 1 19

i	x_i	y_i	x_i^2	x_iy_i
1	1	1	1	1
2	2	4	4	8
3	3	9	9	27
4	4	16	16	64
5	5	25	25	125
6	6	36	36	216
7	7	49	49	343
8	8	64	64	512
合计	36	204	204	1 296

进而，可以求得

$$b=\frac{1\,296-8\times4.5\times25.5}{204-8\times4.5\times4.5}=9,$$

$$a=-15.$$

于是，线性回归方程为

$$y=-15+9x.$$

思考交流

在例 2 中，从表中提供的数据很容易看出：$y=x^2$，而我们用最小二乘法进行估计时得出的是线性方程(图 1-35). 这样的估计已经失去了意义. 你觉得问题出在哪儿？应当怎样去避免？

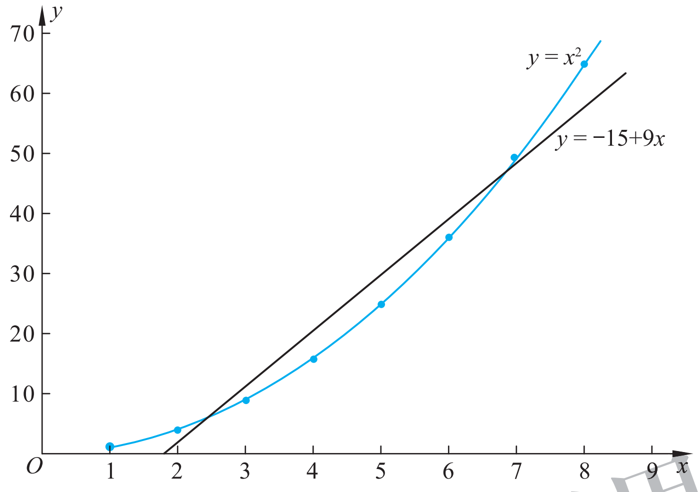


图 1-35

抽象概括

利用最小二乘法估计时，要先作出数据的散点图. 如果散点图呈现一定的规律性，我们再根据这个规律性进行拟合. 如果散点图呈现出线性关系，我们可以用最小二乘法估计出线性回归方程；如果散点图呈现出其他的曲线关系，我们就要利用其他的工具进行拟合.

信息技术应用

利用计算机求线性回归方程

利用计算机电子表格软件(如 Excel)中的计算功能，可以求出回归方程的系数. 具体的操作步骤如下.

(1) 在新建的一个电子表格软件中的 B2 至 C7 的位置输入原始数据.

(2) 在 D2 位置输入计算公式：=B2^2，然后将此公式复制到 D3 至 D7 的位置，计算出 x_i^2 的值.

(3) 在 E2 位置输入计算公式:=B2 * C2,然后将此公式复制到 E3 至 E7 的位置,计算出 $x_i \cdot y_i$ 的值.

(4) 在第 8 行的 B8 位置处输入列的求和公式:=SUM(B2:B7),然后复制到此行的相应位置处,计算出相应列的和.

(5) 在第 9 行的 B9 位置处输入 B2 至 B7 数据的平均数公式:=AVERAGE(B2:B7),并将此公式复制到 C9 位置处.

(6) 在第 10 行的 B10 位置处输入:=(E8-6 * B9 * C9)/(D8-6 * B9^2)(如图 1-36),计算出回归系数 b.

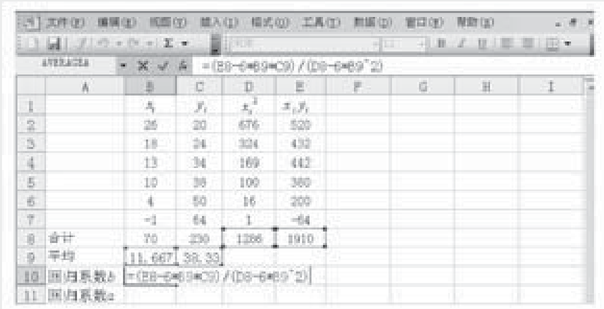


图 1-36

(7) 在第 11 行的 B11 位置处输入:=C9-B10 * B9(如图 1-37),计算出回归系数 a.

图 1-37

习 题 1—8

1. 利用上一节习题中各题的数据,用最小二乘法估计相应变量之间的线性回归方程.
2. 一位运动生理学家根据训练水平(单位:kg · m/min,即每分时间将1 kg物体升高 1 m)来预测血液输出量(每分时间由心脏输出的血液的升数).他选取四个训练水平:0,300,600,900(kg · m/min).随机抽取 20 人的一个样本,随机分成四组,每个水平一组,每组 5 人;训练 15 min 后,测量他们的心脏血液输出量,结果如下表.求这两个变量之间的最小二乘回归直线,若给定训练水平为 700 kg · m/min,预测心脏血液输出量的值.

个体	训练水平/(kg·m/min)	心脏血液输出量/(L/min)
1	0	4.4
2	0	5.6
3	0	5.2
4	0	5.4
5	0	4.4
6	300	9.1
7	300	8.6
8	300	8.5
9	300	9.3
10	300	9.0
11	600	12.8
12	600	13.4
13	600	13.2
14	600	12.6
15	600	13.2
16	900	17.0
17	900	17.3
18	900	16.5
19	900	16.8
20	900	17.2

3. 下面是上海市统计局提供的主要年份的研究生人数情况.

统计年鉴

上海统计网首页 > 统计年鉴 > 上海统计年鉴2003目录 > 表格

表17.5 主要年份研究生人数

单位：人

年份	获博士学位人数	获硕士学位人数	研究生					
			招生数		在读人数		毕业生数	
			高等学校	研究所(院)	高等学校	研究所(院)	高等学校	研究所(院)
1985	57	1 371	4 264	83	8 163	170	1 543	37
1986	37	1 243	3 933	88	10 195	238	1 642	27
1987	70	1 902	3 444	70	10 766	243	2 692	58
1988	249	3 196	3 406	374	10 332	1 354	3 524	474
1989	302	3 198	2 679	325	9 120	1 167	3 367	417
1990	300	2 746	2 803	324	8 533	1 035	2 953	369
1991	342	2 739	2 717	302	8 020	949	2 936	320
1992	343	2 124	3 323	345	8 858	997	2 262	264
1993	368	2 489	3 919	363	10 037	1 008	2 569	315
1994	442	2 363	4 665	465	11 905	1 185	2 608	251
1995	606	2 742	4 776	525	13 378	1 335	3 038	317
1996	627	3 233	5 915	592	15 307	1 528	3 537	323
1997	890	3 585	6 163	562	16 841	1 619	4 117	358
1998	1 090	3 552	7 281	593	19 499	1 663	4 253	389
1999	1 323	4 288	8 758	655	22 656	1 764	5 196	415
2000	1 307	4 546	11 796	856	28 582	2 032	5 435	433
2001	1 487	5 330	14 751	1 075	36 528	2 515	6 380	437
2002	1 735	6 191	17 848	1 363	45 713	3 183	7 481	445

(1) 画出获硕士学位人数和获博士学位人数之间的散点图.

(2) 上面的二者之间是否是线性相关的？如果是，请用最小二乘估计求出博士学位人数关于硕士学位人数的线性回归方程.

(3) 如果 2005 年获硕士学位人数为 10 000 人，请估计获博士学位的人数.

(4) 如果硕士研究生的在读时间为 3 年，当年获硕士学位的人数与第三年后获博士学位的人数之间应当有一个什么关系？请画出散点图，如果成线性关系，用最小二乘法给出线性回归方程.

(5) 比较(4)中的回归方程与(2)中的回归方程之间的差异，你认为哪个方程更具有代表性？

(6) 请试着分析其他数据之间的关系.

阅读材料

统计小史

在人类的发展史上，统计与概率的思想在很多方面对人们决策起到了重要的作用. 比如，18 世纪英国政府为了确定如何开展人寿保险业，对各个年龄段人的死亡情况进行了统计和分析，进而为后来人寿保险的发展提供了重要的科学依据. 再如，生物学上孟德尔遗传学理论的建立就依赖于统计分析，在长期统计分析的基础上形成了科学的理论，为以后的数量遗传学提供了科学的思考方法.

目前，社会上的各行各业都离不开统计学：生物学上有生物统计学；经济学上有计量经济学；教育与心理方面有教育统计学、心理统计学；产品的生产过程中会用到质量控制的有关理论与方法，这也是统计学在起作用；就连律师为了提供有力的证据也离不开统计学；在医学上，为了评估有争议的医学报告，也少不了利用统计学进行分析与论证；在天文学上，需要对大量的天文观测进行统计分析以获取可靠的结论. 目前，一些新兴研究领域也离不开统计与概率，如对策论、风险投资、随机模拟技术等.

统计学在生活中的应用越来越广泛，对某个问题的调查最简单的方法就是普查，但这种调查方法的局限性很大，出于对费用和时间的考虑(如前面所说的，调查对象的量太大或是一些破坏性调查等)，人们认识到需要在调查中进行抽查. 但对抽查结果的可靠性一开始是有怀疑的. 第二次世界大战期间，各交战国为适应急剧变化的战局，急需及时而有效地收集情报，除抽样调查外别无他法，这就促进了对抽样调查的理论和方法的研究. 战后不久，出现了这方面的专著，F. 耶茨受联合国统计抽样专业委员会的委托，为协助 1950 年世界农业和世界人口调查而写的《人口调查与一般调查的抽样方法》就是其中之一. 20 世纪 50 年代后，世界各国已逐渐把抽样调查作为一种重要的调查方法. 这是因为，普查的工作量太大，往往为人力财力时间所不允许，在实施过程中易出现人为的误差；经验表明，有时一个精心设计的抽样方案，其实施的效果甚至可以胜过普查.

课题学习

调查通俗歌曲的流行趋势

问题情境

1987 年的春节联欢晚会，使费翔的“冬天里的一把火”成为当时风靡一时的歌曲，也流行了很长一段时间．但是，现在的中学生对这首歌就不一定很认同，而更多的是喜欢目前流行的歌曲．这就是通俗歌曲流行的趋势．

为了方便分析，我们将一个人对歌曲的喜欢程度进行量化，分为 10 个等级：1，2，3，4，5，6，7，8，9，10．其中“10”表示非常喜欢，“1”表示非常不喜欢．

说　明

本课题学习强调统计活动的全过程，因此任务提出较细化，请同学们按要求认真完成．

根据你和同学们的了解，确定每年最具有代表性的一首通俗歌曲．由调查对象根据他自己的喜好给每首歌曲打分．调查时，要求记下调查对象的性别与年龄，以便为分析提供可靠的证据．

任务 1　请你与同学们一起讨论一个调查方案，然后按照设计好的方案进行调查．

任务 2　根据调查的数据，分析每首通俗歌曲的喜好程度与性别是否有关系．

任务 3　根据调查的数据，分析每首通俗歌曲的喜好程度与年龄是否有关系．

任务 4　根据调查的数据，计算填写下页的表格．

通俗歌曲的名称	通俗歌曲首次推出的年份(A)	被调查人的出生年份(B)	$C=B-A$	喜好程度(D)

以变量 C 为横坐标、以变量 D 为纵坐标，作出散点图，并由此分析变量 D 随着变量 C 的变化趋势.

任务 5 根据调查数据和分析结果，写出调查报告，并在全班进行交流.

实施建议

1. 可以组成课题学习小组，集体讨论，互相启发，分工合作，形成具体可行的调查方案. 调查方案的设计与讨论是非常必要，也是非常重要的，讨论要充分，设计要细致.

2. 在设计调查方案时，一定要讨论调查问卷的设计. 问卷上栏目的设计直接影响调查的结果，要尽可能避免一些敏感性的问题.

3. 调查报告的呈现形式可以参考下一页的表.

4. 成果交流.

建议以小组为单位，选出代表，在班级中报告研究成果，交流研究体会.

5. 评价建议.

在评价中，采用自评、互评、教师评价相结合的形式，应善于发现别人工作中的特色，可主要考虑以下几个方面：

(1) 求解过程和结果：合理、清楚、简捷、正确；

(2) 独到的思考和发现；

(3) 提出有价值的求解设计和有见地的新问题；

(4) 发挥组员的特长，体现合作学习的效果.

________年级　____班

1. 课题组成员、分工、贡献	
成员姓名	分工与完成情况
2. 调查的过程和结果	
3. 主要参考资料	
4. 成果的自我评价（请说明方法或原理的合理性、特色或创新点、不足之处等）	
5. 在调查的过程中发现和提出了哪些新问题？是如何解决的？得到哪些很得意的结论？	
6. 描述在调查研究中的感受	

回顾与反思

1. 你学习了哪些抽样方法？它们各自有什么特点？

2. 表示数据的统计图表有哪些？它们各有什么特点？请分别举例说明.

3. 样本的数字特征有哪些？分别是如何刻画的？

4. 在从事统计活动的过程中，你经历了哪些主要的步骤？

5. 如何看待变量之间的相关性？如何进行刻画呢？

◆ 本章小结

一、内容提要

本章的学习内容主要由随机抽样、用样本估计总体以及变量的相关性这三个部分组成.

1. 在“随机抽样”的内容中，首先从具体的实例展开，主要从总体的量很大，对总体的抽样具有破坏性两个方面讨论了抽样的必要性；接着从一些统计误导的例子谈起，体会什么样的样本才具有代表性；最后介绍了三种比较典型和常用的抽样方法，即简单随机抽样、分层抽样和系统抽样，通过具体的问题体会不同抽样方法各自的优越性与局限性，并针对不同的问题选择适当的抽样方法.

2. 在“用样本估计总体”的内容中，首先复习了象形统计图、条形统计图、折线统计图、扇形统计图等，学习了一种新的统计图——茎叶图，通过具体的问题不断体会它们各自的特点和用途，并有针对性地选择合适的统计图表；接着介绍了数据的数字特征，在平均数、中位数、众数、极差、方差等的基础上，着重介绍了数字特征——标准差，并结合具体情境理解不同数字特征的意义，以及根据问题的需要选择适当的数字特征来表达数据的信息；最后介绍了用样本估计总体的方法——用样本的频率分布(频率分布表、频率分布直方图、频率折线图)估计总体分布，用样本的基本数字特征估计总体的基本数字特征，并初步体会样本频率分布和数字特征的随机性.

3. 在“变量的相关性”的内容中，首先通过例子介绍了现实生活中存在的不满足函数关系的一些量，如人的身高与体重、人的身高与右手一拃长；通过变量之间的散点图，探索用不同的方法确定线性回归直线；在此基础上，介绍了最小二乘的方法，体会最小二乘的思想，并会根据给出的公式求线性回归方程；最后，再通过具体的例子初步体会用样本数据拟合直线方程的实际意义以及拟合结果的随机性.

当然，统计的学习最好通过案例来进行. 因此，教科书还设计了两个大的统计活动：结婚年龄的变化和通俗歌曲的流行趋

势，并在活动的要求上设计了一定的层次．通过这两个统计活动，经历较为系统的数据处理全过程，在此过程中学习一些数据处理的方法，并运用所学的知识和方法去解决实际问题．

在完成了这一章内容的学习后，建议同学先回顾本章的内容，梳理知识结构，再互相讨论，并共同建立以下的知识框架图．

实际问题
↓
确定调查对象
↓
收集数据 —— 普查；抽样调查：简单随机抽样、分层抽样、系统抽样
↓
整理数据 —— 统计图表：条形统计图、条形统计图、折线统计图、扇形统计图；茎叶图、频率分布表、频率分布直方图、频率折线图
↓
分析数据 —— 数字特征：集中趋势：平均数、中位数、众数；离散程度：极差、方差、标准差
↓
作出推断 —— 用样本估计总体：用样本的频率分布估计总体的分布；用样本的数字特征估计总体的数字特征

二、学习要求和需要注意的问题

1．学习要求

（1）结合具体的实际问题情境，理解随机抽样的必要性和重要性．在参与解决统计问题的过程中，学会用简单随机抽样方法从总体中抽取样本；通过对实例的分析，了解分层抽样和系统抽样方法．

（2）通过实例体会分布的意义和作用，在表示样本数据的过程中，学会列频率分布表，画频率分布直方图、频率折线图、茎叶图，体会它们各自的特点；会从样本数据中提取基本的数字特征（如平均数、标准差），并作出合理的解释．

（3）体会用样本估计总体的思想，会用样本的频率分布估计总体分布，会用样本的基本数字特征估计总体的基本数字特征；初步体会样本频率分布和数字特征的随机性．

（4）利用散点图直观认识变量间的相关关系，经历用不同估算方法描述两个变量线性相关的过程；知道最小二乘法的思想，并能根据给出的线性回归方程系数公式建立线性回归方程.

2. 需要注意的问题

（1）突出处理的对象和数据的现实背景. 在具体情境中体会不同抽样方法各自的优越性与局限性，体会不同统计图的特点和用途，理解不同数字特征的意义，并能根据问题的需要选择适当的抽样方法收集数据，选择适当的统计图表和数字特征来表达数据的信息.

（2）统计的学习，本质上是统计活动的学习，而不是概念和公式的学习. 要特别注重统计的过程，即经历“收集数据—整理数据—分析数据—作出推断”的全过程. 统计学习的要点是“做”而不是记忆和运算，在活动中鼓励使用计算器或计算机处理复杂的数据.

（3）感受统计的意义和价值，发展初步的统计观念，培养统计意识，体会统计思维与确定性思维的差异. 在用统计解决问题、作出决策之前，思考需要收集哪些信息，如何收集信息，如何表达信息. 不断体会为什么要用统计，统计能给我们带来什么，在统计活动中真正理解统计对决策的作用.

复 习 题 一

A 组

1. 下表给出了某个地区近年来 1 000 次火灾发生的不同原因,请你用自己的方式表示其中的数据.

原 因	烹 饪	电 器	明 火	吸 烟	纵 火	儿童玩火	其 他
次 数	168	270	80	176	140	60	106

2. 某产品售后服务中心随机选取了 20 个工作日,分别记录下每天接到的客户服务电话的数量:

63　38　25　42　56　48　53　39　28　47

45　52　59　48　51　62　48　50　52　38

(1) 请用不同的方式分别表示上面的数据;

(2) 分别计算以上数据的平均数、中位数和众数;

(3) 根据以上结果,你能为该产品的售后服务中心提供什么建议?

3. 下表给出了 2000 年西安和桂林的月降水量(单位:mm):

	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
西安	9.2	4.9	5.4	18.6	38.0	106.3	54.4	128.9	62.9	73.6	26.2	10.6
桂林	41.4	53.3	178.8	273.5	384.9	432.4	67.5	228.5	201.4	147.3	28.0	19.1

(1) 请用适当的统计图表示上面的数据;

(2) 分别计算西安和桂林 2000 年月降水量的平均数和标准差;

(3) 试比较西安和桂林两地的降水量.

4. 某学校对男女学生进行有关“习惯与礼貌”的评分,记录如下:

男:54, 70, 57, 46, 90, 58, 63, 46, 85, 73, 55, 66, 38, 44, 56, 75, 35, 58, 94, 52;

女:77, 55, 69, 58, 76, 70, 77, 89, 51, 52, 63, 63, 69, 83, 83, 65, 100, 74.

(1) 请用茎叶图表示上面的数据,并从图中分别比较男女生得分的平均数、标准差的大小;

(2) 分别计算男女生得分的平均数、标准差,由此,你能得出什么结论?

5. 为了解 A,B 两种轮胎的性能,某汽车制造厂分别从这两种轮胎中随机抽取了 8 个进行测试,下面列出了每一个轮胎行驶的最远里程数(单位:1 000 km):

轮胎 A:96, 112, 97, 108, 100, 103, 86, 98;

轮胎 B:108, 101, 94, 105, 96, 93, 97, 106.

(1) 分别计算 A,B 两种轮胎行驶的最远里程的平均数、中位数;

(2) 分别计算 A,B 两种轮胎行驶的最远里程的极差、标准差;

(3) 你认为哪种型号的轮胎性能更加稳定?

6. 为了解某种干电池的寿命,电池厂随机抽取了 50 节进行测试,下面列出了每一节电池的使用寿命(单位:h):

11　14　25　13　11　20　15　30　9　16

13　10　14　11　10　16　19　12　0　20

16　10　15　14　22　19　10　33　3　12

16　19　23　15　20　11　17　14　23　15

12　15　12　10　13　11　9　8　13　17

（1）完成下表，并画出相应的频率分布直方图和频率折线图；

使用寿命分组(Δx_i)	频　数(n_i)	频　率(f_i)	$\frac{f_i}{\Delta x_i}$
0～5 h			
5～10 h			
10～15 h			
15～20 h			
20～25 h			
25～30 h			
30～35 h			

（2）以上电池使用寿命的平均数、中位数、众数分别是多少？

（3）由此，你能估计这种干电池的使用寿命吗？

7．下面给出了 1999 年世界一些国家国民平均寿命（单位：岁）：

国　家	平均寿命	国　家	平均寿命	国　家	平均寿命	国　家	平均寿命
中国	70.1	印度	63.2	印度尼西亚	65.7	日本	80.6
韩国	72.9	马来西亚	72.3	巴基斯坦	62.5	泰国	68.6
土耳其	69.5	埃及	66.8	南非	48.5	尼日利亚	47.5
墨西哥	72.1	美国	76.9	加拿大	79.0	阿根廷	73.6
巴西	67.2	法国	78.5	德国	77.0	意大利	78.3
荷兰	77.7	波兰	73.2	罗马尼亚	69.5	俄罗斯	65.8
西班牙	77.9	乌克兰	67.3	英国	77.2	澳大利亚	78.8

（1）将数据进行适当的分组，并画出相应的频率分布直方图和频率折线图；

（2）以上所有国家国民平均寿命的平均数是多少？标准差呢？

（3）由此，你能估计世界人民的平均寿命吗？

8．下面是某年世界卫生组织公布的几个国家男性与女性的平均寿命情况（单位：岁）：

国　家	男性平均寿命(x)	女性平均寿命(y)
中国	66.7	70.4
韩国	67.7	75.7
马来西亚	68.7	73.0
美国	72.0	78.9
法国	72.9	81.1
新加坡	74.0	78.3
日本	76.2	82.5

（1）将上表中的数据制成散点图；

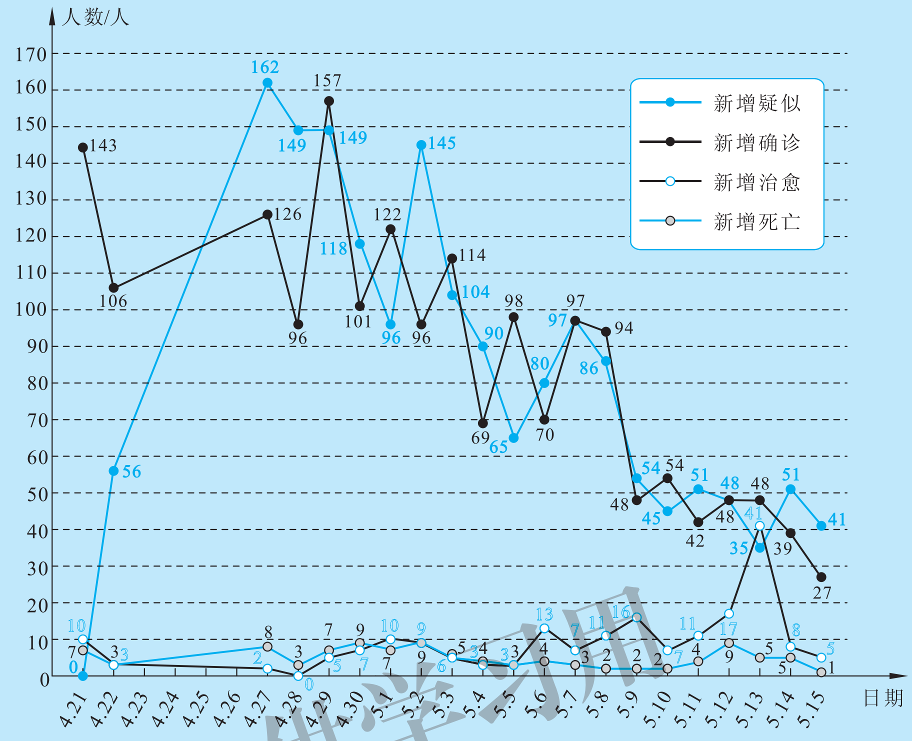
（2）你能从散点图中发现男性年龄与女性年龄近似成什么关系吗？

（3）如果近似成线性关系，请用最小二乘法估计这条直线的方程；

（4）如果瑞典男性平均寿命是 75.5 岁，预测瑞典女性的平均寿命.

B　组

1. 下面给出了 2003 年 4 月 21 日至 5 月 15 日上午 10 时，北京市非典型性肺炎疫情新增数据走势图.



（第 1 题）

（1）哪一天新增确诊的人数最多？哪一天新增疑似的人数最多？

（2）哪一天新增治愈的人数最多？哪一天新增死亡的人数最少？

（3）从图中，你能预测这次北京市非典型性肺炎疫情的发展趋势吗？

2. 1798 年英国科学家 Henry Cavendish 对地球密度进行了测量，下面是地球密度相对于水密度的记录结果：

5.50　5.61　4.88　5.07　5.26　5.55　5.36　5.29　5.79　5.10　5.27　5.39　5.42　5.47　5.58
5.65　5.57　5.53　5.62　5.29　5.44　5.34　5.63　5.34　5.46　5.30　5.75　5.68　5.85

请用适当的统计图表示上述测量的数据，并估计地球的密度(单位：kg/m^3).

3. 要调查睡眠时间与年龄之间的关系，请选择合适的调查方法，并将调查数据画出散点图，看这二者之间是否成线性关系. 如果是，请用最小二乘法进行估计，写出回归方程.

4. 请选择一种合适的调查方法，在你周围的人群中，调查吸烟人的心率与吸烟时间的关系，并画出散点图，看这二者之间是否成线性关系. 如果是，请用最小二乘法进行估计，求出线性回归方程. 写出一个调查报告，对吸烟人提出建议.

5. 设有 n 个数据：$x_1, x_2, \cdots, x_n$，利用二次函数的性质，试求当 a 取何值时，$(x_1-a)^2+(x_2-a)^2+\cdots+(x_n-a)^2$ 达到最小值.

6. 设有 n 个点：$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$，试利用最小二乘法和二次函数的性质，推导这 n 个点的线性回归方程.

第二章 算法初步

随着计算机科学和信息技术的飞速发展，算法的思想已经渗透到社会的方方面面. 在以前的学习中，虽然没有出现算法这个名词，但实际上在数学教学中已经渗透了大量的算法思想，如四则运算的过程、求解方程的步骤等. 完成这些工作都需要一系列程序化的步骤，这就是算法的思想.

在本章中，我们将介绍算法的基本思想、基本结构和描述算法的基本语句. 这对我们理解数学与计算机技术的关系是很有帮助的.

在中国漫长的历史中，数学曾有许多重要的发展……第一个算学家是刘徽……第二个算学家是祖冲之.

20 世纪数学的……另一个现象是计算机的介入. 计算机引发了许多新的课题，如递归函数、复杂性、分形等.

——著名数学家

陈省身

§1　算法的基本思想

在本节中，我们将通过几个实例来体会算法的基本思想.

> **算　法**
> 是对问题求解方法的精确描述.

例 1　在电视台的某个娱乐节目中，要求参与者快速猜出物品价格. 主持人出示某件物品，参与者每次估算出一个价格，主持人只能回答高了、低了或者正确.

在某次节目中，主持人出示了一台价值在 1 000 元以内的随身听，并开始了竞猜. 下面是主持人和参与者之间的一段对话：

参与者：800 元！

主持人：高了！

参与者：400 元！

主持人：低了！

参与者：600 元！

主持人：低了！

……

如果你是参与者，你接下来会怎么猜？

分析理解

如果用 P 表示商品的价格，则参与者的猜测结果为：

由主持人的第一次回答，P 为 0～800 元；

由主持人的第二次回答，P 为 400～800 元；

由主持人的第三次回答，P 为 600～800 元.

根据参与者的猜测，我们知道，参与者首先需要确定的是商品价格的范围，数学上一般可以用区间来表示，然后根据主持人的回答，报出区间中点，将价格区间缩小一半.

因此，下一步参与者要猜的数应是 700 元，然后根据主持人的回答继续猜，直至猜出正确价格.

抽象概括

实际上，可以把上述过程概括如下：

1. 报出首次价格 T_1；

2. 根据主持人的回答确定价格区间：

　(1) 若报价小于商品价格，则商品的价格区间为$(T_1, 1\ 000)$；

　(2) 若报价大于商品价格，则商品的价格区间为$(0, T_1)$；

　(3) 若报价等于商品价格，则游戏结束.

3. 如果游戏没有结束，则报出上面确定的价格区间的中点 T_2.

按照上述方法，继续判断，直到游戏结束. 像这样的一系列步骤通常称为解决这个问题的一个算法.

例 2　在给定素数表的条件下，设计算法，将 936 分解成素因数的乘积.（4 000 以内的素数表见附录 1）

分析理解

1. 查表判断 936 是否为素数：

(1) 如果 936 是素数，则分解结束；

(2) 如果 936 不是素数，则进行第 2 步.

2. 确定 936 的最小素因数：2.　　$936=2\times468$.

3. 查表判断 468 是否为素数：

(1) 如果 468 是素数，则分解结束；

(2) 如果 468 不是素数，则重复上述步骤，确定 468 的最小素因数.

重复进行上述步骤，直到找出 936 的所有素因数.

解　算法步骤如下：

1. 判断 936 是否为素数：否.

2. 确定 936 的最小素因数：2.　　$936=2\times468$.

3. 判断 468 是否为素数：否.

4. 确定 468 的最小素因数：2.　　$936=2\times2\times234$.

5. 判断 234 是否为素数：否.

6. 确定 234 的最小素因数：2.　　$936=2\times2\times2\times117$.

7. 判断 117 是否为素数：否.

8. 确定 117 的最小素因数：3.　　$936=2\times2\times2\times3\times39$.

9. 判断 39 是否为素数：否.

10. 确定 39 的最小素因数：3.　　$936=2\times2\times2\times3\times3\times13$.

11. 判断 13 是否为素数：13 是素数，所以分解结束.

分解结果是：

$$936=2\times2\times2\times3\times3\times13.$$

短除法可以使这个过程更清晰，见图 2-1.

短除法

$$\begin{array}{r|l}
2 & 9\ \ 3\ \ 6 \\ \hline
2 & 4\ \ 6\ \ 8 \\ \hline
2 & 2\ \ 3\ \ 4 \\ \hline
3 & 1\ \ 1\ \ 7 \\ \hline
\end{array}$$
$$\begin{array}{r|l}
3 & 3\ \ 9 \\ \hline
 & 1\ \ 3
\end{array}$$

图 2-1

按照以上程序，完成了对 936 的素因数分解. 实际上，在给定素数表的基础上，对任意自然数 n，都可以按照上述办法进行分解.

以上步骤是解决素因数分解问题的一个过程，只要依照这一系列步骤，都能解决这个问题. 我们把这一系列步骤称为解决这个问题的一个算法.

例 3　设计一个算法，求 840 与 1 764 的最大公因数.

分析理解

我们已经学习了对自然数进行素因数分解的方法，下面的算法就是在此基础上设计的. 解答这个问题需要按以下思路进行.

首先，对两个数分别进行素因数分解：

$840=2^3\times3\times5\times7$，　　$1\ 764=2^2\times3^2\times7^2$.

其次，确定两数的公共素因数：2，3，7.

接着，确定公共素因数的指数：对于公共素因数 2，2^2 是 1 764 的因数，2^3 是 840 的因数，因此 2^2 是这两个数的公因数，这样就确定了公共素因数 2 的指数为 2. 同样，可以确定出公因数 3 和 7 的指数均为 1. 这样，就确定了 840 与 1 764 的最大公因数为：

$$2^2\times3^1\times7^1=84.$$

说　明

在数学软件或图形计算器中，命令 isPrime(　) 可以判断给定的一个自然数是否为素数；命令 factor(　) 可以对给定自然数进行素因数分解.

解　算法步骤如下：

1. 先将 840 进行素因数分解：$840=2^3\times3\times5\times7$；
2. 然后将 1 764 进行素因数分解：$1\ 764=2^2\times3^2\times7^2$；
3. 确定它们的公共素因数：2，3，7；
4. 确定公共素因数的指数：公共素因数 2，3，7 的指数分别为 2，1，1；
5. 最大公因数为：$2^2\times3^1\times7^1=84$.

以上步骤就是求两个正整数的最大公因数的一个算法. 这个算法的思想具有一般性，它可以帮助设计求三个或者三个以上正整数的最大公因数的算法. 在这个算法的设计中，我们首先分别对 840 和 1 764 进行素因数分解，那么，如何将 840 或 1 764 的所有素因数都找出来呢？例 2 给出了具体的算法. 我们通常会把“对自然数进行素因

数分解”的算法，做成一个“程序包”，又称为一个“平台”，在需要“对自然数进行素因数分解”时，把做好的“程序包”拿来使用即可. 同样的，我们也可以把“求两个自然数的最大公因数”做成一个“程序包”或“平台”，在解决其他问题时，如果需要“求两个自然数的最大公因数”，我们就可以把做好的程序包直接拿来使用. 通常把这种思想称为“平台思想”，“平台”的思想在算法设计中是一个最基本的思想，也是数学中思考问题的一个重要思想.

通过以上的例子可以看出，算法是解决某类问题的一系列步骤或程序，只要按照这些步骤执行，都能使问题得到解决. 一般来说，“用算法解决问题”都是可以利用计算机帮助完成的.

练 习 1

1. 模仿例 3，设计一个算法，求 324，440，556 的最大公因数.
2. 设计算法，求 1 356 和 2 400 的最小公倍数.

例 4 “韩信点兵”问题.

韩信是汉高祖刘邦手下的大将，他英勇善战，智谋超群，为建立汉朝立下了汗马功劳. 据说他在点兵的时候，为了保住军事机密，不让敌人知道自己部队的实力，采用下述点兵方法：先令士兵从 1～3 报数，结果最后一个士兵报 2；再令士兵从 1～5 报数，结果最后一个士兵报 3；又令士兵从 1～7 报数，结果最后一个士兵报 4. 这样，韩信很快就算出了自己部队士兵的总人数. 请设计一个算法，求出士兵至少有多少人.

韩信

分析理解

士兵从 1～3 报数，最后一个士兵报 2. 这句话说明士兵的总人数除以 3 余 2.

算法的第 1 步是将所有除以 3 余 2 的正整数找出来，按照从小到大的顺序排成一列数.

士兵从 1～5 报数，最后一个士兵报 3. 这句话说明士兵的总人数除以 5 余 3.

算法的第 2 步是从上列数中找出除以 5 余 3 的一列数.

最后，在满足前两个条件的上列数中再找出除以 7 余 4 的一列数，这列数中最小的数，即为我们所求的数.

解 算法步骤如下：

1. 首先确定最小的满足除以 3 余 2 的正整数：2；

2. 依次加 3 就得到所有除以 3 余 2 的正整数：

$$2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32,$$
$$35, 38, 41, 44, 47, 50, 53, 56, \cdots$$

3. 在上列数中确定最小的满足除以 5 余 3 的正整数:8;

4. 然后依次加上 15,得到

$$8, 23, 38, 53, \cdots$$

不难看出,这些数既满足除以 3 余 2,又满足除以 5 余 3;

5. 在第 4 步得到的一列数中找出满足除以 7 余 4 的最小数 53,这就是我们要求的数.

在完成上述步骤后,就找到了所求的数 53,这 5 个步骤称为解决“韩信点兵”问题的一个算法.

大家可以想一想,能否在这个算法的基础上作一些改进,以更快地获得结果?

实际上,我们颠倒一下 3,5,7 的顺序,就可以得到另一个算法:

1. 首先确定最小的除以 7 余 4 的正整数:4;

2. 依次加 7 就得到所有除以 7 余 4 的正整数:

$$4, 11, 18, 25, 32, 39, 46, 53, 60, \cdots$$

3. 在第 2 步得到的一列数中确定最小的除以 5 余 3 的正整数:18;

4. 然后依次加上 35,得到

$$18, 53, 88, \cdots$$

5. 在第 4 步得到的一列数中找出最小的满足除以 3 余 2 的正整数:53.

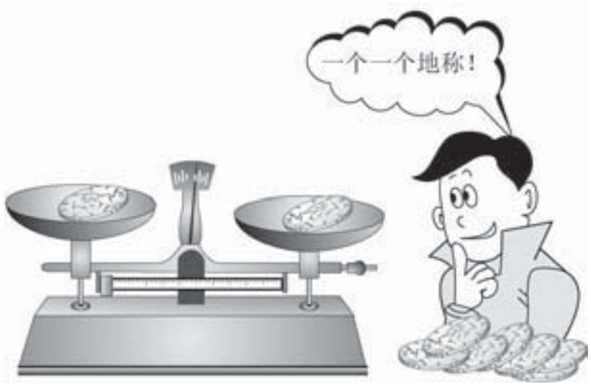
思考交流

从例 4 的两个算法中,你能得到哪些启示?

例 5　一位商人有 9 枚银元,其中有 1 枚略轻的是假银元.你能用天平(不用砝码)将假银元找出来吗?

分析理解

最容易想到的解决这个问题的一种方法是:把 9 枚银元按顺序排成一列,先称前 2 枚,若不平衡,则可找出假银元;若平衡,则 2 枚银元都是真的,再依次与剩下的银元比较,就能找出假银元.

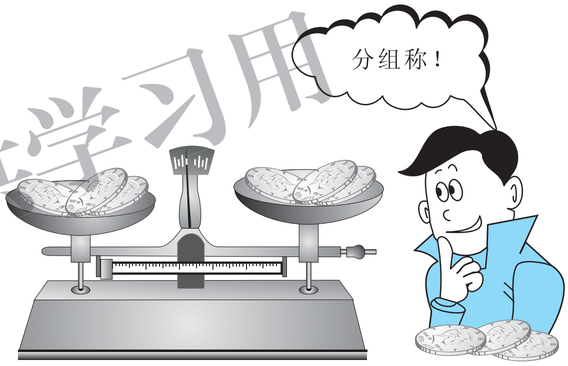


解 按照下列步骤，就能将假银元找出来：

1. 任取 2 枚银元分别放在天平的两边. 如果天平左右不平衡，则轻的一边就是假银元；如果天平平衡，则进行第 2 步.

2. 取下右边的银元，放在一边，然后把剩余的 7 枚银元依次放在右边进行称量，直到天平不平衡，偏轻的那一枚就是假银元.

这种算法最少要称 1 次，最多要称 7 次. 是不是还有更好的办法，使得称量次数少一些？我们可以采用下面的方法：



1. 把银元分成 3 组，每组 3 枚.

2. 先将两组分别放在天平的两边. 如果天平不平衡，那么假银元就在轻的那一组；如果天平左右平衡，则假银元就在未称的第 3 组里.

3. 取出含假银元的那一组，从中任取两枚银元放在天平的两边. 如果左右不平衡，则轻的那一边就是假银元；如果天平两边平衡，则未称的那一枚就是假银元.

经分析发现，这种算法只需称量 2 次，这种做法要明显好于前一种做法. 当然，这两种方法都具有一般性，同样适用于 n 枚银元的情形. 这是信息论中的一个模型，可以帮助我们找出某些特殊信息.

从以上两个问题中可以看出，同一个问题可能存在着多种算法，其中一些可能要比另一些好. 在实际问题和算法理论中，找出好的算法是一项重要的工作.

练 习 2

1. 有一把围棋子，5 个 5 个地数，最后余下 2 个；7 个 7 个地数，最后余下 3 个；9 个 9 个地数，最后余下 4 个. 请设计一种算法，求出这把棋子至少有多少个.
2. 一个大油瓶装 8 kg 油. 还有两个空油瓶，一个能装 5 kg 油，另一个能装 3 kg 油. 请设计一种算法，将这 8 kg 油平均分成两份.

对于一元二次方程，可以用熟悉的求根公式来求解. 但是，绝大部分方程不存在求根公式[1].

在实际问题中，通常只要能获得满足一定精度的近似解就可以了，因此，讨论方程近似解的算法具有重要的现实意义.

在函数的应用部分，我们学习了用二分法求方程 $f(x)=0$ 的近似解(如图 2-2). 二分法求方程近似解的基本思想是：将方程的有解区间平分为两个小区间，然后判断解在哪个小区间；继续把有解的区间一分为二进行判断，如此周而复始，直到求出满足精度要求的近似解.

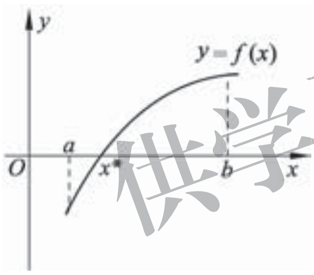


图 2-2

其算法步骤如下(要求近似解的精度为 10^{-n})：

1. 确定有解区间 $[a,b]$ ($f(a)\cdot f(b)<0$).
2. 取 $[a,b]$ 的中点 $x=\dfrac{a+b}{2}$.
3. 计算函数 $f(x)$ 在中点处的函数值 $f\left(\dfrac{a+b}{2}\right)$.
4. 判断函数值 $f\left(\dfrac{a+b}{2}\right)$ 是否为 0：

(1) 如果为 0，$x=\dfrac{a+b}{2}$ 就是方程的解，问题就得到了解决；

(2) 如果不为 0，则分下列两种情形：

① 若 $f(a)\cdot f\left(\dfrac{a+b}{2}\right)<0$，则确定新的有解区间为 $\left(a,\dfrac{a+b}{2}\right)$；

② 若 $f(a)\cdot f\left(\dfrac{a+b}{2}\right)>0$，则确定新的有解区间为 $\left(\dfrac{a+b}{2},b\right)$.

[1] 例如，五次和五次以上的高次方程就不存在求根公式，因此，讨论求方程近似解的方法具有重要的意义，例如二分法、切线法、割线法等. 讨论五次以上高次方程的求解问题曾经极大地推动了数学的发展，数学家伽罗瓦在这方面作出了重大贡献.

伽罗瓦(Galois，1811—1832)生于法国巴黎郊区布拉伦镇. 这位优秀的数学家一生坎坷，英年早逝. 在他 1829 年至 1831 年完成的几篇论文中，建立了判别方程根式可解的充分必要条件，从而宣告了方程根式可解这一经历了 300 年的难题的彻底解决，开创了群论的先河.

5. 判断新的有解区间的长度是否大于精度：

(1)如果新的有解区间长度大于精度，则在新的有解区间的基础上重复上述步骤；

(2)如果新的有解区间长度小于或等于精度，则这个有解区间中的任意一个数均为方程的满足精度的近似解.

下面，结合具体实例，体会二分法算法的实现过程.

例 6 求方程 $x^3+x^2-1=0$ 在[0,1]上的近似解，精度为 0.1.

解 根据上述分析，可以通过下列步骤求得方程的近似解：

设 $f(x)=x^3+x^2-1$，

1. 因为 $f(0)=-1$，$f(1)=1$，$f(0)\cdot f(1)<0$，则区间[0,1]为有解区间；
2. 取[0,1]的区间中点 0.5；
3. 计算 $f(0.5)=-0.625$；
4. 由于 $f(0.5)\cdot f(1)<0$，可得新的有解区间[0.5,1]，
   $1-0.5=0.5>0.1$；
5. 取[0.5,1]的区间中点 0.75；
6. 计算 $f(0.75)=-0.015\,625$；
7. 由于 $f(0.75)\cdot f(1)<0$，可得新的有解区间[0.75,1]，
   $1-0.75=0.25>0.1$；
8. 取[0.75,1]的区间中点 0.875；
9. 计算 $f(0.875)=0.435\,546\,875$；
10. 由于 $f(0.75)\cdot f(0.875)<0$，可得新的有解区间[0.75，0.875]，
    $0.875-0.75=0.125>0.1$；
11. 取[0.75，0.875]的区间中点 0.812 5；
12. 计算 $f(0.812\,5)=0.196\,533\,203\,125$；
13. 由于 $f(0.75)\cdot f(0.812\,5)<0$，可得新的有解区间[0.75，0.812 5]，
    $0.812\,5-0.75=0.062\,5<0.1$.

所以，区间[0.75，0.812 5]中的任一数值，都可以作为方程的近似解.

经过分析，不难发现，二分法求方程近似解的基本思想为：逐渐缩小有解区间的长度，直到满足精度的要求. 这种方法可以应用于求一元非线性方程[1]的近似解.

[1]一元非线性方程包括一元高次方程、对数方程、指数方程、三角方程等.

抽象概括

通过对以上几个问题的分析，我们对算法有了一个初步的了解. 在解决某些问题时，需要设计出一系列可操作或可计算的步骤，通过实施这些步骤来解决问题，通常把这些步骤称为解决这些问题的算法. 这种描述不是算法的严格定义，但是反映了算法的基本思想.

在以上的学习中，同学们会有一种突出的感觉：解决这些问题的基本思想并不复杂，很清晰，但是叙述起来很烦琐，有的步骤非常多，有的计算量很大，有时候完全依靠人工完成这些工作很困难. 但是，这恰恰是计算机的特长，它能不厌其烦地完成枯燥的、重复的、烦琐的工作.

因此，现代算法的作用之一是使计算机能代替人完成某些工作，这是学习算法的重要原因之一. 下面，我们将讨论算法设计的基本结构和表述算法的基本语句，进一步体会算法的基本思想，体会算法与计算机的联系.

练习 3

1. 请设计二分法算法，求方程 $x^3-x-1=0$ 在区间 $[1,1.5]$ 内的解（精度为 0.01）.
2. 设计一个算法，求函数 $y=\log_2 x$，当 $x=3$ 时的函数值（精度为 0.001）.

习题 2—1

A　组

1. 设计算法，找出 3 个数中的最大数.
2. 设计算法，将 3 个数按从大到小的顺序排列.
3. 描述一元二次方程求解的算法.
4. 设计一个算法，求方程 $5x+2y=22$ 的正整数解.
5. 设计算法，作一个以已知线段长为边长的正三角形.
6. 设计一个算法，求方程组 $\begin{cases}4x+5y+2z=30,\\5x-2y+4z=21\end{cases}$ 的非负整数解.

B　组

1. 已知 2006 年 5 月 26 日是星期五，请设计算法，确定 2007 年 5 月内的任意一天是星期几.
2. 描述二元一次方程组求解的算法.
3. 查阅中国古代数学家刘徽割圆术求圆面积的有关资料，试着用语言描述这种算法.

阅读材料

物不知数

古希腊数学家崇尚演绎，中世纪的东方数学则表现出强烈的算法精神，特别是中国和印度数学，不讲究命题的形式推导，注重算法的实现. 这些特点在中国的古代数学著作中有突出的体现.

“韩信点兵”问题在我国古代数学史上有不少有趣的名称. 如“物不知其数”“鬼谷算”“隔墙算”“大衍求一术”等. 其中在《孙子算经》中，有“物不知其数”一问，原文如下：“今有物不知其数，三三数之剩二，五五数之剩三，七七数之剩二，问物有几何？”

程大位[1]在《算法统宗》中用下列诗句来描述了这个问题的算法：

三人同行七十稀，
五树梅花廿一枝.
七子团圆正半月，
除百零五便得知.

这几句诗句的意思是：

用 70 乘被 3 除所得的余数：$70\times2=140$；

用 21 乘被 5 除所得的余数：$21\times3=63$；

用 15 乘被 7 除所得的余数：$15\times2=30$；

然后将这三个数加起来：$70\times2+21\times3+15\times2=233$；

最后，$233-105-105=23$.

你知道 70，21，15 这三个数是怎么来的吗？它们具有什么特征？这个算法的原理是什么？


程大位

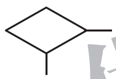
[1] 程大位（1533—1606），中国明朝著名数学家，他编撰的《算法统宗》传入日本、朝鲜及东南亚，并对这些地区的数学发展产生了很大的影响.

§2　算法框图的基本结构及设计

在算法设计中，算法框图（也叫程序框图）可以准确、清晰、直观地表达解决问题的思路和步骤，本节将介绍算法框图的三种基本结构：顺序结构、选择结构和循环结构.

表 2-1 列出了几个基本的框图和它们各自表示的功能.

表 2 1

	框　图	功　　能
终端框 （起止框）		表示一个算法的起始和结束
输入、输出框		表示一个算法输入和输出的信息
处理框		赋值、计算
判断框		判断某一条件是否成立

2.1　顺序结构与选择结构

问题提出

首先，请看一个问题.

例 1　尺规作图，确定线段 AB 的一个 5 等分点.

分析理解

确定线段 AB 的 5 等分点，是指在线段 AB 上确定一点 M，使得 $AM=\frac{1}{5}AB$. 同学们都熟悉解决这个问题的方法：

第一，从 A 点出发作一条与原直线不重合的射线；

第二，任取射线上一点 C，并在射线上作线段 AD，使 $AD=5AC$；

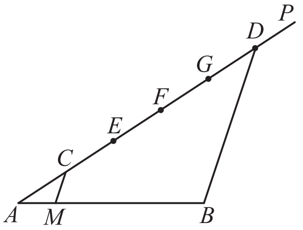


图 2-3

第三，连接 DB，并过 C 点作 BD 的平行线交 AB 于点 M，点 M 就是要找的 5 等分点.

这个过程也需要一步一步来实现.

作法　作图步骤如下：

1. 如图 2-3，从已知线段的左端点 A 出发，作一条射线 AP；
2. 在射线上任取一点 C，得线段 AC；
3. 在射线上作线段 $CE=AC$；
4. 在射线上作线段 $EF=AC$；
5. 在射线上作线段 $FG=AC$；
6. 在射线上作线段 $GD=AC$，那么线段 $AD=5AC$；
7. 连接 DB；
8. 过 C 作 BD 的平行线，交线段 AB 于点 M，这样点 M 就是线段 AB 的一个 5 等分点.

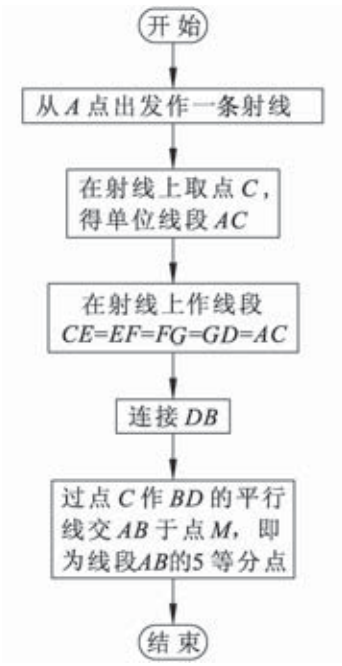


图 2-4

这个实现过程可以用图 2-4 来表示.

这一系列步骤就是解决这个作图问题的一个算法. 只要依次执行这些步骤，就能确定线段的 5 等分点. 这个算法具有一般性，对于任意自然数 n，都可以按照这个算法的思想，设计出确定线段 n 等分点的步骤，得到解决这个问题的一般算法.

像这样，按照步骤依次执行的一个算法，称为具有“**顺序结构**”的算法，或者称为算法的顺序结构.

通常，为了使算法结构更加清晰，可借助图来帮助描述算法. 图的特点是直观、清楚，便于检查和交流. 顺序结构的图见图 2-5. 通常，像这样的图叫作框图.

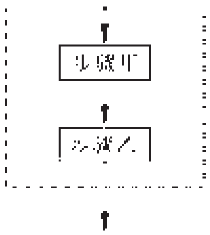


图 2-5

思考交流

(1) 你能确定任意给定线段的 8 等分点吗？16 等分点、64 等分点、100 等分点呢？

(2) 你还有更好的算法吗？

例 2　通常说一年有 365 天，它表示地球围绕太阳转一周所需要的时间，但事实并不是这样简单. 根据天文资料，地球围绕太阳转一周所需要的精确时间是 365.242 2 天，称之为天文年. 这个误差看似不大，却引起季节和日历之间难以预料的大变动. 在历法上规定四年一闰，百年少一闰，每四百年又加一闰. 如何判断某一年是不是闰年呢？请设计一个算法，解决这个问题，并用框图描述这个算法.

分析理解

设 y 为年份，按照历法的规定，如果 y 为闰年，那么或者 y 能被 4 整除不能被 100 整除，或者 y 能被 400 整除.

对于给定的年份 y，要确定它是否为闰年，需要进行判断，判断的结果决定后面的步骤，像这样的结构通常称作**选择结构**. 选择结构的算法框图可以用图 2-6 来表示.

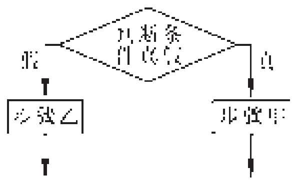


图 2-6

解　算法步骤如下：

1. 若 y 不能被 4 整除，则输出“y 不是闰年”.
2. 若 y 能被 4 整除，则判断 y 是否能被 100 整除：

(1) 若 y 不能被 100 整除，则输出“y 是闰年”；

(2) 若 y 能被 100 整除，则判断 y 是否能被 400 整除：

① 若 y 能被 400 整除，则输出“y 是闰年”；

② 若 y 不能被 400 整除，则输出“y 不是闰年”.

这个算法的框图如图 2-7：

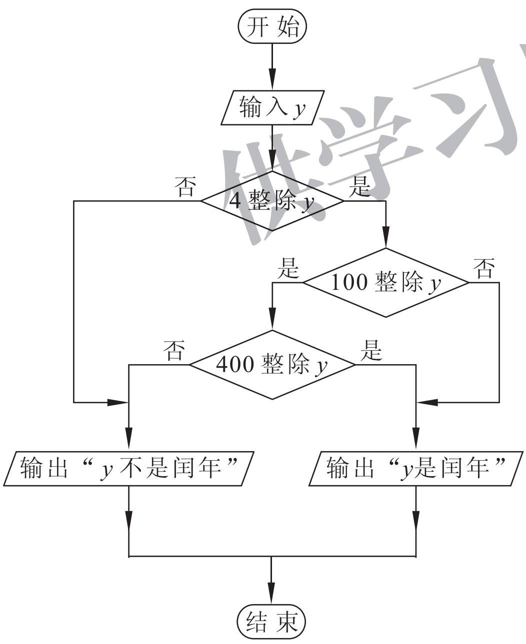


图 2-7

信息技术建议

利用计算机语言（如 BASIC 语言）可以方便地实现判断某一年是否是闰年的算法框图. 具体操作步骤见本节的“信息技术应用”栏目.

信息技术应用

确定年份是否为闰年

使用计算机语言可以实现确定给定年份是否为闰年的算法，设 years 表示给定年份，下面以 BASIC 语言为例给出其具体程序：

说　明

在数学软件或程序语言中，算术运算符 Mod 用于返回两数相除的余数. 例如，10 Mod 6 所得值为 10 除以 6 的余数 4.

```
If(years Mod 4=0 And years Mod 100<>0)Or years Mod 400=0 Then
    Print years,"是闰年"
Else
    Print years,"不是闰年"
End If
```

练 习

1. 对一批货物征收税金：价格在 10 000 元以上的货物征税 5%；在5 000 元以上、10 000 元以下(含 10 000 元)的征税 3%；在 1 000 元以上、5 000 元以下(含 5 000 元)的货物征税 2%；1 000 元以下(含 1 000 元)的货物免税. 请设计一个算法，根据货物价格输出税金，画出算法框图.
2. 设计算法，求出方程 $ax+b=0$ 的解，画出算法框图.

2.2 变量与赋值

我们已经注意到，在设计算法的过程中，解决问题的基本思想常常很简单、很清楚，但是表述却很麻烦. 为了解决这个问题，需要引入变量与赋值. 在初中时，我们就已经了解变量的含义：在研究问题的过程中可以取不同数值的量称为变量.

变量和函数是中学数学里最重要和最基本的概念，在算法和程序设计中，它们仍然发挥着重要和基本的作用，它们会使算法的表述变得非常简洁、清楚.

下面介绍如何设置变量和给变量赋值.

例 3 设计一个算法，从 5 个不同的数中找出最大数，用框图描述这个算法.

分析理解

解决这个问题的思路很简单，先选 2 个数进行比较，去掉小的，留下大的；再取第 3 个数与留下的数进行比较，去掉小的，留下大的；继续进行，直到每个数都被比较，最后留下的数就是最大数.

解 记这 5 个不同的数分别为 a_1, a_2, a_3, a_4, a_5，算法步骤如下：

1. 比较 a_1 与 a_2，将较大的数记作 b.

(在这一步中，b 表示的是前 2 个数中的最大数.)

2. 再将 b 与 a_3 进行比较，将较大的数记作 b.

（执行完这一步后，b 的值就是前 3 个数中的最大数.）

3. 再将 b 与 a_4 进行比较，将较大的数记作 b.

（执行完这一步后，b 表示的是什么?）

4. 再将 b 与 a_5 进行比较，将较大的数记作 b.

（执行完这一步后，b 表示的是什么?）

5. 输出 b，b 的值即为所求的最大数.

以上算法步骤如图 2-8 所示.

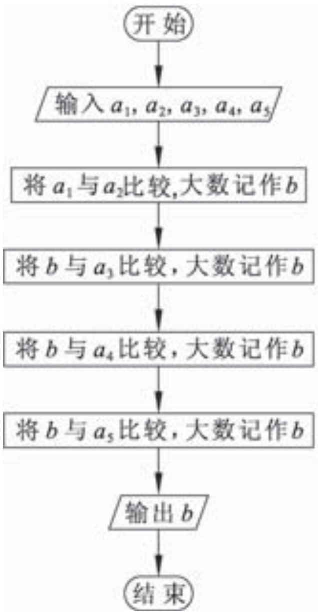


图 2-8

在上述算法的前 4 个步骤中，每步都要与上一步中得到的最大数 b 进行比较，得出新的最大数，将其也记作 b. b 可以取不同的值，通常把 b 称作变量.

在算法中，经常需要讨论变量取某些值的情况. 以下几种情况经常使用.

第一种情况，“让变量 n 取数值 3”，在算法中，可以表示为 $n=3$，其意义是把“＝”右边的值 3，赋予“＝”左边的变量 n，读作“把 3 赋值给 n”.

第二种情况，“先将 3 赋值给变量 n，再将 n 得到的数值加 1，并赋值给变量 m”，在算法中，可以表示为以下顺序结构（图 2-9）：

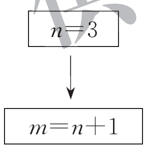


图 2-9

不难理解，首先把 3 赋值给 n，接着，又把 $3+1=4$ 赋值给 m.

第三种情况，在算法中，经常会出现以下赋值：$n=n+1$. 在这个赋值中，表示将原来 n 的数值加 1，再赋值给“＝”左边的 n.

抽象概括

变量和赋值的概念在算法中十分重要. 可以把变量想象成一个盒子，赋值就相当于往盒子里放东西. 这个盒子可以装不同的数值，但是一次只能装一个，当赋予它新值的时候，原来的值将被新值取代. 当变量参与运算和操作时，它表示的是想象中盒子里装的值.

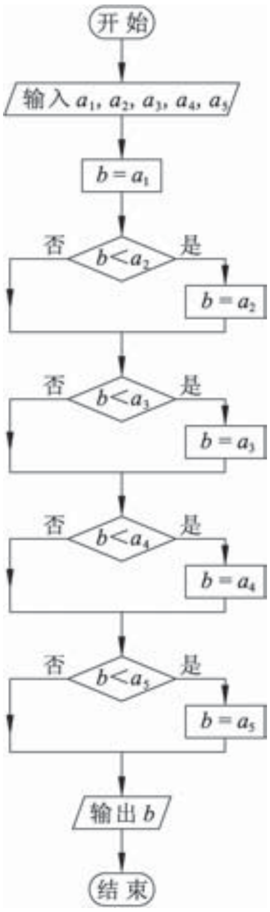


图 2-10

下面，我们用变量与赋值来表示例 3 的算法步骤：

1. $b=a_1$；
2. 比较 b 与 a_2，如果 $b<a_2$，则 $b=a_2$；
3. 比较 b 与 a_3，如果 $b<a_3$，则 $b=a_3$；
4. 比较 b 与 a_4，如果 $b<a_4$，则 $b=a_4$；
5. 比较 b 与 a_5，如果 $b<a_5$，则 $b=a_5$；
6. 输出 b，b 就是这 5 个数中的最大数.

算法的框图如图 2-10.

例 4 用赋值语句写出下列算法，并画出框图：摄氏温度 C 为 23.5 ℃，将它转换成华氏温度 F，并输出. 已知 $F=\frac{9}{5}C+32$.

分析理解

这个算法需要设置两个变量：C 和 F，分别代表输入的摄氏温度和输出的华氏温度. 算法可以描述如下：

$C=23.5$；

$F=\frac{9}{5}C+32$；

输出 F.

框图如图 2-11.

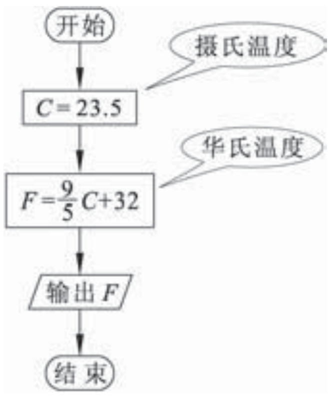
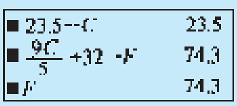


图 2-11

从以上的例子可以清楚地看到：使用了变量和赋值语句，算法的表示变得非常简洁和清晰.

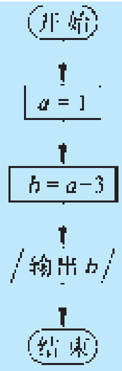
说　明

在一些计算器中，按键 $\boxed{\text{STO}}$ 可以对变量进行赋值，例如：$\boxed{\text{a}}$ $\boxed{\text{STO}}$ $\boxed{\text{b}}$ 表示 a 赋值给 b. 例 4 可以如下方式执行算法：



练 习 1

1. 输入 3 个数，设计算法找出这 3 个数中的最小数，并画出相应的框图.
2. 阅读如图所示的框图，说明输出结果.
3. 王妈妈开了一家小型餐馆，为了节约服务生收费时间，她购进红、黄、蓝、绿四种颜色的盘子，用这几种颜色的盘子分别盛 5 元、8 元、10 元和 12 元的食品，这样结账的时候，只要数一下盘子就可以了. 请利用赋值语句描述用餐记费的算法.



（第 2 题）

例 5 经过市场调查分析得知，1999 年第一季度内，某地区对某件商品的需求量为 12 000 件. 为保证商品不脱销，商家在每月月初将商品按相同数量投放市场. 已知年初商品的库存量为 50 000 件，用 S 表示商品的库存量，请设计一个算法，求出第一季度结束时商品的库存量，并画出框图.

分析理解

因为第一季度商品的需求量为 12 000 件，而且每个月以相同数量投放市场，因此每个月向市场投放 4 000 件商品. 这样，1 月的库存为年初库存量 50 000 件减去市场投放量 4 000 件，故为 46 000 件. 可以用下表表示库存量随着月份的变化情况.

表 2 2

月份	1	2	3
库存量 S/件	46 000	42 000	38 000

还可以用下列赋值语句来表示库存量的变化：

$S=S-4\ 000.$

赋值号左边的变量 S 可看作盒子，如果它表示的是这个月的存储量，那么右边的变量 S 表示的是上个月的存储量.

这是对变量 S 的赋值，赋值的目的是改变变量 S 的值，将变量 S 上次的值减 4 000 再次赋予变量 S.

解　算法框图如图 2-12：

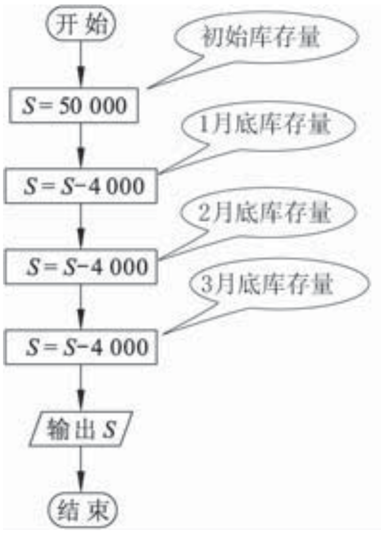


图 2-12

例 6　有关专家建议，在未来几年，中国的通货膨胀率保持在 3%左右将对中国经济的稳定有利无害. 所谓通货膨胀率为 3%，指的是每年消费品的价格增长率为 3%. 在这种情形下，某种品牌的钢琴 2004 年的价格是 10 000 元，请用框图描述这种钢琴今后 4 年的价格变化情况，并输出 4 年后钢琴的价格.

分析理解

用 P 表示钢琴的价格，不难算出：

2005 年　$P=10\ 000(1+3\%)=10\ 300$；

2006 年　$P=10\ 300(1+3\%)=10\ 609$；

2007 年　$P=10\ 609(1+3\%)=10\ 927.27$；

2008 年　$P=10\ 927.27\ (1+3\%)\approx 11\ 255.09$.

因此，价格的变化情况见表 2-3.

表 2 3

年份	2004	2005	2006	2007	2008
钢琴价格 P/元	10 000	10 300	10 609	10 927.27	11 255.09

这个变化情况可以用下列赋值语句来表示：

$P=\ P\ (1+3\%)$.

如果左边的变量 P 表示的是今年的钢琴价格，那么右边的变量 P 表示的是去年的钢琴价格.

解　算法框图如图 2-13：

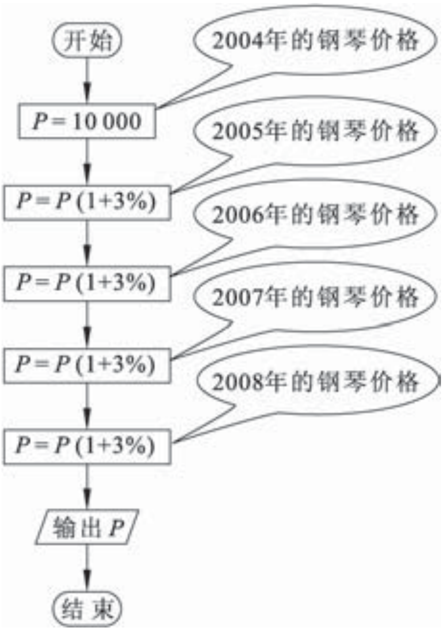


图 2-13

练 习 2

1. 一个弹球自由下落撞击地板后，它的反弹高度是原来高度的 85%. 假设初始高度为 h，请用框图说明弹球高度的变化情况，并输出反弹 4 次后的高度.
2. 一个地区的人口现为 45 647 人，并且每年的人口增长率为2.8%，请用框图表示这个地区的人口变化情况，并输出该地区 5 年后的人口数.

2.3　循环结构

前面介绍了算法的两种重要结构——顺序结构和选择结构，并学习了运用变量和赋值来描述算法. 下面介绍算法的另一种重要结构——**循环结构.**

例 7　设计算法，输出 1 000 以内能被 3 和 5 整除的所有正整数，画出算法框图.

分析理解

这个问题很简单，凡是能被 3 和 5 整除的正整数都是 15 的倍数，由于 1 000＝15×66＋10，因此 1 000 以内一共有 66 个这样的正整数.

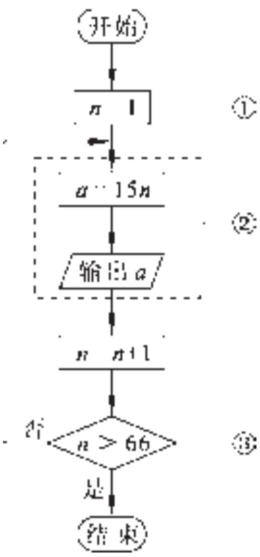


图 2-14

解　引入变量 a 表示待输出的数，则

$$a=15n\quad(n=1,2,3,\cdots,66).$$

n 从 1 变到 66，反复输出 a，就能输出 1 000 以内的所有能被 3 和 5 整除的正整数.

算法框图如图 2-14 所示.

像这样的算法结构称为循环结构，其中反复执行的第②部分称为**循环体**.

变量 n 控制着循环的开始和结束，称为**循环变量**，第①部分就是赋予循环变量初始值，预示循环开始.

第③部分判断是否继续执行循环体，称为**循环的终止条件**.

例 8　阅读图 2-15 中所示的框图，回答下列问题：

(1) 变量 y 在这个算法中的作用是什么？

(2) 这个算法的循环体是哪一部分，功能是什么？

(3) 这个算法的处理功能是什么？

解　(1) 变量 y 是循环变量，控制着循环的开始和结束；

(2) 框图中的第②部分是循环体，其功能是判断年份 y 是否是闰年，并输出结果；

(3) 由前面的分析，我们知道，这个算法的处理功能是：判断 2000～2500(包括 2500)年中，哪些年份是闰年，哪些年份不是闰年，并输出结果.

在例 7、例 8 的算法中，需要反复进行相同的操作，如果按照顺序结构来描述，算法显得十分烦琐，不利于阅读，如果采取循环结构来描述，算法就显得简洁、清楚. 循环结构是一种简化算法叙述的结构.

例 9　设计算法，求 100 个数中的最大数，画出算法框图.

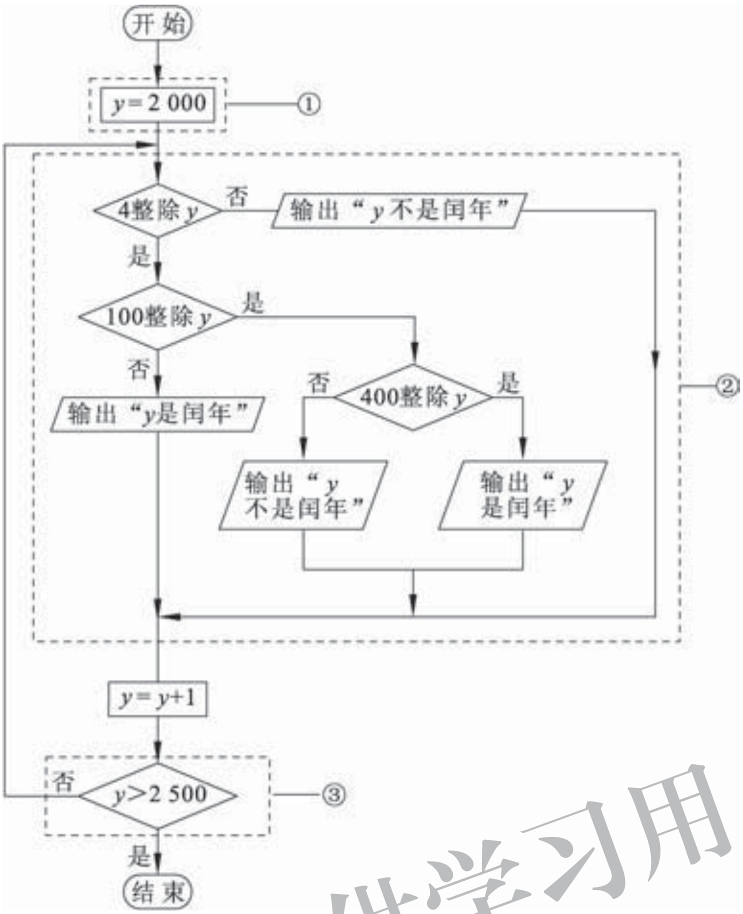


图 2-15

分析理解

在例 3 中，求 5 个数中的最大数的算法将数进行了 4 次比较，按照这种算法，求 100 个数中的最大数，就得进行 99 次比较. 虽然比较次数很多，但是操作都是一样的，因此，可以采用循环结构来描述.

引入变量 i，b，并用 $a_i(i=1,2,3,\cdots,100)$ 表示待比较的数. 赋予 b 的初始值 $b=a_1$，算法中反复执行的部分为：

比较 b 与 a_i，如果 $b<a_i$，则 $b=a_i$. 框图如图 2-16.

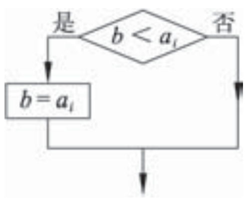


图 2-16

这就是循环体.

在循环体中，操作都是一样的，待比较的数可能有变化. 因此用变量 i 来表示待比较的数的序号，则 i 的初始值为 2，终止值为 100，它是循环变量.

这个循环体不能无限制地执行，当 $i=100$ 时，进行的是最后一次比较，因此循环的终止条件是 $i>100$.

解　算法框图如图 2-17：

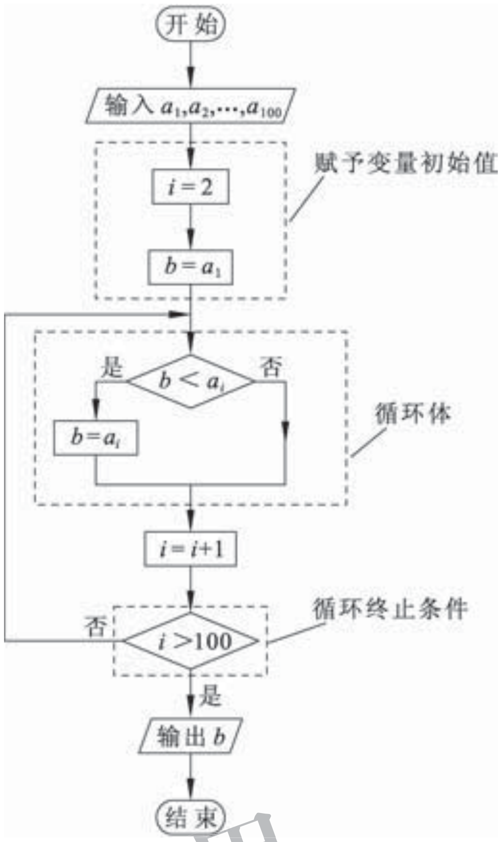


图 2-17

抽象概括

本小节介绍了如何用循环结构来描述算法，一般来说，在画出算法框图之前，需要确定三件事：

(1) 确定循环变量和初始条件；

(2) 确定算法中反复执行的部分，即循环体；

(3) 确定循环的终止条件.

这样，循环结构的算法框图的基本模式如图 2-18：

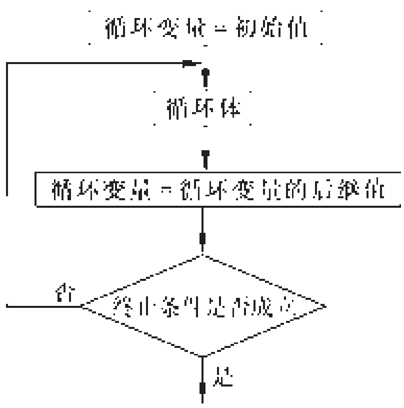
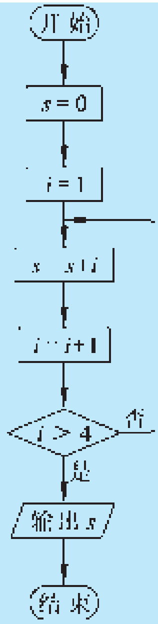


图 2-18

练 习 1

1. 设计一个算法，输出 500 以内能被 4 整除的正整数.
2. 请观察给出的框图，这是一个求和算法的框图，请运行几步看一看，指出该循环结构的循环体、循环变量和循环的终止条件.



（第 2 题）

例 10　菲波那契数列表示的是这样一列数：0，1，1，2，3，5，…，从第三项起每一项等于前两项的和. 设计一个算法框图，输出这个数列的前 50 项.

分析理解

设置 50 个变量：A_1，A_2，A_3，…，A_{50}，表示菲波那契数列的前 50 项.

如果设 A_{i-2}，A_{i-1}，A_i 分别表示数列中连续的三项，则

$$A_i=A_{i-1}+A_{i-2}.$$

由这个递推关系式知道，只要已知这个数列的前两项，就能将后面的所有项都输出来. 因为在算法中，反复计算和输出的步骤都是一样的，因此，可以用循环结构来描述这个算法，循环体如图 2-19.

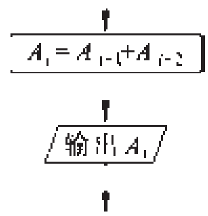


图 2-19

（1）循环变量和初始条件

设下标 i 为循环变量；

初始条件为：3 为 i 的初始值.

（2）循环体

算法中反复执行部分为：

$$A_i=A_{i-1}+A_{i-2};$$

输出 A_i.

(3) 终止条件

当 $i>50$ 时，算法结束.

解法 1　算法框图如图 2-20.

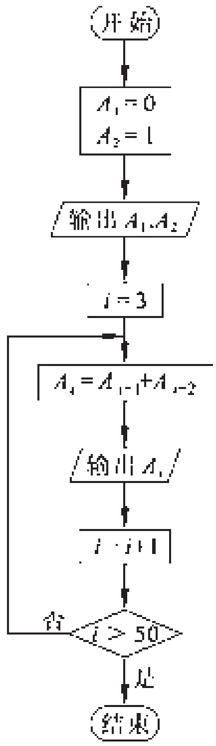


图 2-20

在上述算法中，一共设置了 50 个变量 $A_1,A_2,A_3,\cdots,A_{50}$，这些变量在完成了输出操作后，不再进行任何其他操作，因此没有保留的必要. 可以采用以下方式，除了变量 i 之外，只设 3 个变量 A,B,C，完成变量输出操作后，及时调整变量的值，节约空间.

$A=0$；

$B=1$；

输出 A,B；

$C=A+B$；

输出 C.

执行至此，已经输出了数列的前三项，可以释放掉这 3 个变量代表的数值，赋予变量新的数值：

$A=B$；（用 A 表示数列的第 2 项）

$B=C$；（用 B 表示数列的第 3 项）

$C=A+B$；

输出 C.（此时实际上输出的是数列的第 4 项）

反复这样做，就可以输出数列中的所有项.

解法 2　算法框图如图 2-21.

信息技术建议

利用计算机语言（如 BASIC 语言）可以方便地输出菲波那契数列. 具体操作步骤见本节的“信息技术应用”栏目.

信息技术应用

输出菲波那契数列的前 50 项

使用计算机语言可以很容易地计算输出菲波那契数列，下面以 BASIC 语言为例给出输出菲波那契数列前 50 项的具体程序：

```
f1=0
f2=1
Print"菲波那契数列为", f1, f2
i=3
Do
    f3=f1+f2
    Print f3
```

```
        f1=f2
        f2=f3
        i=i+1
    Loop While i≤50
```

存储空间是计算机的重要资源，在设计算法时，要尽可能减少变量，节省空间. 这是算法设计的重要原则.

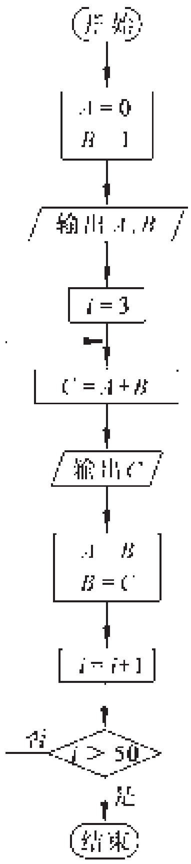


图 2-21

例 11　设区间[0,1]是方程 $f(x)=0$ 的有解区间，画出用二分法算法求方程 $f(x)=0$ 在区间[0,1]上的一个近似解的框图，要求精度为 0.01.

分析理解

我们知道，对于给定的一元方程 $f(x)=0$，要求精度为 0.01 的近似解的算法如下：

1. 确定有解区间$[a_0,b_0]$($f(a_0)\cdot f(b_0)<0$).

2. 取$[a_0,b_0]$的中点$\dfrac{a_0+b_0}{2}$.

3. 计算函数 $f(x)$在中点处的函数值 $f\left(\dfrac{a_0+b_0}{2}\right)$.

4. 判断函数值 $f\left(\dfrac{a_0+b_0}{2}\right)$是否为 0：

(1) 如果为 0，$x=\dfrac{a_0+b_0}{2}$就是方程的解，问题就得到了解决.

(2) 如果不为 0，则分下列两种情形：

① 若 $f(a_0)\cdot f\left(\dfrac{a_0+b_0}{2}\right)<0$，则确定新的有解区间为$\left(a_0,\dfrac{a_0+b_0}{2}\right)$；

② 若 $f(a_0)\cdot f\left(\dfrac{a_0+b_0}{2}\right)>0$，则确定新的有解区间为$\left(\dfrac{a_0+b_0}{2},b_0\right)$.

5. 判断新的有解区间的长度是否不大于 0.01：

(1)如果区间长度不大于 0.01，则此区间内任意值均可作为方程的近似解；

(2)如果区间长度大于 0.01，则在新的有解区间的基础上重复上述步骤.

在上述算法中：

(1) 循环变量和初始条件

设两个变量 a,b，分别表示有解区间的左端点和右端点，初始值分别为 0 和 1.

（2）循环体

算法中反复执行的部分是判断函数值 $f\left(\frac{a+b}{2}\right)$ 是否为 0：

① 如果 $f\left(\frac{a+b}{2}\right)=0$，输出 $\frac{a+b}{2}$.

② 如果 $f\left(\frac{a+b}{2}\right)$ 不为 0，则判断 $f(a)\cdot f\left(\frac{a+b}{2}\right)$ 的符号：

i　如果 $f(a)\cdot f\left(\frac{a+b}{2}\right)<0$，$b=\frac{a+b}{2}$；

ii　如果 $f(a)\cdot f\left(\frac{a+b}{2}\right)>0$，$a=\frac{a+b}{2}$.

③ 判断是否 $b-a\leqslant 0.01$.

（3）终止条件

$f\left(\frac{a+b}{2}\right)=0$ 或 $b-a\leqslant 0.01$.

解　算法框图如图 2-22.

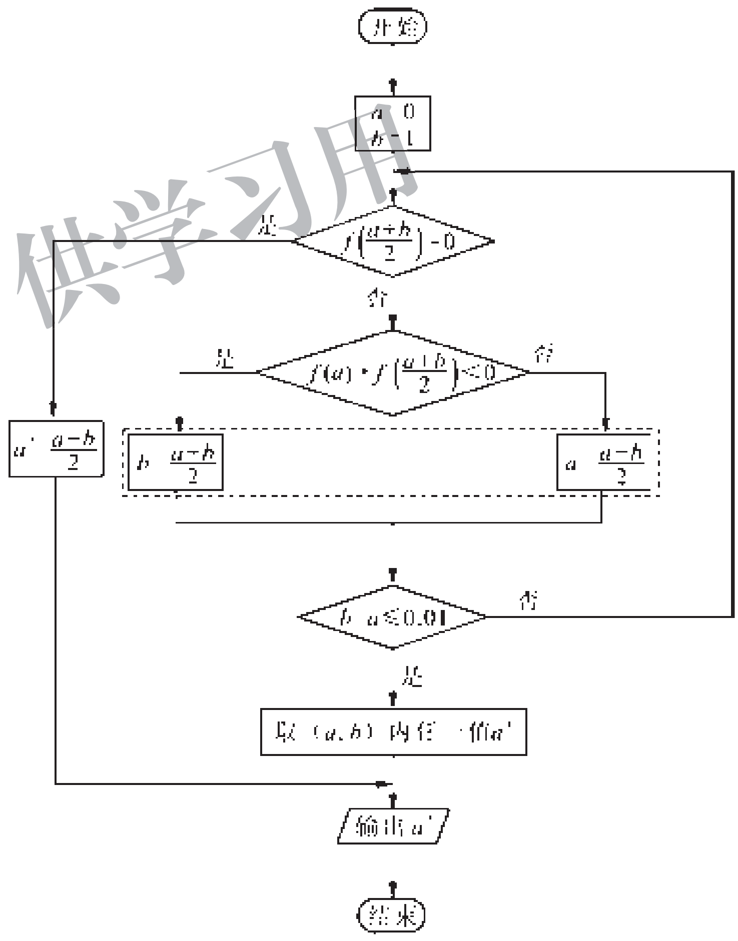


图 2-22

思考交流

上述框图中画虚线的部分在算法中的作用是什么？

从以上实例的分析中可以初步体会到：循环结构可以大大地简化算法的表述；循环变量在构造循环结构中发挥了关键作用，本质上，这就是“函数的思想”.

抽象概括

顺序结构、选择结构和循环结构是算法框图的基本结构.

顺序结构的主要特征是：完成一个步骤，再进行另一个步骤，即按顺序完成一组工作. 如图 2-5 所示.

选择结构的主要特征是：根据对条件的判断决定下一步工作. 如果条件成立，则进行步骤甲，否则，进行步骤乙. 如图 2-6 所示. 它体现了数学中分类讨论思想.

循环结构是针对重复完成一部分工作的算法设计，通常称这部分工作为循环体. 循环结构中有不变的东西，也有变的东西，不变的东西是指这部分工作的步骤，变的东西是循环变量的取值. 循环结构的作用是简化算法. 如图 2-18 所示.

练 习 2

1. 设计算法找出 100 个数中的最小数，画出算法框图.
2. 已知一列数满足后一项等于前两项的平方和：0，1，1，2，5，29，866，750 797，…请设计算法框图，输出该数列的前 20 项.

习 题 2—2

A　组

1. 请设计算法框图完成从英里到千米的换算.
2. 随着人的年龄的增加，成年人的肺活量会逐渐减少，假如我们用 V 表示人的肺活量（单位：升），

用 h 表示人的身高（单位：英寸），a 表示年龄，则这几个量近似地满足关系式：$V=0.104h-0.018a-2.69$. 请设计算法框图，输入身高、年龄，输出肺活量.

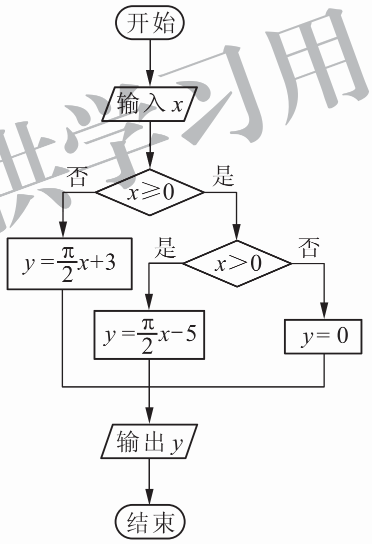
3.《中华人民共和国个人所得税法》规定，公民全月工资、薪金所得不超过 3 500 元的部分不必纳税，超过 3 500 元的部分为全月纳税所得额. 此项税款按下表分段累进计算，请用框图表示：输入工资，输出税后工资的算法.

全月应纳税所得额	税率
不超过 1 500 元的部分	3%
超过 1 500 元至 4 500 元的部分	10%
超过 4 500 元至 9 000 元的部分	20%
…	…

4. 民乐乐团筹备了一场新年音乐会，12 月 31 日晚在中山音乐礼堂演出，并对外售票，成人票 5 元，学生票 3 元. 假设有 n 个成人和 m 个学生参加新年音乐会，请设计算法框图，完成售票计费工作，要求输出最后票房收入.

5. 徐童买了一辆价值为 15 万元的汽车，汽车将以每年 20% 的速度折旧. 请用算法框图描述汽车的价值变化，并输出 5 年后汽车的价值.

6. 观察所给算法框图，说明它所表示的函数.



（第 6 题）

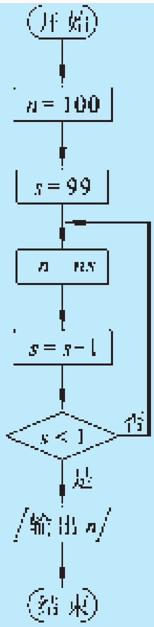
7. 函数 $f(x)=\begin{cases}-x+1, & x>0,\\ 0, & x=0,\\ x+3, & x<0.\end{cases}$ 请设计算法框图，要求输入自变量，输出函数值.

8. 高中某班一共有 40 名学生，设计算法框图，统计班上数学成绩良好（$80<$分数$\leqslant 90$）和优秀（分数>90）的学生人数.

9. 设计算法框图，输出 2 000 以内除以 3 余 1 的正整数.

B 组

1. 观察如图所示的算法框图，说明该算法的处理功能.
2. 设计算法判断一个数是否是素数，并画出算法框图.
3. 一个三位数，各位数字互不相同，十位数字比个位、百位数字之和还要大，且十位、百位数字不是素数. 设计算法，找出所有符合条件的三位数，要求画出算法框图.



（第 1 题）

阅读材料

美索不达米亚人的开方算法

汹涌湍急的底格里斯河与幼发拉底河所灌溉的美索不达米亚平原，是人类文明的发祥地之一. 美索不达米亚人长于计算，他们创造了优良的计数系统. 美索不达米亚的学者在发展程序化算法方面表现出了熟练技巧，他们创造了许多成熟的算法，开方计算中有一个例子——求正数平方根近似的算法是最具代表性的. 他们设计的算法是这样的：

1. 确定平方根的首次近似值 $a_1\{a$ 可以任取一个正数$\}$；

2. 由代数式 $b_1=\dfrac{a}{a_1}$ 求出 b_1；

3. 取二者的算术平均数 $a_2=\dfrac{a_1+b_1}{2}$ 为第二次近似值；

4. 由方程 $b_2=\dfrac{a}{a_2}$ 求出 b_2；

5. 取算术平均数 $a_3=\dfrac{a_2+b_2}{2}$ 作为第三次近似值；

……

反复进行上述步骤，直到获得满足精度的近似值.

下面来看看这个算法的原理.

设 $x=a^{\frac{1}{2}}$ 表示所求的平方根，并设 a_1 是这个根的首次近似值. 由方程 $b_1=\dfrac{a}{a_1}$ 求出 b_1，若 $a_1^2<a$，则 $b_1^2>a$，反之亦然. 接着，再取二者的算术平均数 $a_2=\dfrac{a_1+b_1}{2}$，则这个近似值更接近所求的平方根.

耶鲁大学收藏的一块古巴比伦泥版（编号 7289），其上载有 $2^{\frac{1}{2}}$ 的近似值，结果精确到六十进制的三位小数，用十进制写出来是1.414 213，这个结果是相当精确的.

你能用循环结构来描述这个算法，画出相应的框图吗？

资料来源：李文林. 数学史概论. 第二版. 北京：高等教育出版社，2002

§3　几种基本语句

我们已经学习了用自然语言和框图来描述算法，要使算法在计算机上实现，还得借助程序语言. 程序语言的种类很多，但是，有一些基本语句是所有语言都要使用的，例如，输入语句、输出语句、赋值语句、条件语句和循环语句. 这些语句在程序语言中是最重要的和最基本的. 输入语句、输出语句很好理解，赋值语句前面已作了介绍，这一节以 BASIC 语言为例主要介绍条件语句和循环语句.

3.1　条件语句

在算法中，选择结构是一种基本结构，条件语句是表达选择结构最常用的语句. 下面通过实例说明条件语句的用法.

例 1　设计算法，根据输入 x 的值，计算 y 的值.

$$y=\begin{cases}x^2+1, & x\leqslant 2.5;\\ x^2-1, & x>2.5.\end{cases}$$

分析理解

这是一个分段函数的求值问题，计算之前，应对自变量进行判断，可以表述为：

(1) 输入 x；

(2) 如果 $x\leqslant 2.5$，则 $y=x^2+1$；

(3) 如果 $x>2.5$，则 $y=x^2-1$；

(4)输出 y.

解　可以用条件语句表示第(2)，(3)步骤：

```
If  x<=2.5    Then
    y=x^2+1
Else
    y=x^2-1
End If
```

这种表示具有一般性，实际上，对于形如框图 2-23 的算法，

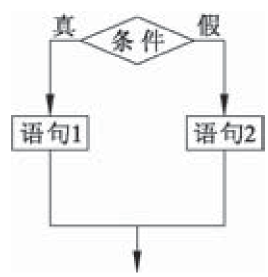


图 2-23

都可以用下列语句来表示：

If　条件　Then

　　语句 1

Else

　　语句 2

End If

例 2　在音乐唱片超市里，每张唱片售价 25 元. 顾客如果购买 5 张以上（含 5 张）唱片，则按照九折收费；如果顾客购买 10 张以上（含 10 张）唱片，则按照八五折收费. 请用语句描述完成计费工作的算法，画出算法框图.

解　假如用变量 a 表示顾客购买的唱片数，用变量 C 表示顾客要缴纳的金额. 则这个算法可以表示为：

(1) 输入 a.

(2) 对 a 进行判断：

　① 若 $a<5$，则 $C=25a$；

　② 若 $5\leqslant a<10$，则 $C=22.5a$；

　③ 若 $a\geqslant 10$，则 $C=21.25a$.

(3) 输出 C.

算法框图如图 2-24 所示：

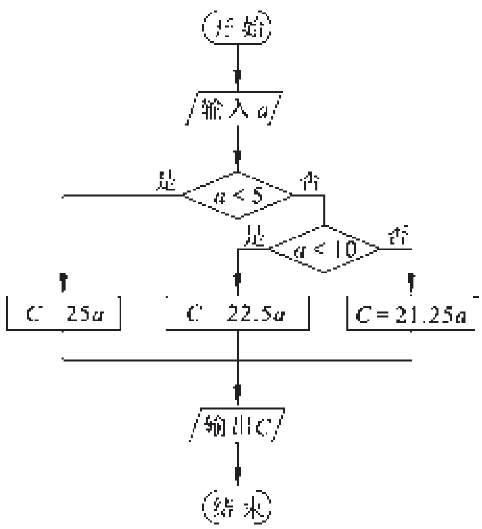


图 2-24

在这个算法中，判断的后面接着判断，我们可用复合 If 语句来描述这个算法：

输入 a；

If　$a<5$ Then

　　$C=25*a$

```
Else
    If  a<10 Then
        C=22.5*a
    Else
        C=21.25*a
    End If
End If
输出 C.
```

实际上，对于形如图 2-25 所示框图描述的算法，都可以用复合 If 语句来表示：

```
If  条件 1  Then
    语句 1
Else
    If  条件 2  Then
        语句 2
    Else
        语句 3
    End If
End If
```

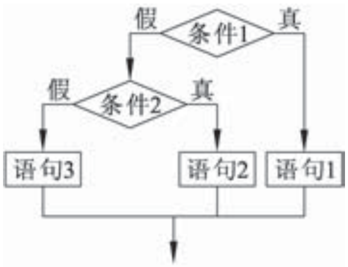


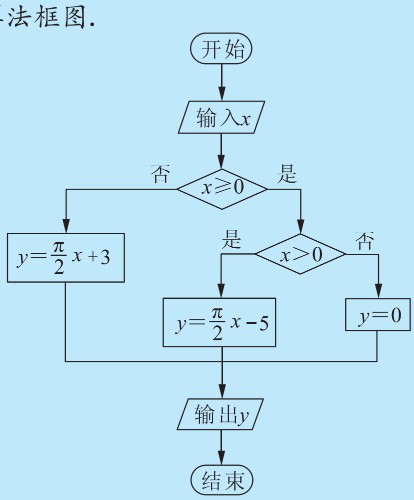
图 2-25

思考交流

请试着用复合 If 语句表示闰年问题的算法.

练 习

1. 请用复合 If 语句描述下面的算法框图.



（第 1 题）

2. 农历 9 月 9 日是中国的重阳节，中华饭店自助餐厅决定在这一天进行优惠酬宾活动. 对于 80 岁以上的老人，享受免费自助餐；70 岁以上的老人享受 5 折优惠；60 岁以上的老人享受 6 折优惠，其余的嘉宾享受 9 折优惠. 请设计算法，完成这一天的计费工作，要求输入用餐者的人数、年龄、消费额，输出应付金额. 运用复合 If 语句描述该算法，并画出算法框图.

3.2 循环语句

说　明

在 For 循环语句里，我们默认循环变量每次的增量为 1. 如果需要考虑增量不为 1 的情况，需要使用参数 Step.

例如，“输出 1 到 100 内的所有奇数”的算法语句为：

```
For i = 1 To 100 Step 2
    输出 i
Next
```

循环结构是算法中的基本结构，For 语句是表达循环结构最常见的语句之一，它适用于预先知道循环次数的循环结构.

For 语句的一般形式是：

```
For 循环变量=初始值 To 终值
    循环体
Next
```

下面通过实例来说明 For 语句的使用方法.

例 3　结合图 2-20 的框图，使用 For 语句描述输出菲波那契数列的前 50 项的算法.

解

```
f1=0
f2=1
输出"菲波那契数列为", f1, f2
For i=3  To 50
    f3=f1+f2
    输出 f3
    f1=f2
    f2=f3
Next
```

例 4　请阅读下列用 For 语句写出的算法，说明该算法的处理功能.

```
S=0
T=1
For i=1 To 20
    S=S+i
    T=T*i
Next
输出 S
```

输出 T

分析理解

由 For 语句的形式，我们知道 i 是循环变量，初始值为 1，终止值为 20，循环体为：

$S=S+i$；

$T=T*i$.

解　算法的框图如图 2-26.

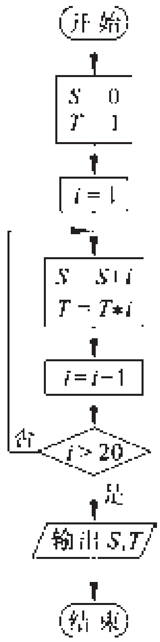


图 2-26

因此，这个算法实际上处理的是求和

$$S=1+2+3+\cdots+20$$

及求积

$$T=1\times2\times3\times\cdots\times20$$

这两件事情.

在一些循环结构中，预先不知道循环的次数，比如使用二分法求方程的近似解，要根据其他形式的终止条件停止循环，在这种情况下，一般用 Do Loop 语句来描述.

Do Loop 语句的一般形式为：

Do

　　循环体

Loop While　条件为真

这节课学习了使用基本语句描述算法，和自然语言相比，基本语句描述的算法显得直观、清楚、明了，它的另一个优点是比较接近程序语言，便于编写程序，上机实现.

练　习

1. 用 Do Loop 语句描述判断一个数是否为素数的算法.
2. 高一(2)班一共有 40 名学生，每次数学考试结束，班主任总是要统计成绩为 85～100 分，60～85 分和 60 分以下的各分数段的人数，请你帮助班主任设计一个解决该问题的算法，并用语句描述.

习 题 2—3

A 组

1. 已知地球半径为 6 371 km，请用相应基本语句写出求地球表面积和体积的算法，并用框图表示.
2. 输入三个数，用基本语句写出计算它们的平均值和三个数的乘积的算法.
3. 用基本语句写一个算法，要求输入 20 个数，输出其中正数、负数和零的个数.
4. 为了节约用水，学校改革澡堂收费制度，开始实行洗澡计时收费，洗澡时间在 30 分以内(含 30 分)，每分收费 0.1 元，洗澡时间在 30 分以上，超出 30 分的部分每分收费 0.2 元. 请设计算法，使用基本语句完成澡堂计费工作，要求输入时间，输出洗澡需交费用.
5. 使用基本语句写一个算法，要求输入三角形的三条边长，输出三角形的面积.（若用 a，b，c 表示三角形的三条边长，则三角形的面积 $S=\sqrt{l(l-a)(l-b)(l-c)}$，其中 $l=\frac{1}{2}(a+b+c)$.）
6. 用基本语句写一个算法，要求输出 1～900 中既能被 3 整除又除以 5 余 1 的所有整数.
7. 使用基本语句写出一元二次方程求解的算法.
8. 使用基本语句写出确定线段 100 等分点的算法.

B 组

1. 设计一个算法，判断给出的自然数是否为素数，并使用相应基本语句加以描述.
2. 用基本语句写一个算法，输出 100～999 中的所有水仙花数.（若三位数$\overline{ABC}=A^3+B^3+C^3$，则称$\overline{ABC}$为水仙花数. 例如 $1^3+5^3+3^3=1+125+27=153$，则 153 为水仙花数）

阅读材料

算 法 的 复 杂 性

从前面的知识知道，解决同一问题可以有多种算法，有优有劣，差异很大. 好的算法可以把解决问题的时间由几天缩短到几秒，因而，算法分析成为计算机科学的一个基本研究方向. 如何比较算法的效率呢？

假设 M 是一种算法，n 表示输入数据的规模. M 所占用的时间和空间是衡量该算法效率的两个主要指标. 一般情况下，我们主要考虑算法的运行时间.

算法复杂性是一个函数 $f(n)$，函数值是算法 M 运行所需要的时间，n 表示输入数据的规模. 如果 $f(n)$是 n 的多项式函数，则称算法是多项式算法；如果 $f(n)$是 n 的指数函数，则称算法是指数算法，目前，这种算法在计算机上是无法实现的. 显然多项式算法要好得多. 改进算法，减少算法运行时间，是算法分析的基本工作.

课题学习

确定线段 n 等分点的算法

我们在 §2 中学习了利用尺规作图确定已知线段的一个 5 等分点的方法，在具体作法中，我们先在射线 AP 上选取固定长度 AC 作为一个单位（标尺），再通过这个单位 AC（标尺）依次作出四段等长的线段 CE，EF，FG，GD，连接 BD，过点 C 作 BD 的平行线，交线段 AB 于点 M，从而点 M 就是线段 AB 的一个 5 等分点.（如图 1 所示）

能不能将上面的作法简化呢？仔细观察图 2，我们在选取固定长度 AC 作为一个单位（标尺）后，可以根据它作出等长线段 CE，现在我们就得到 2 个单位长的线段 AE，从而可以根据 AE 作出等长线段 EF，再作与 AC 等长的线段 FD，连接 BD，过点 C 作 BD 的平行线，交线段 AB 于点 M，从而点 M 就是线段 AB 的一个 5 等分点.

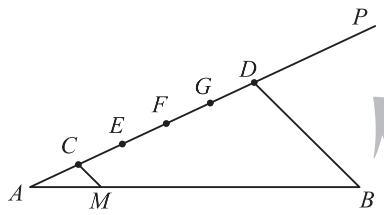


图 1

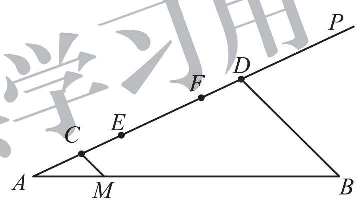


图 2

1. 比较上面两种作法，发现第二种作法少作了一个点，如果需要确定已知线段的一个 12 等分点、24 等分点，你怎么作呢？

2. 如果确定线段的一个 100 等分点呢？

*3. 若要确定 $n(n\in \mathbf{N}_+, n>2)$ 等分点，需要在射线 AP 上作多少个点？如何操作？这种方法是否能够推广，有无其他应用价值？

如果我们将十进制的数 n 用二进制表示：

$$n=a_0 2^0+a_1 2^1+a_2 2^2+\cdots+a_k 2^k,$$

其中 $a_0, a_1, \cdots, a_k$ 为 0 或 1，令 $m=k+a_0+a_1+a_2+\cdots+a_k$.

可以验证，若要确定 n 等分点，只要在射线 AP 上作 m 个点就足够了.

◆ 本章小结

一、内容提要

本章的学习内容主要由算法初步、算法的基本结构和几种基本语句三个部分组成.

1. 对“算法初步”的内容，希望同学们根据本章提供的算法实例，或再选取一些熟知的算法实例，从以下两个方面加深理解：

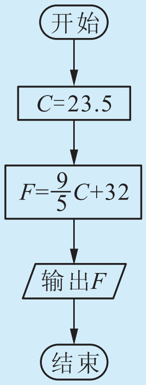
（1）形成算法的初步印象，体会算法是问题解决的“机械”程序，并能在有限步内获得问题的解决；

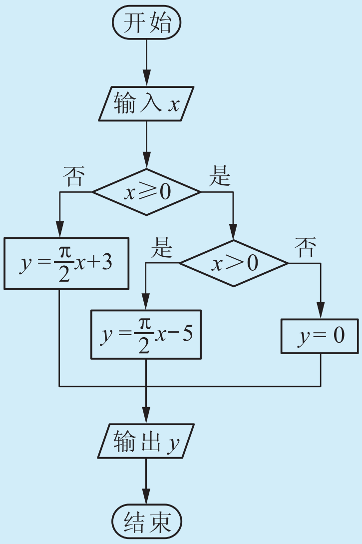
（2）感受算法学习的必要性，并初步认识二分法. 通过这三个课时的学习力图把握算法的基本思想.

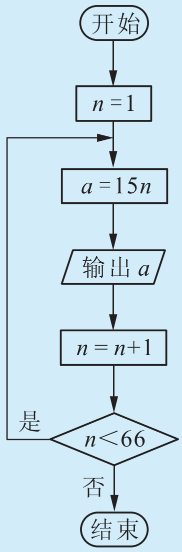
2. “算法的基本结构”的内容，首先介绍了顺序结构和选择结构，学习了用框图来描述算法，通过具体的问题体会算法的这种描述方式的优点；接着介绍了算法中的重要概念——变量，以及如何给变量赋值，学习将常数赋予变量、将含其他变量的表达式赋予变量、将含有变量自身的表达式赋予变量，理解这些赋值方式的意义，切实学会通过赋值的方式改变变量的值.

3. “几种基本语句”的内容，介绍了三种基本语句：赋值语句、条件语句和循环语句，其中条件语句介绍了两种：If 语句和复合 If 语句，循环语句介绍了两种：For 语句和 Do Loop 语句，运用这些语句描述了本章前面涉及的一些算法.

在完成这一章内容的学习后，建议同学们先回顾本章的内容，梳理知识结构，再相互讨论，并共同建立以下的知识框架图.

<table>
<tr><td rowspan="3">本章基本结构</td><td>算法思想</td><td colspan="3">一些问题的解决常常需要设计出一系列可操作的步骤，只要按顺序执行这些步骤，都能完成任务，通常把这种解决问题的思想称为程序化思想或者算法思想.</td></tr>
<tr><td>算法的基本结构</td><td>

顺序结构</td><td>

选择结构</td><td>

循环结构</td></tr>
<tr><td>语句</td><td></td><td>If 语句</td><td>For 语句
Do Loop 语句</td></tr>
</table>

二、学习要求和需要注意的问题

1. 学习要求

(1) 结合熟悉的算法，把握算法的基本思想，学会用自然语言来描述算法.

(2) 通过模仿、操作、探索，经历设计程序框图表达解决问题的过程. 在具体问题的解决过程中(如：三元一次方程组求解等问题)，理解程序框图的三种基本逻辑结构：顺序结构、选择结构、循环结构.

(3) 通过实际问题的学习，了解构造算法的基本程序.

(4) 经历将具体问题的程序框图转化为程序语句的过程，理解几种基本算法语句——赋值语句、条件语句、循环语句，体会算法的基本思想.

2. 需要注意的问题

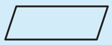
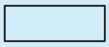
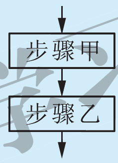
(1) 从熟知的问题出发，体会算法的程序化思想，而不是简单呈现一些算法.

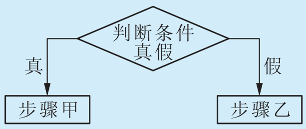
(2) 变量和赋值是算法学习的重点之一，因为设置恰当的变量，学习给变量赋值，是构造算法的关键，应作为学习的重点.

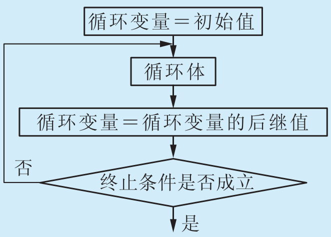
(3) 不刻意追求最优的算法，把握算法的基本结构和程序化思想才是我们的重点.

(4) 本章所指的算法基本上是能在计算机上实现的算法.

算法框图与基本语句

	框　图	基本语句
开始		
输入 输出		输入(Input) 输出(Output)
处理		赋值 $x=a$，计算
顺序 结构		步骤甲 步骤乙

选择
结构
		If　条件　Then
		　步骤甲
		Else
		　步骤乙
		End If

循环 结构		For 语句 　For 循环变量＝初始值 To 终值 　　循环体 　Next Do Loop 语句 　Do 　　循环体 　Loop While 条件为真
结束		

复习题二

A　组

1. 设计算法框图，求解方程 $x^3+4x-10=0$ 在区间[0，2]内的解(精度为 10^{-5}).
2. 设计一个算法框图，求函数 $f(x)=2^x$ 在 $x=\frac{1}{2}$ 的近似值(精度为 0.01).
3. 铁路托运行李规定：行李重不超过 50 千克的，托运费按每千克0.15 元计费；如超过 50 千克，超过部分每千克加收 0.10 元. 设计算法框图完成自动计费工作，要求输入行李质量，输出费用.
4. 某超市为了促销，规定：一次性购物 50 元以下(含 50 元)的，按原价付款；超过 50 元但在 100 元以下(含 100 元)的，超过部分按九折付款；超过 100 元的，超过部分按八折付款. 设计算法框图，完成超市的自动计费工作，要求输入购物金额，输出应付款.
5. 设计算法，要求输入两个正整数，输出它们的最大公因数和最小公倍数，画出算法框图，并用基本语句描述该算法.
6. 设计输出 x，$\sin x$，$\cos x$，$\tan x$ 的值的算法框图，其中 $x=0°$，$10°$，$20°$，$30°$，…，$180°$.
7. 为了改善办学条件，学校于今年购买了 75 台新型图形计算器，还计划从明年开始每年追加 25 台. 请用算法框图表示学校拥有新型图形计算器总量的变化趋势，并输出 5 年后拥有的新型图形计算器总量.

B　组

辗转相除法是求两个数最大公因数的一种算法，请查阅相关书籍了解这种算法，说明它的算理，画出这个算法框图，并用语句来描述这个算法.

C　组

1. 《张邱建算经》[1]卷下最后一题通常称为“百鸡问题”：“今有鸡翁一，值钱五；鸡母一，值钱三；鸡雏三，值钱一. 凡百钱买鸡百只. 问鸡翁、母、雏各几何?”通俗地说就是：一只公鸡 5 元，一只母鸡 3 元，三只小鸡 1 元，如果花 100 元钱买 100 只鸡，公鸡、母鸡、小鸡各几只? 写出该问题的算法，并画出算法框图.

   [1]《张邱建算经》共三卷，据考证大约成书于公元 466—485 年，作者张邱建是北魏人.

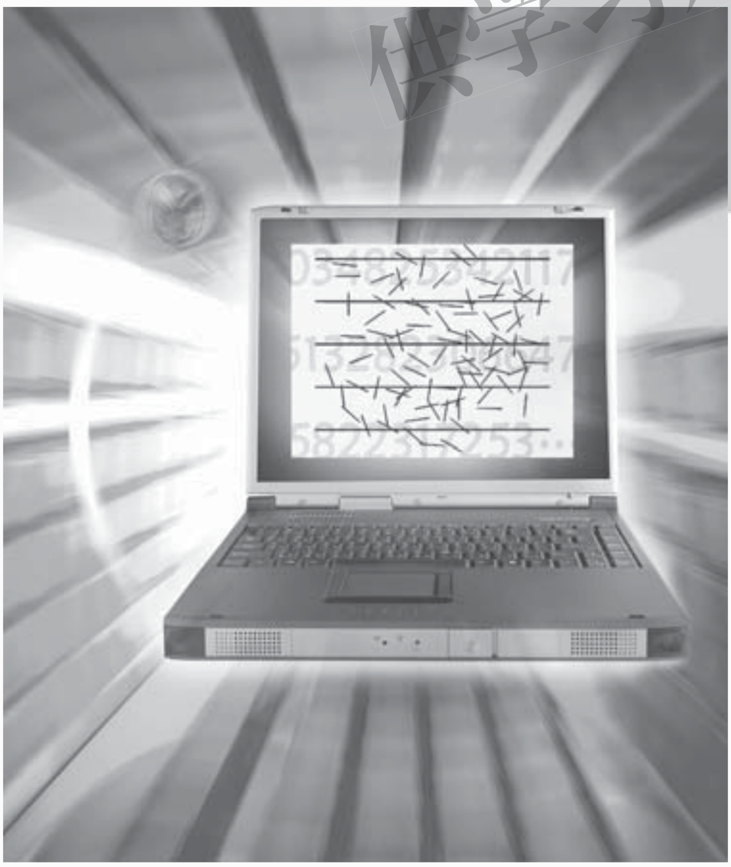
2. 设计算法框图计算 $10!+7!\times8!$，其中 $10!=10\times9\times8\times7\times6\times5\times4\times3\times2\times1$，$7!=7\times6\times5\times4\times3\times2\times1$，$8!=8\times7\times6\times5\times4\times3\times2\times1$.

第三章　概　率

我们生活在一个充满机会和风险的世界里，比如彩票中奖、天气预报、投资风险等. 如何把握机会，减少风险？解决这些问题需要学习和掌握概率知识.

概率论起源于对赌博中一些问题的研究，现在已广泛地应用于自然科学、社会科学和人们的日常生活之中，是一个非常有生命力的数学分支.

本章将进一步学习概率的概念，介绍一些基本的概率模型，并通过讨论实例，加深对随机现象及其规律的理解.



π 这个数渗透了整个数学.

你能想象这样也可以计算出 π 的近似值吗？向等距的平行线上投针，针长为线距的一半，投针数为 n，与线相交的针数为 m，则$\pi \approx n/m$.

§1　随机事件的概率

1.1　频率与概率

问题提出

在初中阶段，我们已经了解了必然事件、不可能事件和随机事件等概率论中的一些基本概念.

随机事件是具有不确定性的事件，但是并不是具有不确定性的事件都是随机事件. 例如“谁能成为下一届联合国的秘书长”，这是不确定性事件，但它并不是随机事件，因为只有可以进行重复试验的不确定性事件才是随机事件.

生活中有很多随机事件的例子，如“明天本地下雨”“买一张彩票中奖”等. 随机事件在一次试验中是否发生具有不确定性，但是在相同条件下的大量重复试验中，它发生的频率是否会呈现出一定的规律性呢?

为了研究这个问题，2003 年北京市某学校高一(5)班的学生做了如下试验：

在相同条件下大量重复掷一枚图钉，观察出现“钉尖朝上”的频率的变化情况(如图 3-1).

图 3-1

(1) 每人手捏一枚图钉的钉尖，钉帽在下，从 1.2 米的高度让图钉自由下落.

(2) 重复 20 次，记录下“钉尖朝上”出现的次数.

图 3-2 是汇总六位同学的数据后画出来的频率图.

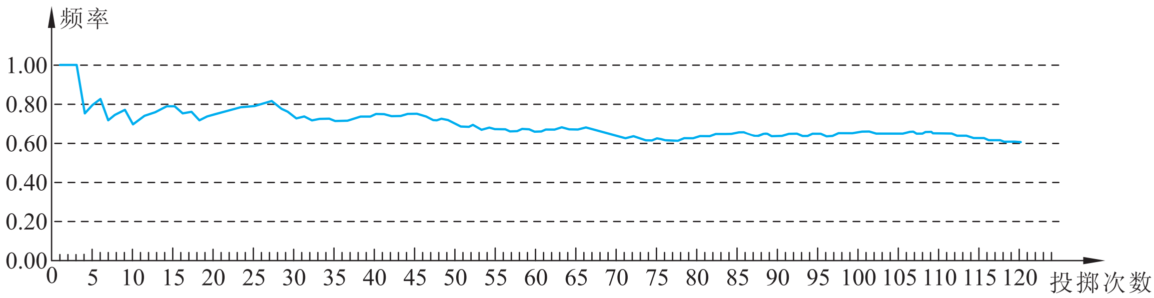


图 3-2

观察图 3-2，出现“钉尖朝上”的频率有什么样的变化趋势?

说　明

在 n 次重复试验中，“钉尖朝上”出现的次数与 n 的比值称为这 n 次试验中“钉尖朝上”的频率.

动手实践

从一定高度按相同的方式让一枚图钉自由下落，图钉落地后可能钉尖朝上，也可能钉尖着地. 大量重复试验时，观察出现“钉尖朝上”的频率的变化情况.

(1) 从一定高度让一枚图钉自由下落并观察图钉落地后的情况，每人重复 20 次，记录下“钉尖朝上”出现的次数.

(2) 汇总每个人所得的数据，并将每个人的数据进行编号，分别得出前 20 次、前 40 次、前 60 次……试验出现“钉尖朝上”的频率.

(3) 在直角坐标系中，横轴表示掷图钉的次数，纵轴表示以上试验得到的频率，将上面算出的结果表示在坐标系中.

(4) 从图上观察出现“钉尖朝上”的频率的变化趋势，你会得出什么结论？

通过上面的试验，我们可以看出：出现“钉尖朝上”的频率是一个变化的量，但是在大量重复试验时，它又具有“稳定性”——在一个“常数”附近摆动.

思考交流

在上面掷图钉的活动中，随着试验次数的增加，出现“钉尖朝上”的频率在这个“常数”附近的摆动幅度是否一定越来越小？

抽象概括

(1) 在大量重复试验的情况下，出现“钉尖朝上”的频率会呈现出稳定性，即频率在一个“常数”附近摆动. 随着试验次数的增加，摆动的幅度具有越来越小的趋势.

(2) 有时候试验也可能出现频率偏离“常数”较大的情形，但是随着试验次数的增大，频率偏离“常数”的可能性会减小.

阅读理解

历史上曾有人做过掷硬币的试验，试验结果如表 3-1.

表 3-1

实验者	抛掷次数 n	正面朝上的次数 m	频率 m/n
德·摩根	2 048	1 061	0.518 1

续表

实验者	抛掷次数 n	正面朝上的次数 m	频率 m/n
蒲丰	4 040	2 048	0.506 9
费勒	10 000	4 979	0.497 9
皮尔逊	24 000	12 012	0.500 5
罗曼诺夫斯基	80 640	40 173	0.498 2

重复抛掷硬币，出现“正面朝上”的频率是事先无法确定的. 但是在大量重复抛掷硬币时，出现“正面朝上”的频率具有稳定性——它在 0.5 附近摆动.

又如，考察新生婴儿的性别：可能是男孩，也可能是女孩. 对大量新生婴儿的统计显示，出现“新生婴儿是男孩”的频率具有稳定性.

著名数学家拉普拉斯对男婴和女婴的出生规律作了详细的研究，他对伦敦、彼得堡、柏林和法国的情形进行了分析，得到了庞大的统计资料. 这些统计资料显示，10 年间，男孩出生的频率在$\frac{22}{43}$附近摆动.

表 3-2 是 20 世纪波兰的一些统计结果[1].

表 3 2

出生年份	总出生数 n	男孩数 m	频率 m/n
1927	958 733	496 544	0.518
1928	990 993	513 654	0.518
1929	994 101	514 765	0.518
1930	1 022 811	528 072	0.516
1931	964 573	496 986	0.515
1932	934 663	482 431	0.516
总计	5 865 874	3 032 452	0.517

表 3-3 是我国历次人口普查总人口性别构成情况[2]，它们与拉普拉斯得到的结果非常地接近.

表 3 3　我国历次人口普查总人口性别构成　　(单位:万人)

普查年份	总人口	男	女	性别比(以女性为 100)
1953	59 435	30 799	28 636	107.56
1964	69 458	35 652	33 806	105.46
1982	100 818	51 944	48 874	106.30
1990	113 368	58 495	54 873	106.60
2000	126 583	65 355	61 228	106.74

❶ 引自:[波兰]M. 费史著. 概率论及数理统计. 王福保译. 上海:上海科学技术出版社,1962

❷资料来源:国家统计局的统计数据.

动手实践

在前面的学习中，我们已经了解了随机数表. 下面我们用随机数表来模拟掷硬币的试验.

用 0，1，…，9 这 10 个数字中的任意 5 个表示"正面朝上"，其余 5 个表示"反面朝上"，每产生一个随机数就完成一次模拟.

例如，可用 0，1，2，3，4 表示"正面朝上"，用 5，6，7，8，9 表示"反面朝上". 具体过程如下：

(1) 制作一个如下形式的表格，在随机数表中随机选择一个开始点，完成 100 次模拟，并将结果记录在表 3-4 中.

表 3 4

试验次数	产生的随机数	对应的正反面情况
1		
2		
3		
4		
⋮	⋮	⋮
100		

(2) 根据表 3-4 的记录，得出 100 次模拟试验中出现"正面朝上"的频率.

(3) 汇总全班同学的结果，给出出现"正面朝上"的频率.

信息技术建议

如果你有科学计算器、图形计算器或计算机，你还可以用它们产生随机数来模拟掷硬币的试验.

根据上面的模拟结果，我们可以看出：出现"正面朝上"的频率是一个变化的量，但是当试验次数比较大时，出现"正面朝上"的频率在 0.5 附近摆动. 这与历史上大量抛掷硬币的试验结果是一致的.

抽象概括

在相同的条件下，大量重复进行同一试验时，随机事件 A 发生的频率会在某个常数附近摆动，即随机事件 A 发生的频率具有稳定性. 这时，我们把这个常数叫作随机事件 A 的**概率**，记作 $P(A)$. 我们有 $0 \leqslant P(A) \leqslant 1$.

频率反映了一个随机事件出现的频繁程度，但频率是随机的，而概率是一个确定的值，因此，人们用概率来反映随机事件发生的可能性的大小. 在实际问题中，某些随机事件的概率往往难以确切得到，因此，我们常常通过做大量的重复试验，用随机事件发生的频率作为它的概率的估计值.

练　习

1. 请举出身边的一些随机事件的例子.
2. 在上面掷图钉的活动中,根据已有的数据,计算出现“钉尖朝上”的概率大约是多少?
3. 课后调查:气象台常常用概率的语言来刻画未来天气的变化情况,比如“今天的降水概率是 60%”.你对这句话是如何理解的?对你身边的人进行调查,看看他们是如何理解的.

小资料

随机与保险

古语“天有不测风云,人有旦夕祸福”说明了生活中的风险具有不确定性.正是由于风险的存在,保险业才应运而生.那么保险业是如何具体运作的呢?

既然风险具有不确定性,如果保险期内风险频繁发生(比如投保意外伤害险的人在保险期内频繁受伤),那么保险公司不就亏本了吗?不必担心!我们知道,随机事件会呈现出一定的规律性.虽然单个人遭受意外伤害具有不确定性,但是,如果考察大量的人,人们遭受意外伤害的频率将具有一定的规律性,即遭受意外伤害的频率具有稳定性.根据统计结果,保险公司可以制定保险费与损失赔款的额度.从长远来看,投保的人越多,保险公司的实际赔付就会越接近预期结果——保险公司收取的保险费多于实际赔付.

1.2　生活中的概率

思考交流

对于 1.1 节练习的第 3 题的调查,请同学们交流各自的调查结果,发表自己的看法,并讨论下面的问题.

(1) 如何理解“今天北京的降水概率是 60%,上海的降水概率是 70%”?有没有可能“北京今天降雨了,而上海没有降雨”?请从概率的角度作出解释.

(2) 据报道:我国 1998 年的洪水是“百年一遇”的大洪水.在这里,“百年一遇”是什么意思?

问题 1　抛掷 10 次硬币,是否一定是 5 次“正面朝上”和 5 次“反面朝上”?

信息技术建议

利用计算机电子表格软件(如 Excel)中的随机函数可以对这个概率试验进行模拟,具体的操作步骤参见本节的“信息技术应用”栏目.

动手实践

全班同学每人抛掷 10 次硬币,统计出现 5 次“正面朝上”的人数,估计出现 5 次“正面朝上”的概率,这个概率大吗?

我们可以利用随机数表来模拟抛掷 10 次硬币的过程. 在表中随机选择一个开始点,用 0,1,2,3,4 表示“正面朝上”,用 5,6,7,8,9 表示“反面朝上”,产生 10 个随机数就完成 1 次模拟.

表 3-5 是某学生利用随机数表完成 10 次模拟的结果.

表 3 5

产生的随机数	对应的正反面情况
9954377560	反反反正正反反反反正
6319193767	反正正反正反正反反反
1145550193	正正正反反反正正反正
7825057919	反反正反正反反反正反
7563233716	反反反正正正正反正反
2354828717	正正反正反正反反正反
9977235699	反反反反正正反反反反
6979008925	反反反反正正反反正反
5516635468	反反正反反正反正反反
4409497745	正正正反正反反反正反

在这 10 次试验中,有 3 次试验恰出现 5 次“正面朝上”.

请完成 20 次这样的模拟,记录下每次模拟的结果. 由模拟得到的数据,估计出现 5 次 “正面朝上”的概率. 汇总班上同学的数据,重新估计出现 5 次 “正面朝上”的概率,与前面的估计相比,哪个更可信(理论上可以算得这个事件的概率约为 0.246)?

思考交流

掷一枚硬币,出现“正面朝上”的概率为$\frac{1}{2}$,是指一枚硬币掷 2 次恰出现 1 次“正面朝上”吗? 如果不是,应如何理解?

练　习　1

掷两枚均匀的硬币，会出现“两枚硬币都是正面朝上”“恰好一枚硬币正面朝上”“两枚硬币都是正面朝下”这三种结果，出现这三种结果的可能性相同吗？

（1）重复试验 50 次，记录结果，并根据试验的结果估计上述事件发生的概率.

（2）汇总你与班上其他同学的数据，得到至少 500 次试验的结果，用这些结果对上述事件的概率重新进行估计.

（3）为了对上述事件的概率给出比较好的估计，你需要怎么做？

（4）根据试验结果，你认为出现“两枚硬币都是正面朝上”“恰好一枚硬币正面朝上”“两枚硬币都是正面朝下”的概率相同吗？

问题 2　有四个阄，其中两个分别代表两件奖品，四个人按顺序依次抓阄来决定这两件奖品的归属. 先抓的人中奖率一定大吗？

为此，2003 年北京市某学校高一(5)班的学生做了如下的模拟活动：

口袋里装有 2 个白球和 2 个黑球，这 4 个球除颜色外完全相同，白球代表奖品，每 4 人一组，按顺序依次从中摸出 1 个球并记录结果. 每组重复试验 20 次.

表 3-6 是汇总了 8 组学生的数据得到的结果.

表 3-6

	第一个人摸到白球	第二个人摸到白球	第三个人摸到白球	第四个人摸到白球
出现的次数	78	83	80	79
出现的频率	0.487 50	0.518 75	0.500 00	0.493 75

根据表 3-6，你认为每个人摸到白球的机会相等吗？

动手实践

口袋里装有 2 个白球和 2 个黑球，这 4 个球除颜色外完全相同，每 4 人一组，按顺序依次从中摸出 1 个球. 每组重复试验 20 次，将结果记录在表 3-7 中.

表 3-7

	第一个人摸到白球	第二个人摸到白球	第三个人摸到白球	第四个人摸到白球
出现的次数				
出现的频率				

（1）根据模拟得到的数据，你能得出什么结论？

（2）汇总全班的数据，估计出第一个人、第二个人、第三个人、第四个人摸到奖品的概率.

（3）你认为摸奖的次序对中奖率有影响吗？

阅读理解

为了推动体育事业和福利事业的发展，我国发行了体育彩票和福利彩票. 两种彩票的发行收入一部分用于奖金，另一部分用于支持体育事业和福利事业.

北京市有一种电脑体育彩票，每张彩票的号码是从 01，02，…，36 这 36 个数中选取 7 个组成的（彩票号码与次序无关，可以由彩民自己选取）. 开奖时，特等奖号码是从这 36 个数中随机地选取 7 个号码组成[1]（号码与次序无关），例如，第 03099 期特等奖中奖号码为：26，10，25，23，34，06，05. 如果某人所选的彩票号码也由这些号码组成，那么这个人就获得了特等奖.

[1] 通常称这 7 个号码为正选号码，开奖时还会产生一个特选号码. 为了公平起见，开奖用的 36 个球应该是均匀的、没有任何差别的. 原则上每期摇奖都应该使用新球，除非能保证用过的球仍和新球一样.

思考下面的问题，并与同学进行交流.

（1）彩民甲研究了近几期这种体育彩票的中奖号码，发现数字 06 和 08 出现的次数最多. 他认为，06 和 08 是“幸运号码”，因此，他在所买的每一注彩票中都选上了 06 和 08. 你认为他这样做有道理吗？

（2）彩民乙对近几期这种体育彩票的中奖号码也进行了研究，发现数字 04 和 09 在近几期中奖号码中一次都没出现. 他认为，既然每个数字出现的机会是相等的，那么下一期就该 04 和 09 出现了，因此，他在所买的每一注彩票中都选上了 04 和 09. 你认为这种做法有助于他中奖吗？

抽象概括

概率和日常生活有着密切的联系，对于生活中的随机事件，我们可以利用概率知识作出合理的判断与决策. 例如，“明天的降水概率为 70%”，在明天出门时我们会选择带上雨伞；“买 1 张体育彩票中特等奖的概率约为$\dfrac{1}{8\ 000\ 000}$”，我们在买体育彩票时就应抱着一种平常的心态，不要沉溺于中特等奖的梦想之中.

信息技术应用

利用计算机进行模拟

利用计算机电子表格软件(如 Excel)中的随机函数(RAND)可以对抛掷 10 次硬币这个概率试验进行模拟. RAND 函数返回大于等于 0 且小于 1 的均匀分布随机数,每次计算工作表时都将返回一个新的数值,其表达式为:RAND( ).

这种模拟的工作原理为:每次产生 10 个 0 或 1 的随机数(假设用“0”表示出现正面,用“1”表示出现反面),统计每次试验中是否正好出现 5 次正面和 5 次反面,也就是统计这 10 个随机数的和是否为 5. 重复这个试验 100 次或更多次,统计出现 5 次正面的次数占总试验次数的百分比. 具体操作步骤如下:

(1) 新建一个电子表格文件,在 A1 的位置处输入:=INT(RAND()+0.5),产生一个 0 或 1 的随机数(如图 3-3).

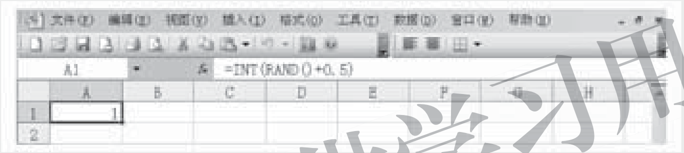


图 3-3

(2) 将 A1 位置的表达式分别复制到 B1 至 J1 处,这样就产生了 10 个 0 或 1 的随机数(如图 3-4).

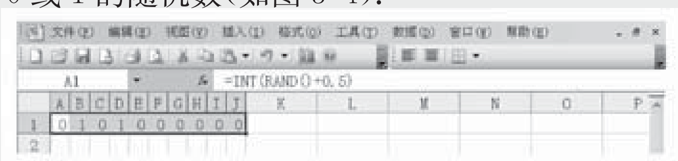


图 3-4

(3) 在 K1 位置处输入:=IF(SUM(A1:J1)=5,1,0),判断在 A1 至 J1 这些单元格里产生的 10 个随机数中是否正好有 5 个“1”和 5 个“0”,如果是,就返回数值 1,否则返回数值 0(如图 3-5).

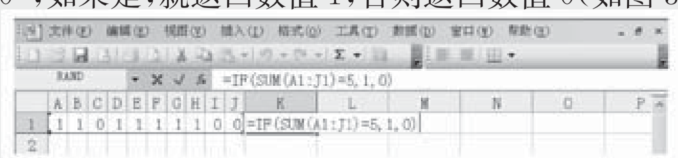
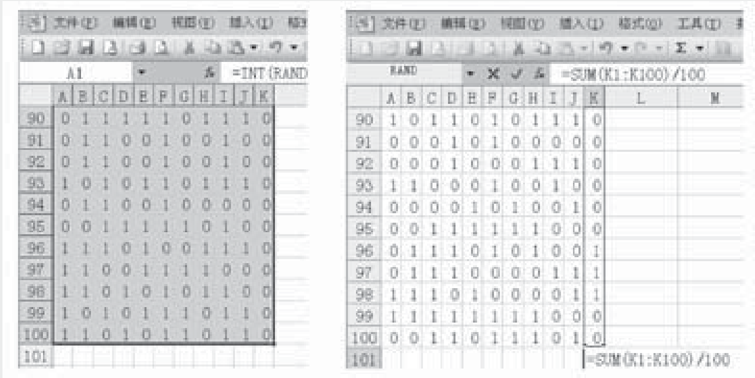


图 3-5

(4) 再将第 1 行的这 11 个表达式复制 100 行,产生 100 组这样的数据,也就是模拟 100 次这样的试验,并统计每次的试验结果

(见图 3-6①).

(5) 在 K101 处输入:=SUM(K1:K100)/100,统计这 100 次重复试验中正好出现 5 正 5 反的频率(如图 3-6②,理论值为:$C_{10}^{5}\left(\frac{1}{2}\right)^{10}=0.246\ 093\ 75$).



① ②

图 3-6

注意 产生数据以后,只要对其中的任意一个单元格进行操作,就会重新产生一次随机数,也就是对数据进行了一次更新. 对于试验而言,也就是重新产生了 100 个随机数. 这一点是非常重要的. 一种最简单的更新数据的操作就是,选中其中的任意一个单元格,然后按一次回车键就可以了.

练 习 2

1. 请举出一些日常生活中与概率有关的例子.
2. 总数为 10 万张的彩票,中奖率为$\frac{1}{1\ 000}$,买1 000张一定中奖吗? 买 10 000 张呢? 与同学交流你的看法.
3. 有两粒不均匀的骰子,第一粒骰子被掷了 500 次,“向上的点数是 6”出现了 78 次,第二粒骰子被掷了 700 次,“向上的点数是 6”出现了 102 次. 如果你希望掷得“向上的点数是 6”,你将选择哪粒骰子?

习 题 3—1

A　组

1. 在一个路口观察从你身边路过的人，连续观察 10 人，记录下出现"男性"的次数. 汇总班上所有同学的数据，估计出现"男性"的概率.
2. 一家保险公司想了解汽车的挡风玻璃破碎的概率. 公司收集了 20 000 部汽车的信息，时间是从某年的 7 月 1 日到下一年的 7 月 1 日，共发现有 600 部汽车的挡风玻璃破碎. 在一年时间里，一部汽车的挡风玻璃破碎的概率近似是多少?
3. 通过试验知道，一枚不均匀的硬币抛掷后易于出现"正面朝上"，为了估计出现"正面朝上"的概率，重复抛掷这枚硬币2 000 次，下表记录了不同抛掷次数相应的出现"正面朝上"的次数.

抛掷次数	100	500	1 000	2 000
出现"正面朝上"的次数	62	361	712	1 404

根据这 2 000 次抛掷的结果，对出现"正面朝上"的概率进行估计.

B　组

掷两粒均匀的骰子，观察并记录两粒骰子向上的点数之和.

(1) 两粒骰子向上的点数之和有多少种可能?

(2) 重复掷两粒骰子 50 次，根据试验的结果，分别估计"点数和为 4""点数和为 7""点数和为 10"的概率.

(3) 汇总班上同学的数据，得到至少 500 次试验的结果，用这些结果对上述概率重新进行估计.

(4) 为了对上述事件的概率给出比较好的估计，你需要怎么做?

(5) 你认为出现"点数和为 4""点数和为 7""点数和为 10"的概率相同吗?

§2　古典概型

2.1　古典概型的特征和概率计算公式

问题提出

在第一节里，我们做了这样一个模拟活动：

口袋里装有 2 个白球和 2 个黑球，这 4 个球除颜色外完全相同，白球代表奖品. 4 个人按顺序依次从中摸球并记录结果，估计第一个人、第二个人、第三个人、第四个人摸到白球的概率.

图 3-7

从这个模拟活动的结果我们知道，四个人顺次抓阄决定两件奖品的归属，先抓的人和后抓的人的中奖率是一样的，即摸奖的顺序不影响中奖率，先抓还是后抓对每个人来说是公平的. 那么，如何从理论上来计算出每个人的中奖率呢?

通常，我们可以通过大量的重复试验，得到某个事件发生的频率，进而来估计其发生的概率. 但是，这种方法费时、费力. 而对于某一类特殊的随机试验，我们可以根据试验结果的对称性来确定随机事件发生的概率. 利用这种方法，我们能够在这类随机事件发生之前就预知其概率.


图 3-8

例如，在以前的学习中，我们知道(如图 3-7)，抛掷一枚均匀的硬币，出现“正面朝上”的概率为$\frac{1}{2}$. 这是因为试验的可能结果只有 2 个：“正面朝上”和“反面朝上”，并且硬币是均匀的，因此，出现这两种结果的可能性是相等的，出现“正面朝上”的概率为$\frac{1}{2}$，出现“反面朝上”的概率也为$\frac{1}{2}$.

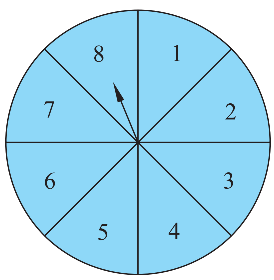


图 3-9

如图 3-8 所示，掷一粒均匀的骰子，出现“向上的点数为 6”的概率为$\frac{1}{6}$. 这是因为试验的结果只有 6 个，即“向上的点数是 1”“向上的点数是 2”……“向上的点数是 6”，并且骰子是均匀的，因此，出现这 6 种结果的可能性是相等的，每一种结果出现的概率都是$\frac{1}{6}$.

如图 3-9 所示，转动一个有 8 等份标记的转盘，出现“箭头指向

4”的概率为$\frac{1}{8}$.这是因为试验的结果只有 8 个,即“箭头指向 1”“箭头指向 2”……“箭头指向 8”,并且转盘是 8 等份,因此,箭头指向每一个数的可能性都相等,每一个结果出现的概率都是$\frac{1}{8}$.

上面三个试验都具有如下两个特征:

(1) 试验的所有可能结果只有有限个,每次试验只出现其中的一个结果;

(2) 每一个试验结果出现的可能性相同.

我们把具有这样两个特征的随机试验的数学模型称为**古典概型**(古典的概率模型).抛掷 1 枚均匀的硬币、掷 1 粒均匀的骰子以及转动一个 8 等份的转盘都可以用古典概型来描述.试验的每一个可能结果称为基本事件,例如,“反面朝上”“向上的点数为 4”“箭头指向 5”分别是上述三个试验的基本事件.

思考交流

(1) 向一个圆面内随机地投一个点,如果该点落在圆内任意一点都是等可能的,你认为可以用古典概型来表述吗?为什么?

(2) 如图 3-10 所示,射击运动员向一靶心进行射击,这一试验的结果只有有限个:命中 10 环、命中 9 环……命中 1 环和命中 0 环(即不命中).你认为可以用古典概型来表述吗?为什么?

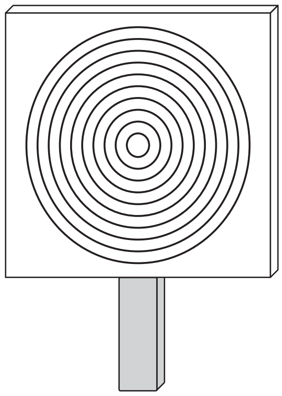


图 3-10

我们来看下面的问题:

掷一粒均匀的骰子,骰子落地时向上的点数为偶数的概率是多少?

用 A 表示事件“向上的点数是偶数”.掷一粒均匀的骰子时,试验的所有可能结果共有 6 个,即“向上的点数是 1”“向上的点数是 2”……“向上的点数是 6”,并且每个结果是等可能出现的.事件 A 由“向上的点数是 2”“向上的点数是 4”“向上的点数是 6”这 3 个可能结果组成,事件 A 发生,是指向上的点数是 2,4,6 这三种情形之一出现,因此可以认为事件 A 的概率 $P(A)=\frac{3}{6}=0.5$.

抽象概括

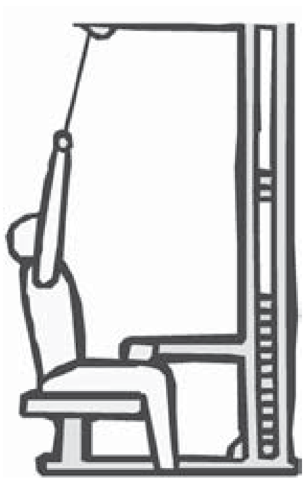
对于古典概型,通常试验中的某一事件 A 是由几个基本事件组成的.如果试验的所有可能结果(基本事件)数为 n,随机事件 A 包含

的基本事件数为 m，那么事件 A 的概率规定为

$$\boxed{P(A)=\frac{\text{事件 }A\text{ 包含的可能结果数}}{\text{试验的所有可能结果数}}=\frac{m}{n}.}$$

在古典概型中，计算事件 A 的概率，关键是计算试验的所有可能结果（基本事件）数 n 和事件 A 包含的可能结果（基本事件）数 m.

例 1 在一个健身房里，用拉力器进行锻炼时，需要选取 2 个质量盘装在拉力器上. 有 2 个装质量盘的箱子，每个箱子中都装有 4 个不同的质量盘：2.5 kg，5 kg，10 kg 和 20 kg，每次都随机地从 2 个箱子中各取 1 个质量盘装在拉力器上后，再拉动这个拉力器.



(1) 随机地从 2 个箱子中各取 1 个质量盘，共有多少种可能的结果？用表格列出所有可能的结果.

(2) 计算选取的 2 个质量盘的总质量分别是下列质量的概率：

(i)20 kg；(ii)30 kg；(iii)不超过 10 kg；(iv)超过 10 kg.

(3) 如果一个人不能拉动超过 22 kg 的质量，那么他不能拉开拉力器的概率是多少？

解 (1) 第 1 个箱子的质量盘和第 2 个箱子的质量盘都可以从 4 种不同的质量盘中任意选取. 我们可以用一个"有序实数对"来表示随机选取的结果. 例如，我们用(10，20)来表示：在一次随机的选取中，从第 1 个箱子取的质量盘是 10 kg，从第 2 个箱子取的质量盘是 20 kg. 表 3-8 列出了所有可能结果.

表 3 8

第 2 个质量盘的质量/kg 第 1 个质量盘的质量/kg	2.5	5	10	20
2.5	(2.5,2.5)	(2.5,5)	(2.5,10)	(2.5,20)
5	(5, 2.5)	(5, 5)	(5, 10)	(5, 20)
10	(10,2.5)	(10, 5)	(10, 10)	(10,20)
20	(20,2.5)	(20, 5)	(20,10)	(20,20)

从表 3-8 中可以看出，随机地从 2 个箱子中各取 1 个质量盘的所有可能结果共有 16 种. 由于选取质量盘是随机的，因此这 16 种结果出现的可能性是相同的，这个试验属于古典概型.

（2）　　**表 3 9**

总质量/kg \ 第 2 个质量盘的质量/kg \ 第1个质量盘的质量/kg	2.5	5	10	20
2.5	5	7.5	12.5	22.5
5	7.5	10	15	25
10	12.5	15	20	30
20	22.5	25	30	40

(i) 用 A 表示事件“选取的 2 个质量盘的总质量是 20 kg”，因为总质量为 20 kg 的所有可能结果只有 1 种，因此，事件 A 的概率

$$P(A)=\frac{1}{16}=0.0625.$$

(ii) 用 B 表示事件“选取的 2 个质量盘的总质量是 30 kg”，从表 3-9 中可以看出，总质量为 30 kg 的所有可能结果共有 2 种，因此，事件 B 的概率

$$P(B)=\frac{2}{16}=\frac{1}{8}=0.125.$$

(iii) 用 C 表示事件 “选取的 2 个质量盘的总质量不超过10 kg”. 总质量不超过 10 kg，即总质量为 5 kg，7.5 kg，10 kg 之一，从表 3-9 中容易看出，所有可能结果共有 4 种，因此，事件 C 的概率

$$P(C)=\frac{4}{16}=\frac{1}{4}=0.25.$$

(iv) 用 D 表示事件“选取的 2 个质量盘的总质量超过 10 kg”. 总质量超过 10 kg，即总质量为 12.5 kg，15 kg，20 kg，22.5 kg，25 kg，30 kg，40 kg 之一，从表 3-9 中可以看出，所有可能结果共有 12 种，因此，事件 D 的概率

$$P(D)=\frac{12}{16}=\frac{3}{4}=0.75.$$

(3) 用 E 表示事件“不能拉开拉力器”，即总质量超过了 22 kg. 总质量超过 22 kg 是指总质量为 22.5 kg，25 kg，30 kg，40 kg 之一. 从表 3-9 中可以看出，这样的可能结果共有 7 种，因此，不能拉开拉力器的概率

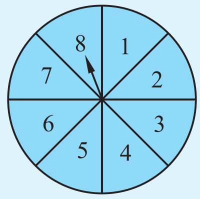
$$P(E)=\frac{7}{16}\approx 0.44.$$

说　明

在这个例子中，我们用列表的方法列出了所有可能的结果. 在计算古典概率时，只要所有可能结果的数量不是很多，列举法是我们常用的一种方法.

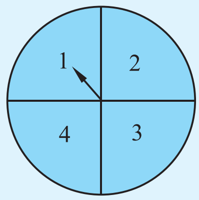
练 习

1. 一枚均匀的硬币连续抛掷 2 次，出现“2 次正面”“2 次反面”“1 次正面、1 次反面”的可能性相同吗？
2. 转动图示转盘，计算下列事件的概率：
   (1) 箭头指向 8；
   (2) 箭头指向 3 或 8；
   (3) 箭头不指向 8；
   (4) 箭头指向奇数；
   (5) 箭头指向偶数；
   (6) 箭头指向 24 的约数；
   (7) 箭头指向 3 的倍数；
   (8) 箭头指向不小于 3 的数.

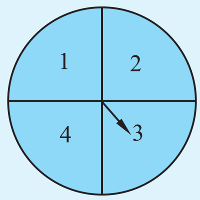


（第 2 题）

3. 同时转动如图所示的两个转盘，记转盘(A)得到的数为 x，转盘(B)得到的数为 y，用列举法列出所有可能的结果 (x,y)，计算下列事件的概率：
   (1) $x+y=5$；　　(2) $x<3$ 且 $y>1$；
   (3) $xy=4$；　　(4) $x=y$.
4. 在例 1 中，随机地从 2 个箱子中各取 1 个质量盘.
   (1) 计算总质量少于 20 kg 的概率；
   (2) 计算总质量至少 20 kg 的概率；
   (3) 计算总质量超过 20 kg 的概率；
   (4) 若某人不能拉动超过 22 kg 的质量，求他能拉开拉力器的概率.



(A)



(B)

（第 3 题）

2.2 建立概率模型

在掷一粒均匀骰子的试验中，如果考虑向上的点数是多少，那么出现 1，2，3，4，5，6 中任何一个的可能性都是 $\frac{1}{6}$；如果考虑向上的点数是奇数还是偶数，那么可以认为试验只有两个结果：“向上的点数是奇数”和“向上的点数是偶数”. 这两个结果是等可能出现的，因此它们出现的概率都是 $\frac{1}{2}$.

一般来说，在建立概率模型时，把什么看作是一个基本事件（即一个试验结果）是人为规定的. 我们只要求：每次试验有一个并且只有一个基本事件出现. 例如，掷一粒均匀的骰子时，根据问题的需要，可以认为有 6 个结果（向上的点数是 1，向上的点数是 2……向上的点数是 6），也可以认为只有 2 个结果（向上的点数是奇数，向上的点数是偶数）. 只要基本事件的个数是有限的，并且它们的发生是等可能的，就是一个古典概型.

在掷一粒均匀骰子的试验中，还可以建立概率模型，使得每一个结果出现的概率都是$\frac{1}{3}$. 例如，可以把这个骰子的 2 个面涂上黑色，另 2 个面涂上红色，剩下 2 个面涂上蓝色，则“向上的面是黑色”“向上的面是红色”和“向上的面是蓝色”出现的概率都是$\frac{1}{3}$. 你还有别的方法吗？

抽象概括

对于同一个随机试验，可以根据需要，建立满足我们要求的概率模型.

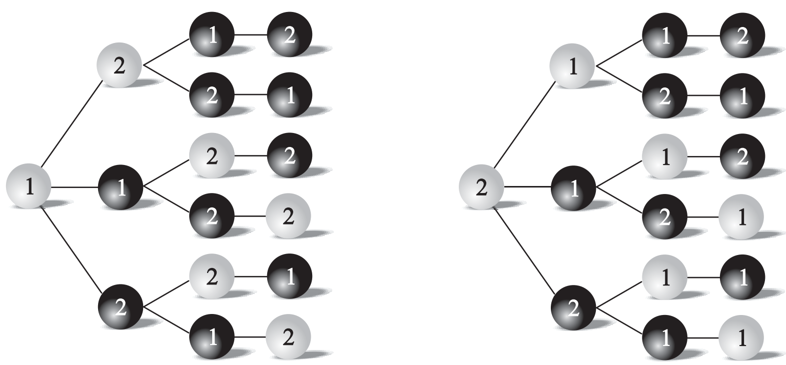
下面来考虑 2.1 节开始提到的问题.

例 2　口袋里装有 2 个白球和 2 个黑球，这 4 个球除颜色外完全相同，4 个人按顺序依次从中摸出 1 个球. 试计算第二个人摸到白球的概率.

下面，我们用四个不同的古典概型给出四种不同的解法.

模型 1　我们只需找出 4 个人按顺序依次摸球的所有可能结果数和第二个人摸到白球的可能结果数. 为此考虑用列举法列出所有可能结果.

用 A 表示事件“第二个人摸到白球”. 把 2 个白球编上序号 1，2；2 个黑球也编上序号 1，2. 于是，4 个人按顺序依次从袋中摸出 1 个球的所有可能结果，可用树状图直观地表示出来(如图 3-11).



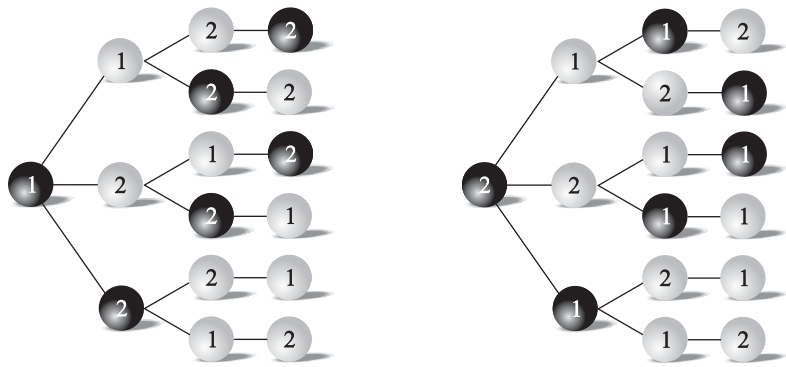


图 3-11

树状图是进行列举的一种常用方法. 从上面的树状图可以看出，试验的所有可能结果数为 24. 由于口袋内的 4 个球除颜色外完全相同，因此，这 24 种结果的出现是等可能的，试验属于古典概型. 在这 24 种结果中，第二个人摸到白球的结果有 12 种，因此“第二个人摸到白球”的概率

$$P(A)=\frac{12}{24}=\frac{1}{2}.$$

这与第一节的模拟结果是一致的.

模型 2 因为是计算“第二个人摸到白球”的概率，所以我们可以只考虑前两人摸球的情况. 前两人依次从袋中摸出 1 个球的所有可能结果可用树状图列举出来(如图 3-12).

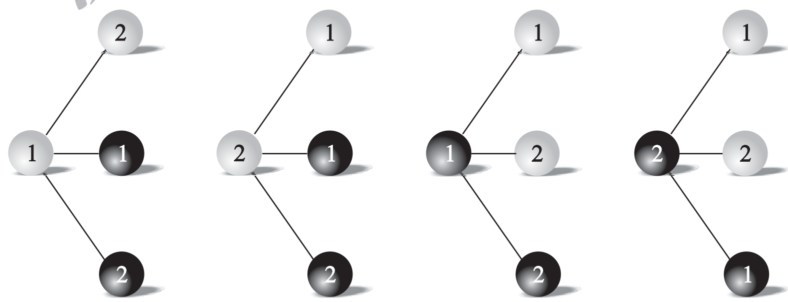


图 3-12

从上面的树状图可以看出，这个模型的所有可能结果数为 12，因为口袋里的 4 个球除颜色外完全相同，因此，这 12 种结果的出现是等可能的，这个模型也是古典概型. 在上面 12 种结果中，第二个人摸到白球的结果有 6 种，因此“第二个人摸到白球”的概率

$$P(A)=\frac{6}{12}=\frac{1}{2}.$$

这里，我们是根据事件“第二个人摸到白球”的特点，利用试验结果的对称性，只考虑前两人摸球的情况，从而简化了模型.

模型 3 还可以从另外一个角度来考虑这个问题. 因为口袋里的 4 个球除颜色外完全相同,因此,可以对 2 个白球不加区别,对 2 个黑球也不加区别,这样建立的模型的所有可能结果数就会更少,由此得到例 2 的另一种解法.

只考虑球的颜色,4 个人按顺序依次从袋中摸出 1 个球的所有可能结果可用树状图列举出来(如图 3-13).

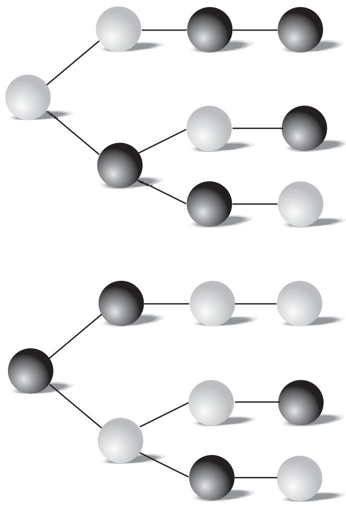


图 3-13

试验的所有可能结果数为 6,并且这 6 种结果的出现是等可能的,这个模型是古典概型. 在这 6 种结果中,第二个人摸到白球的结果有 3 种,因此"第二个人摸到白球"的概率

$$P(A)=\frac{3}{6}=\frac{1}{2}.$$

下面再给出一种更为简单的解法.

模型 4 只考虑第二个人摸出的球的情况,他可能摸到这 4 个球中的任何一个,这 4 种结果出现的可能性是相同的. 第二个人摸到白球的结果有 2 种,因此"第二个人摸到白球"的概率

$$P(A)=\frac{2}{4}=\frac{1}{2}.$$

抽象概括

(1) 从以上 4 种解法中,我们分别规定了不同的结果作为基本事件,只要我们在模型中规定了有限个基本事件,并且它们发生是等可能的,都是古典概型的问题.

(2) 这个问题说明不论第几次摸球,摸到白球的概率都是$\frac{1}{2}$,即抽签与顺序无关. 也说明统计中的简单随机抽样可以保证每一个样本被抽到的概率是相同的.

思考交流

计算第 $k(k=1,3,4)$个人摸到白球的概率. 得到的结果说明什么问题?

练 习

1. 一副扑克牌（去掉大、小王，共 52 张）有 4 种花色（梅花、方块、红心、黑桃），每一种花色有 13 张牌（A，2，3，4，5，6，7，8，9，10，J，Q，K）. 方块和红心称为红色牌，梅花和黑桃称为黑色牌. 从一副扑克牌中随机选取 1 张，计算下列事件的概率：
   (1) 这张牌是 A；　　(2) 这张牌是 K，Q 或 J；
   (3) 这张牌是红色 A；　　(4) 这张牌是梅花；
   (5) 这张牌是黑色牌.
2. 小军、小燕和小明是同班同学，假设他们三人早上到校先后的可能性是相同的.
   (1) 事件“小燕比小明先到校”的概率是多少？
   (2) 事件“小燕比小明先到校，小明又比小军先到校”的概率是多少？

2.3　互斥事件

问题提出

在例 1 中，随机地从 2 个箱子中各取 1 个质量盘，“总质量至少 20 kg”与“总质量不超过 10 kg”能否同时发生？

在一个随机试验中，我们把一次试验下不能同时发生的两个事件 A 与 B 称作**互斥事件**.

在例 1 中，“总质量至少 20 kg”与“总质量不超过 10 kg”这两个事件就是互斥事件. 生活中还有很多类似于这样的事件，它们是不能同时发生的. 例如，在图 3-14 中：

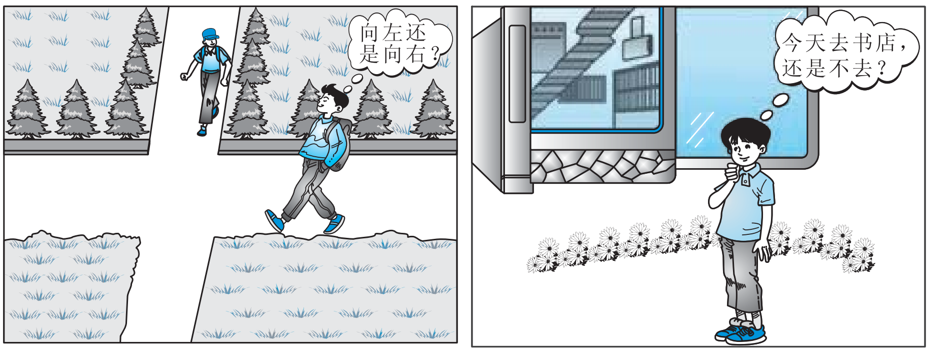


图 3-14

“向左拐弯”与“向右拐弯”是互斥事件，“去书店”与“不去书店”也是互斥事件. 你还能举出一些生活中类似的例子吗？

例 3　在例 1 中，随机地从 2 个箱子中各取 1 个质量盘，下面的事件 A 和事件 B 是否是互斥事件？

(1) 事件 A=“总质量为 20 kg”，事件 B=“总质量为 30 kg”；

(2) 事件 A=“总质量为 7.5 kg”，事件 B=“总质量超过 10 kg”；

(3) 事件 A=“总质量不超过 10 kg”，事件 B=“总质量超过 10 kg”；

(4) 事件 A=“总质量为 20 kg”，事件 B=“总质量超过10 kg”.

解　在(1)(2)(3)中，事件 A 与事件 B 不能同时发生，因此事件 A 与事件 B 是互斥事件.

对于(4)中的事件 A 和事件 B，随机地从 2 个箱子中各取 1 个质量盘，当总质量为 20 kg 时，事件 A 与事件 B 同时发生，因此，事件 A 与事件 B 不是互斥事件.

在例 3 的(1)中，A 表示事件“总质量为 20 kg”，B 表示事件“总质量为 30 kg”，我们把事件“总质量为 20 kg 或 30 kg”记作 $A+B$.

给定事件 A，B，我们规定 $A+B$ 为一个事件，**事件 $A+B$ 发生是指事件 A 和事件 B 至少有一个发生.**

思考交流

(1) 对于例 3 的(2)和(3)中的事件 A 和事件 B，$A+B$ 表示什么事件？

(2) 对例 3 的(1)，(2)和(3)中的每一对事件，通过计算完成表 3-10.

表 3 10

	(1)	(2)	(3)
$P(A)$			
$P(B)$			
$P(A)+P(B)$			
$P(A+B)$			

根据表 3-10 中的结果，你发现 $P(A+B)$ 与 $P(A)+P(B)$ 有什么样的大小关系？

抽象概括

在一个随机试验中，如果随机事件 A 和事件 B 是互斥事件，那么有

$$\boxed{P(A+B)=P(A)+P(B).}$$

例 4 从一箱产品中随机地抽取一件产品，设事件 A=“抽到的是一等品”，事件 B=“抽到的是二等品”，事件 C=“抽到的是三等品”，且已知 $P(A)=0.7$，$P(B)=0.1$，$P(C)=0.05$. 求下列事件的概率：

(1) 事件 D=“抽到的是一等品或三等品”；

(2) 事件 E=“抽到的是二等品或三等品”.

解 (1) 事件 D 即事件 $A+C$，因为事件 A=“抽到的是一等品”和事件 C=“抽到的是三等品”是互斥事件，由互斥事件的概率加法公式，

$P(D)=P(A+C)=P(A)+P(C)=0.7+0.05=0.75.$

(2) 事件 E 即事件 $B+C$，因为事件 B=“抽到的是二等品”和事件 C=“抽到的是三等品”是互斥事件，由互斥事件的概率加法公式，

$P(E)=P(B+C)=P(B)+P(C)=0.1+0.05=0.15.$

思考交流

事件 $D+E$ 表示的是什么？它的概率 $P(D+E)$ 等于 $P(D)+P(E)$ 吗？

容易看出，事件 $D+E$ 表示“抽到的产品是一等品或二等品或三等品”. 事件 D 和事件 E 不是互斥事件，因此不满足互斥事件的概率加法公式. 事实上，$P(D+E)=P(A)+P(B)+P(C)=0.85$，而 $P(D)+P(E)=[P(A)+P(C)]+[P(B)+P(C)]=0.9$，“抽到的是三等品”的概率 $P(C)$ 在 $P(D)$ 和 $P(E)$ 中各算了一次，因此，事件 $D+E$ 的概率 $P(D+E)$ 不等于 $P(D)+P(E)$.

例 5 某地政府准备对当地的农村产业结构进行调整，为此政府进行了一次民意调查. 100 个人接受了调查，他们被要求在赞成调整、反对调整、对这次调整不发表看法中任选一项. 调查结果如表 3-11 所示：

表 3 11

	男	女	总计
赞成	18	9	27
反对	12	25	37
不发表看法	20	16	36
总计	50	50	100

随机选取一个被调查者，他对这次调整表示反对或不发表看法的概率是多少?

解　用 A 表示事件“对这次调整表示反对”，B 表示事件“对这次调整不发表看法”，则 A 和 B 是互斥事件，并且 $A+B$ 就表示事件“对这次调整表示反对或不发表看法”，由互斥事件的概率加法公式，

$$P(A+B)=P(A)+P(B)=\frac{37}{100}+\frac{36}{100}=\frac{73}{100}=0.73,$$

因此，随机选取的一个被调查者对这次调整表示反对或不发表看法的概率是 0.73.

在例 1 中，随机地从 2 个箱子中各取 1 个质量盘，如果 A=“总质量超过 10 kg”，那么我们把事件“总质量不超过 10 kg”称为事件 A 的**对立事件**[1]，记作 $\overline{A}$.

在每一次试验中，相互对立的事件 A 和事件 $\overline{A}$ 不会同时发生，并且一定有一个发生.

例如，在例 1 中，随机地从 2 个箱子中各取 1 个质量盘，若 A=“总质量少于 20 kg”，则 $\overline{A}$=“总质量至少 20 kg”；任意抽取 1 件产品，若 A=“产品合格”，则 $\overline{A}$=“产品不合格”.

若 A=“产品合格”的概率 $P(A)=0.90$，显然，$\overline{A}$=“产品不合格”的概率 $P(\overline{A})=1-0.90=0.10$；若 B=“射击命中”的概率 $P(B)=0.05$，则 $\overline{B}$=“射击不命中”的概率 $P(\overline{B})=1-0.05=0.95$.

一般地，我们有

$$\boxed{P(\overline{A})=1-P(A).}$$

在例 5 中，若事件 C=“对这次调整表示赞成”，则其对立事件 $\overline{C}$=“对这次调整表示反对或不发表看法”，因此，随机选取一个被调查者，他对这次调整表示反对或不发表看法的概率还可以按如下方法计算：

[1] 对立事件也称为逆事件.

$$P(\overline{C})=1-P(C)=1-\frac{27}{100}=\frac{73}{100}=0.73.$$

例 6 某学校成立了数学、英语、音乐 3 个课外兴趣小组，3 个小组分别有 39，32，33 个成员，一些成员参加了不止 1 个小组，具体情况如图 3-15 所示. 随机选取 1 个成员：

(1) 他至少参加 2 个小组的概率是多少？

(2) 他参加不超过 2 个小组的概率是多少？

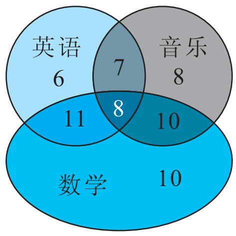


图 3-15

解 (1) 从图 3-15 可以看出，3 个课外兴趣小组总人数为 60. 用 A 表示事件"选取的成员只参加 1 个小组"，则 $\overline{A}$ 就表示"选取的成员至少参加 2 个小组"，于是，

$$P(\overline{A})=1-P(A)=1-\frac{6+8+10}{60}=\frac{3}{5}=0.6.$$

因此，随机选取的 1 个成员至少参加 2 个小组的概率是 0.6.

(2) 用 B 表示事件"选取的成员参加 3 个小组"，则 $\overline{B}$ 就表示"选取的成员参加不超过 2 个小组"，于是，

$$P(\overline{B})=1-P(B)=1-\frac{8}{60}=\frac{13}{15}\approx 0.87.$$

所以，随机选取的 1 个成员参加不超过 2 个小组的概率约等于 0.87.

在例 1 中我们知道，随机地从 2 个箱子中各取 1 个质量盘，事件 A："总质量不超过 10 kg"即总质量为 5 kg，7.5 kg，10 kg 之一. 如果 A_1 表示事件"总质量为 5 kg"，A_2 表示事件"总质量为 7.5 kg"，A_3 表示事件"总质量为 10 kg"，那么我们把事件"总质量不超过 10 kg"记作 $A_1+A_2+A_3$，事件 $A_1+A_2+A_3$ 发生是指 A_1，A_2，A_3 中至少有一个发生.

若 A_1，A_2，A_3 中任意两个都是互斥事件，容易得出，

$$P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3).$$

一般地，如果随机事件 A_1，A_2，…，A_n 中任意两个是互斥事件，那么有

$$\boxed{P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n).}$$

在例 1 中，随机地从 2 个箱子中各取 1 个质量盘，如果一个人不能拉动超过 22 kg 的质量，他不能拉开拉力器的概率还可按下面的方法计算.

总质量超过 22 kg，即总质量为 22.5 kg，25 kg，30 kg，40 kg 之一.

用 A_1 表示事件“总质量为 22.5 kg”,A_2 表示事件“总质量为25 kg”,A_3 表示事件“总质量为 30 kg”,A_4 表示事件“总质量为 40 kg”,则 $A_1+A_2+A_3+A_4$ 就表示事件“总质量超过 22 kg”. 而 A_1,A_2,A_3,A_4 中任意两个都是互斥事件,由互斥事件的概率加法公式,随机地从 2 个箱子中各取 1 个质量盘,总质量超过 22 kg 的概率为

$$\begin{aligned}P(A_1+A_2+A_3+A_4)&=P(A_1)+P(A_2)+P(A_3)+P(A_4)\\&=\frac{2}{16}+\frac{2}{16}+\frac{2}{16}+\frac{1}{16}\\&=\frac{7}{16}\approx 0.44.\end{aligned}$$

因此,随机地从 2 个箱子中各取 1 个质量盘,此人不能拉开拉力器的概率约为 0.44.

练　习　1

1. 掷一粒均匀的骰子,用 A 表示事件“向上的点数至少为 5”,则

   (1) $\overline{A}$ 指什么事件?　　(2)$\overline{A}$ 的对立事件是什么?

2. 从一副扑克牌(去掉大、小王,共 52 张)中随机选取 1 张,下列每组事件是否是互斥事件?若是互斥事件,那么是否是互为对立事件?若不是互为对立事件,分别说出事件 A、事件 B 的对立事件.

   (1) A=“这张牌是红心”,B=“这张牌是方块”;

   (2) A=“这张牌是红心”,B=“这张牌是 K”;

   (3) A=“这张牌是红色牌”,B=“这张牌是黑色牌”;

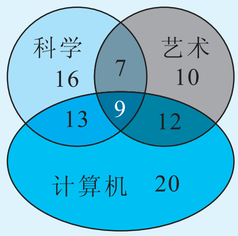
   (4) A=“这张牌牌面是 2,3,4,6,10 之一”,B=“这张牌是方块”;

   (5) A=“这张牌牌面是 2,3,4,5,6,7,8,9,10 之一”,
   B=“这张牌牌面是 A,K,Q,J 之一”;

   (6) A=“这张牌是牌面为 2,3,4,5,6,7 之一的一张方块”,B=“这张牌是牌面为 8,9,10,J,Q,K,A 之一的一张方块”.

3. 一种计算机芯片可以正常使用的概率为 0.994,它不能正常使用的概率是多少?

4. 一所大学有科学、艺术、计算机 3 个学生协会,它们分别有 45,38,54 个成员,一些成员属于不止 1 个协会,具体情况如图所示. 随机选取 1 个成员,他属于不止 1 个协会的概率是多少?



(第 4 题)

例 7　小明的自行车用的是密码锁,密码锁的四位数密码由 4 个数字 2,4,6,8 按一定顺序构成. 小明不小心忘记了密码中 4 个数字的顺序,试问:随机地输入由 2,4,6,8 组成的一个四位数,不能打开锁的概率是多少?

解　用 A 表示事件“输入由 2,4,6,8 组成的一个四位数,不是密

码”，A 比较复杂，可考虑它的对立事件，即“输入由 2，4，6，8 组成的一个四位数，恰是密码”，它只有一种结果. 利用树状图可以列出输入由 2，4，6，8 组成的一个四位数的所有可能结果（如图 3-16）.

从图 3-16 可以看出，所有可能结果数为 24，并且每一种结果出现的可能性是相同的，这是一个古典概型. $P(\bar{A})=\dfrac{1}{24}$，因此，

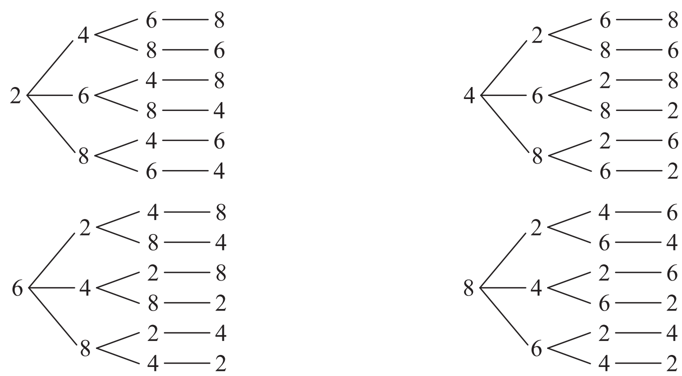


图 3-16

$$P(A)=1-P(\bar{A})=\frac{23}{24}\approx 0.958,$$

即小明随机地输入由 2，4，6，8 组成的一个四位数，不能打开锁的概率约为 0.958.

抽象概括

在概率计算的问题中，当事件 A 比较复杂而 $\bar{A}$ 比较简单时，我们往往通过计算 $\bar{A}$ 的概率 $P(\bar{A})$ 来求得 A 的概率 $P(A)$.

例 8 班级联欢时，主持人拟出了如下一些节目：跳双人舞、独唱、朗诵等. 指定 3 个男生和 2 个女生来参与，把 5 个人分别编号为 1，2，3，4，5，其中 1，2，3 号是男生，4，5 号是女生. 将每个人的号分别写在 5 张相同的卡片上，并放入一个箱子中充分混合，每次从中随机地取出一张卡片，取出谁的编号谁就参与表演节目.

（1）为了取出 2 人来表演双人舞，连续抽取 2 张卡片，求取出的 2 人不全是男生的概率.

（2）为了取出 2 人分别表演独唱和朗诵，抽取并观察第一张卡片后，又放回箱子中，充分混合后再从中抽取第二张卡片. 求：

（i）独唱和朗诵由同一个人表演的概率.

（ii）取出的 2 人不全是男生的概率.

解 （1）利用树状图我们可以列出连续抽取 2 张卡片的所有可能结果（如图 3-17）.

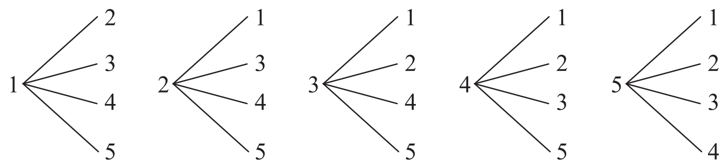


图 3-17

由图 3-17 可以看出，试验的所有可能结果数为 20. 因为每次都是随机地抽取，因此这 20 种结果出现的可能性是相同的，试验属于古典概型.

解法 1　用 A_1 表示事件“连续抽取 2 张卡片，取出的 2 人中恰有 1 位女生”，A_2 表示事件“连续抽取 2 张卡片，取出的 2 人都是女生”，则 A_1 与 A_2 互斥，并且 A_1+A_2 表示事件“连续抽取 2 张卡片，取出的 2 人不全是男生”. 由列出的所有可能结果可以看出，A_1 的结果有 12 种，A_2 的结果有 2 种，由互斥事件的概率加法公式，

$$P(A_1+A_2)=P(A_1)+P(A_2)=\frac{12}{20}+\frac{2}{20}=\frac{7}{10}=0.7,$$

即连续抽取 2 张卡片，取出的 2 人不全是男生的概率为 0.7.

解法 2　用 A 表示事件“连续抽取 2 张卡片，取出的 2 人全是男生”，则 $\bar{A}$ 就表示“连续抽取 2 张卡片，取出的 2 人不全是男生”，A 的结果有 6 种，因此，

$$P(\bar{A})=1-P(A)=1-\frac{6}{20}=\frac{7}{10}=0.7,$$

即连续抽取 2 张卡片，取出的 2 人不全是男生的概率为 0.7.

解法 3　如果我们不考虑抽取的顺序，而只看抽取的结果，这样建立的模型的所有可能结果数就会比原来减少，从而简化运算.

不考虑抽取的顺序，用记号[2,4]表示“取出的 2 人是 2 号和 4 号”，则所有可能结果可列举如下：

[1,2]	[1,3]	[1,4]	[1,5]	[2,3]
[2,4]	[2,5]	[3,4]	[3,5]	[4,5]

可以看出，试验的所有可能结果数为 10，并且每一种结果出现的可能性是相同的，这也是一个古典概型. 事件 A=“连续抽取 2 张卡片，取出的 2 人全是男生”，其结果有 3 种，因此，“连续抽取 2 张卡片，取出的 2 人不全是男生”的概率

$$P(\bar{A})=1-P(A)=1-\frac{3}{10}=\frac{7}{10}=0.7.$$

(2) 有放回地连续抽取 2 张卡片，需注意同一张卡片可再次被取出，并且它被取出的可能性和其他卡片相同. 我们用一个有序实数对来表示抽取的结果，例如，“第一次取出 2 号，第二次取出 4 号”就用

(2,4)来表示. 所有的可能结果可以用表 3-12 列出.

试验的所有可能结果数为 25,并且这 25 种结果出现的可能性是相同的,试验属于古典概型.

(i) 用 A 表示事件"独唱和朗诵由同一个人表演". 由表 3-12 可以看出,A 的结果共有 5 种,因此独唱和朗诵由同一个人表演的概率

$$P(A)=\frac{5}{25}=\frac{1}{5}=0.2.$$

表 3 12

第二次抽取 / 第一次抽取	1	2	3	4	5
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)

(ii) **解法 1**　用 A_1 表示事件"有放回地连续抽取 2 张卡片,取出的 2 人中恰有一位女生",A_2 表示事件"有放回地连续抽取 2 张卡片,取出的 2 人都是女生",则 A_1 与 A_2 互斥,并且 A_1+A_2 表示事件"有放回地连续抽取 2 张卡片,取出的 2 人不全是男生".

从表 3-12 中列出的所有可能结果可以看出,A_1 的结果有 12 种,A_2 的结果有 4 种,由互斥事件的概率加法公式,

$$P(A_1+A_2)=P(A_1)+P(A_2)=\frac{12}{25}+\frac{4}{25}=\frac{16}{25}=0.64.$$

所以,有放回地连续抽取 2 张卡片,取出的 2 人不全是男生的概率为 0.64.

解法 2　用 A 表示事件"有放回地连续抽取 2 张卡片,取出的 2 人全是男生",则 $\bar{A}$ 就表示"有放回地连续抽取 2 张卡片,取出的 2 人不全是男生",A 的结果有 9 种,因此,

$$P(\bar{A})=1-P(A)=1-\frac{9}{25}=\frac{16}{25}=0.64.$$

思考交流

有放回地连续抽取 2 张卡片,"取出的都是 1 号"和"取出的是 2 号和 4 号"出现的可能性相同吗? 我们能否也不考虑抽取的顺序来建立古典概型?

练 习 2

1. 在一次抽奖活动中，中奖者必须从一个箱子中取出一个数字来决定他获得什么奖品. 5 种奖品的编号如下：①一次欧洲旅行；②一辆摩托车；③一套高保真音响；④一台数字电视；⑤一台微波炉.
   (1) 他获得去欧洲旅行的概率是多少？
   (2) 他获得高保真音响或数字电视的概率是多少？
   (3) 他不获得微波炉的概率是多少？
2. 一个口袋里装有 2 个白球和 2 个黑球，这 4 个球除颜色外完全相同，不放回地从中连续抽取 2 次，每次取出 1 球. 计算下列事件的概率：
   (1) 第一次取出黑球，第二次取出白球；
   (2) 取出的 2 球颜色不同；
   (3) 取出的 2 球中至少有 1 个白球.

习 题 3—2

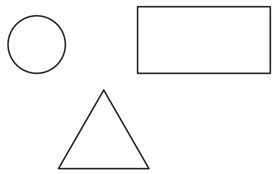
A 组

1. 一个箱子中装有铁球若干个，总重 4 kg，其中红色球总重3 kg，白色球总重 1 kg，并且每个球除颜色外完全相同. 从箱子中任取 1 球，取到红球的概率是多少？
2. 下表给出了某大学数学专业的 85 名学生的年龄和性别(如有 15 名学生为男生且年龄不超过 20 岁)，从该专业的学生中随机地选取 1 名，计算下列事件的概率：

性别 \ 年龄	不超过 20 岁	超过 20 岁
男生	15	30
女生	20	20

   (1) 选取的学生是不超过 20 岁的男生；
   (2) 选取的学生是男生；
   (3) 选取的学生不超过 20 岁；
   (4) 选取的学生是女生或超过 20 岁的男生.
3. 掷一对不同颜色的均匀的骰子.
   (1) 用列表的方法列出所有可能结果，共有多少种可能结果？
   (2) 两粒骰子向上的点数之和有多少种可能？出现哪种点数和的可能性最大？其概率是多少？
   (3) 计算下列事件的概率：
      (i) 点数和不大于 7；　　(ii) 点数和大于 7；
      (iii) 点数和为 6 或 7；　　(iv) 点数和不小于 6；
      (v) 点数和是奇数；　　(vi) 点数和是偶数；
      (vii) 点数和等于 3 的倍数.

4. 三个小组参加植树活动，第一小组由 2 名男生 3 名女生组成，第二小组由 3 名男生 2 名女生组成，第三小组由 4 名男生 1 名女生组成. 为了确定由谁来给树浇水，每个人将自己的名字写在一张纸片上，并把所有纸片放入一个帽子中，将纸片在帽子中充分混合，从帽子中随机地取出一张纸片，取到谁的名字谁就给树浇水. 计算下列事件的概率：
(1) 取到第二小组的 1 名成员；
(2) 取到 1 名男生；
(3) 取到的不是第三小组的成员.
5. 假设你班上每个人的生日在一年 365 天中的任何一天的可能性是相同的，从你班上随机选取一人，计算下列事件的概率：
(1) 他的生日是 11 月 29 日；
(2) 他的生日是在 11 月；
(3) 他的生日是在 1 月 15 日和 2 月 15 日之间(不包括 1 月 15 日和 2 月 15 日)；
(4) 他的生日是在这一年的前 3 个月；
(5) 他的生日不是 4 月 15 日；
(6) 他的生日不在 7 月.
6. 甲盒子里装有分别标有 1，3，5，7，9 的 5 张卡片，乙盒子里装有分别标有 1，4，9 的 3 张卡片，从两个盒子中各随机地取出 1 张卡片，计算 2 张卡片上的数字之和能被 3 整除的概率.
7. 幼儿园的一个小朋友正在给一个圆、一个三角形和一个长方形着色，有红、蓝两种颜色可供选择，对于每一个图形，他都随机地选择一种颜色涂上.



(第 7 题)

(1) 利用树状图列出所有的可能结果；
(2) 计算下列事件的概率：
(i) 三个图形都被涂上红色；
(ii) 圆被涂上红色；
(iii) 三角形和长方形被涂上不同的颜色；
(iv) 三个图形的颜色不全相同.
8. 从一副扑克牌(去掉大、小王，共 52 张)中随机地选取 1 张，计算下列事件的概率：
(1) 这张牌是红色牌；
(2) 这张牌是黑色 A；
(3) 这张牌是黑色 K，黑色 Q 或黑色 J；
(4) 这张牌牌面是 5 的倍数且是红色；
(5) 这张牌不是方块.
9. 我国西部一个地区的年降水量在下列区间内的概率如下表所示：

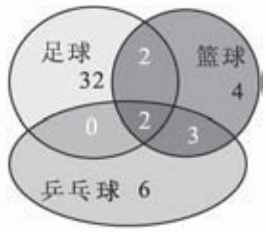
年降水量/mm	[100，150)	[150，200)	[200，250)	[250，300]
概率	0.21	0.16	0.13	0.12

(1) 计算年降水量在[200，300](mm)范围内的概率；
(2) 计算年降水量在[100，250)(mm)范围内的概率.

10. 一个学校的足球队、篮球队和乒乓球队分别有 36，11，11 名成员，一些成员参加了不止 1 支球队，具体情况如图所示. 随机选取 1 名成员：

(1) 他属于不止 1 支球队的概率是多少？

(2) 他属于不超过 2 支球队的概率是多少？



(第 10 题)

11. 一个盒子中装有 1 个红球和 2 个白球，这 3 个球除颜色外完全相同，有放回地连续抽取 2 次，每次从中任意地取出 1 个球. 用列表的方法列出所有可能结果，计算下列事件的概率：

(1) 取出的 2 个球都是白球；

(2) 第一次取出白球，第二次取出红球；

(3) 取出的 2 个球是 1 红 1 白；

(4) 取出的 2 个球中至少有 1 个白球.

B　组

1. (1) 如果口袋里装有 m 个白球和 n 个黑球，这 $m+n$ 个球除颜色外完全相同，$m+n$ 个人按顺序依次从中摸出 1 个球，则第 $k(1\leqslant k\leqslant m+n)$ 个人摸到白球的概率是多少？

(2) 有 $m+n$ 个阄，其中 m 个分别代表 m 件奖品，如果 $m+n$ 个人按顺序依次抓阄来决定这 m 件奖品的归属，则第 $m+n$ 个人中奖的概率是多少？

2. 假设小军、小燕和小明所在的班级共有 50 名学生，并且这 50 名学生早上到校先后的可能性是相同的.

(1)事件“小燕比小明先到校”的概率是多少？

(2)事件“小燕比小明先到校，小明又比小军先到校”的概率是多少？

3. 楼道管理员共有 5 把钥匙，其中有两把是房门钥匙. 晚上因楼道停电，他随机地取了两把去开房门(两把锁都打开，房门才能打开，如果没有打开，允许交换两把的顺序再试一次)，他不能打开房门的概率是多少？

4. 某工厂周一到周六轮到由甲、乙、丙 3 人值班，每人值两天，3 人通过抽签决定每个人在哪两天值班，则周六由乙值班的概率为多少？

5. 有 4 张面值相同的债券，其中有 2 张中奖债券.

(1) 有放回地从债券中任取 2 次，每次取出 1 张，计算取出的 2 张都是中奖债券的概率.

(2) 无放回地从债券中任取 2 次，每次取出 1 张，计算取出的 2 张都是中奖债券的概率.

(3) 有放回地从债券中任取 2 次，每次取出 1 张，计算取出的 2 张中至少有 1 张是中奖债券的概率.

(4) 无放回地从债券中任取 2 次，每次取出 1 张，计算取出的 2 张中至少有 1 张是中奖债券的概率.

§3 模拟方法——概率的应用

从前面的学习中我们知道，可以通过做大量的重复试验，用随机事件发生的频率来估计其概率. 但是，人工进行试验费时、费力，并且有时很难实现. 因此，我们常常借助模拟方法来估计某些随机事件发生的概率. 用模拟方法可以在短时间内完成大量的重复试验. 在第一节中我们用随机数表产生随机数来模拟抛掷硬币的试验，以及通过 4 个人依次摸球来模拟摸奖的活动，都属于模拟方法.

模拟方法是一种非常有效而且应用广泛的方法. 当现实中的试验难以实施或者不可能实施时，模拟可以给我们提供一个解决方案. 下面我们介绍模拟方法的基本思想.

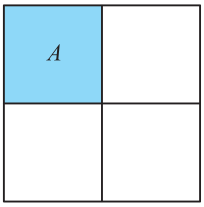


图 3-18

向图 3-18 所示的正方形中随机地撒一把芝麻，假设每一粒芝麻落在正方形内的每一个位置的可能性都是相同的. 由于区域 A 的面积是整个正方形的面积的 $\frac{1}{4}$，因此，大约有 $\frac{1}{4}$ 的芝麻落在区域 A 中. 例如，向图 3-18 所示的正方形中随机地撒 100 粒芝麻，则大约有 25 粒落在区域 A 内. 因此，近似地有

$$\frac{\text{落在区域 } A \text{ 内的芝麻数}}{\text{落在正方形内的芝麻数}}=\frac{\text{区域 } A \text{ 的面积}}{\text{正方形的面积}}.$$

图 3-19

反之，向图 3-19 所示的长方形中随机地撒一把芝麻，例如，撒了 100 粒，这些芝麻均匀地落在长方形中，如果落在区域 B 中的芝麻数是 20，那么区域 B 的面积近似地是整个长方形的面积的 20%.

利用上述思想，根据

$$\frac{\text{落在区域 } A \text{ 内的芝麻数}}{\text{落在正方形内的芝麻数}}=\frac{\text{区域 } A \text{ 的面积}}{\text{正方形的面积}},$$

我们可以求出某些不规则图形的近似面积.

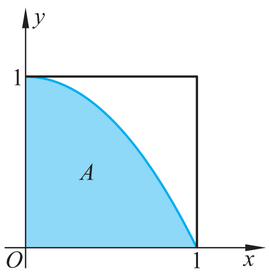


图 3-20

如图 3-20，曲线 $y=-x^2+1$ 与 x 轴，y 轴围成一个区域 A，直线 $x=1$，直线 $y=1$，x 轴，y 轴围成一个正方形，向正方形中随机地撒一把芝麻，数出落在区域 A 内的芝麻数与落在正方形内的芝麻数，由

$$\frac{\text{落在区域 } A \text{ 内的芝麻数}}{\text{落在正方形内的芝麻数}}=\frac{\text{区域 } A \text{ 的面积}}{\text{正方形的面积}},$$

就可求出区域 A 的近似面积.

抽象概括

向平面上有限区域(集合)G 内随机地投掷点 M,若点 M 落在子区域 $G_1 \subsetneqq G$ 的概率与 G_1 的面积成正比,而与 G 的形状、位置无关,即

$$P(\text{点 } M \text{ 落在 } G_1)=\frac{G_1 \text{ 的面积}}{G \text{ 的面积}},$$

则称这种模型为**几何概型**.

几何概型中的 G 也可以是空间中或直线上的有限区域,相应的概率是体积之比或长度之比.

阅读理解

如果有条件的话,你可以利用计算机来模拟向图 3-20 中撒芝麻的试验,并统计出落在区域 A 内的芝麻数与落在正方形中的芝麻数.

如表 3-13,由计算机产生了两列 0～1 的随机数,它们分别表示随机点 (x,y) 的坐标. 如果一个点 (x,y) 满足 $y \leqslant -x^2+1$,就表示这个点落在区域 A 内,在表 3-13 中最后一列相应地就填上 1,否则填 0.

表 3 13

x	y	计数
0.598 895	0.940 794	0
0.512 284	0.118 961	1
0.496 841	0.784 417	0
0.112 796	0.690 634	1
0.359 600	0.371 441	1
0.101 260	0.650 502	1
⋮	⋮	⋮
0.947 386	0.902 127	0
0.117 618	0.305 673	1
0.516 465	0.222 907	1
0.596 393	0.969 695	0

统计出落在区域 A 内的随机点的个数与落在正方形中的随机点的个数,我们就可以求得区域 A 面积的近似值.

下面我们通过例子来说明如何用模拟方法估计随机事件的概率.

问题提出

小明家的晚报在下午 5:30～6:30 的任何一个时间随机地被送到，小明一家人在下午 6:00～7:00 的任何一个时间随机地开始晚餐.

(1) 你认为晚报在晚餐开始之前被送到和在晚餐开始之后被送到哪一种可能性更大?

(2) 晚报在晚餐开始之前被送到的概率是多少?

晚报在 5:30～6:00 送到，或晚餐在 6:30～7:00 开始，这两种情况都使得晚报的送达在晚餐开始之前，因此晚报在晚餐开始之前被送到的可能性更大.

我们可以用模拟方法来估计晚报在晚餐开始之前被送到的概率.

动手实践

我们用图 3-21 所示的两个转盘来模拟上面的过程，一个转盘用于模拟晚报的送达，另一个转盘用于模拟开始晚餐，两个转盘各转动一次并记录下结果就完成一次模拟.

(1) 转动每个转盘 50 次，并记录下每次转动的结果.

(2) 汇总全班的结果. 根据全班的模拟结果，估计“晚报在晚餐开始之前被送到”的概率.

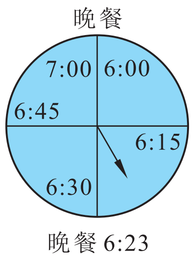


图 3-21

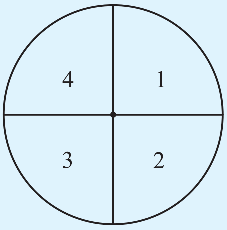
思考交流

(1) 设晚报在下午 5:45～6:45 的任何一个时间随机地被送到，而小明一家人还是在下午 6:00～7:00 的任何一个时间随机地开始晚餐. “晚报在晚餐开始之前被送到”的概率相对上面的问题来说是变大还是变小了? 将你的结论与同学进行交流.

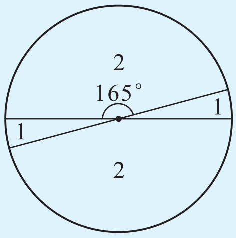
(2) 用两个转盘去模拟上面的过程，至少完成 50 次模拟. 根据你模拟的结果估计“晚报在晚餐开始之前被送到”的概率. 你的估计与你上面的结论吻合吗?

练 习

1. 哪种类型的试验可以用抛掷一枚硬币作为模拟模型？说明理由.
2. 怎样用随机数表来模拟转动下面每个转盘的试验？



(1)



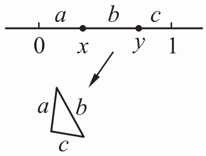
(2)

（第 2 题）

习 题 3—3

A 组

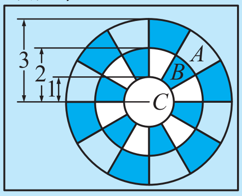
1. 设计模拟方法估计 6 个人中至少有 2 个人的生日在同一个月的概率(假设每个人生日在任何一个月的可能性是相同的).
2. 如图,在 0～1 随机选择两个数 x, y,这两个数对应的点把 0～1 的线段分成了三条线段 a, b, c,用模拟方法估计这三条线段 a, b, c 能构成三角形的概率.



（第 2 题）

B 组

1. 北京市在市民中发起了无偿献血活动.假设每个献血者到达采血站是随机的,并且每个献血者到达采血站和其他的献血者到达采血站是互相没有影响的.在所有人中,通常 45%的人的血型是 O 型,如果一天内有 10 个献血者到采血站献血,用随机数表产生随机数来模拟观察 10 个献血者的血型,完成 10 次模拟,估计至少 4 个献血者的血型是 O 型的概率.
2. 向图示的圆形镖靶连续投掷 3 枚飞镖,假设圆形镖靶上任意一点被投到的可能性都相同.用随机数表产生随机数进行模拟,完成 10 次模拟,试估计:
   (1) 恰好有 2 枚镖落在环形区域 B 内的概率;
   (2) 恰好有 1 枚镖落在圆形区域 C 内的概率.



（第 2 题）

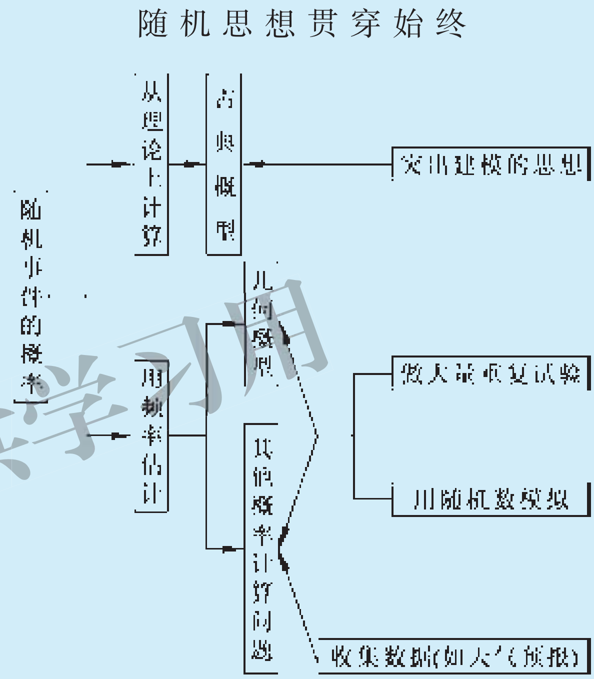
3. 用模拟方法估计 $y=e^x$ 与 x 轴,$x=0$,$x=5$ 围成的区域的面积.

◆ 本章小结

一、内容提要

本章的主要内容是随机事件的概率、古典概型以及模拟方法估计概率.

1. 本章的结构框图如下：



2. 生活中有很多的随机事件. 随机事件 A 在一次试验中是否发生具有不确定性，但是在相同条件下的大量重复试验中，随机事件 A 发生的频率会在某个常数附近摆动，即随机事件 A 发生的频率具有稳定性. 这个常数就叫作随机事件 A 的概率，记作 $P(A)$. 在实际问题中，某些随机事件的概率往往难以确切得到，因此常常通过做大量的重复试验，用随机事件发生的频率来估计它的概率.

3. 生活中会遇到一些对概率的错误认识，通过动手试验以及模拟活动，在正确理解随机事件发生的不确定性及其频率的稳定性的基础上，澄清这些错误认识.

4. 古典概型具有如下两个特征：

(1) 试验的所有可能结果只有有限个；

(2) 每一个试验结果出现的可能性相同.

对于古典概型，如果试验的所有可能结果（基本事件）数为 n，随机事件 A 包含的可能结果（基本事件）数为 m，那么事件 A 的概率规定为 $P(A)=\frac{m}{n}$.

5. 对于同一个随机试验，可以根据需要建立概率模型. 在建立概率模型时，把什么看作是一个基本事件（即一个试验结果）是人为规定的，我们只要求每次试验有一个并且只有一个基本事件出现. 只要基本事件的个数是有限的，并且它们的发生是等可能的，就是一个古典概型.

对于一个实际问题，有时我们从不同的角度去考虑，可以将问题转化为不同的古典概型来解决；而所得到的古典概型的所有可能结果数越少，问题的解决就变得越简单.

6. 不能同时发生的两个事件 A 与 B 称作互斥事件. 在一个随机试验中，如果随机事件 A 和 B 是互斥事件，那么

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

事件"A 不发生"称为 A 的对立事件，记作 $\bar{A}$. A 和 $\bar{A}$ 互为对立事件. 在每一次试验中，事件 A 和 $\bar{A}$ 不会同时发生，并且一定有一个发生.

$$P(\bar{A})=1-P(A).$$

在计算随机事件 A 的概率时，常常把 A 化成几个彼此互斥的事件之和，再利用互斥事件的概率加法公式来计算；当事件 A 比较复杂而 $\bar{A}$ 比较简单时，往往通过计算 $\bar{A}$ 的概率 $P(\bar{A})$ 来求得 A 的概率 $P(A)$.

7. 虽然可以通过做大量重复试验，用随机事件发生的频率来估计其概率，但是，人工进行试验费时、费力，并且有时是不可能实现的. 因此，我们常常借助模拟方法来估计某些随机事件发生的概率. 用模拟方法可以在短时间内完成大量的重复试验. 对于某些无法确切知道概率的问题，模拟方法能帮助我们得到其概率的近似值. 模拟方法在实际中有很多应用.

二、学习要求和需要注意的问题

1. 学习要求

(1) 了解随机事件发生的不确定性和频率的稳定性，了解概率的意义以及频率与概率的区别.

(2) 理解古典概型的两个基本特征，掌握古典概型的概率计算公式，会用列举法计算一些随机事件所包含的基本事件数及事件发生的概率.

(3) 了解两个互斥事件的概率加法公式及对立事件的概率计算公式.

(4) 能够运用模拟方法估计概率，了解模拟方法估计概率的实际应用，初步体会几何概型的意义.

2. 需要注意的问题

(1) 学习概念时，应注意弄清一些容易混淆的概念(比如频率与概率、互斥事件与对立事件)的区别与联系.

(2) 建立古典概型时，一定要注意满足古典概型的两个基本特征. 比如一枚均匀的硬币连续抛掷 2 次，出现“2 次正面”“2 次反面”“1 次正面、1 次反面”的可能性是不相同的，因此，把这 3 个事件看成基本事件建立的模型就不是古典概型. 对于一个实际问题，应尝试建立不同的古典概型来解决.

(3) 在用列举法计算古典概型的基本事件数时，要学会用画树状图或者列表的方法.

(4) 对于古典概型的概率计算公式 $P(A)=\frac{m}{n}$，互斥事件的概率加法公式 $P(A+B)=P(A)+P(B)$ 及对立事件的概率计算公式 $P(\bar{A})=1-P(A)$，在运用时须注意它们的适用条件. 求随机事件 A 的概率时，常常把 A 化成几个彼此互斥的事件之和；当事件 A 比较复杂而 $\bar{A}$ 比较简单时，应先求 $P(\bar{A})$.

复习题三

A　组

1. 50 粒不同颜色的石子放入一个瓶子并且完全混合在一起，其中有 25 粒蓝色石子，20 粒绿色石子和 5 粒红色石子，如果闭上眼睛从瓶中任取一粒石子，计算下列事件的概率：
   (1) 取出的石子是红色石子；
   (2) 取出的石子不是绿色石子.
2. 有一个公共汽车站，2 路、3 路、6 路、9 路、13 路这五路车都在这个站停靠. 一人在此候车，他可乘坐 6 路或 13 路到达目的地. 假设每路汽车首先到站的可能性相同，计算下列事件的概率：
   (1) 首先到站的是 13 路车；
   (2) 首先到站的是能送他去目的地的汽车.
3. 一架客机的经济舱有 52 排座位. 每排座位有 7 个，被两条过道分开，在过道的两边各有 2 个座位，中间有 3 个座位. 除了靠近机翼的两排以外，其他排最外边的座位都是靠窗的座位.
   (1) 经济舱共有多少个座位？
   (2) 随机地选择一个座位，计算下列事件的概率：
      (i) 它是一个靠窗的座位；
      (ii) 它是一个靠过道的座位；
      (iii) 它既不是靠窗的座位也不是靠过道的座位.
4. 一粒均匀的骰子有 3 个面被涂上了紫色，2 个面被涂上了黑色，另一个面被涂上了橙色. 掷这粒骰子，计算下列事件的概率：
   (1) 向下的面是紫色；
   (2) 向下的面不是橙色；
   (3) 向上的面是黑色.
5. 随机地排列数字 1，5，6 得到一个三位数，计算下列事件的概率：
   (1) 所得的三位数大于 400；
   (2) 所得的三位数是偶数.
6. 从长度分别为 1 cm，3 cm，5 cm，7 cm，9 cm 的 5 条线段中，任意取出 3 条，计算取出的 3 条线段能构成三角形的概率.
7. 把 a，b 两个球随机地放入编号为 1，2，3 的 3 个盒子中，每个盒子中放的球数不限. 计算在 1，2 号盒子中各有 1 个球的概率是多少？
8. 抛掷一枚均匀的硬币 4 次，一定出现 2 次正面朝上和 2 次反面朝上吗？计算恰抛得 2 次正面朝上的概率.
9. 某中学高一年级共 4 个班，每班人数如下表：

班级	高一(1)班	高一(2)班	高一(3)班	高一(4)班
学生人数	45	42	44	43

随机地选取该校 1 位高一年级学生：

(1) 他是该校高一(1)班的学生的概率是多少？

(2) 他是该校高一(3)班或高一(4)班的学生的概率是多少？

(3) 他不是该校高一(2)班的学生的概率是多少？

10. 一个小组的 3 个学生在分发数学作业时，从他们 3 人的作业中各随机地取出了 1 份作业.

(1) 每个学生恰好拿到自己作业的概率是多少？

(2) 3 个学生不都拿到自己作业的概率是多少？

(3) 每个学生拿的都不是自己作业的概率是多少？

11. 口袋里装有红球、白球、黑球各 1 个，这 3 个球除颜色外完全相同，有放回地连续抽取 2 次，每次从中任意地取出 1 个球，计算下列事件的概率：

(1) 取出的球全是红球；

(2) 取出的球不全是红球；

(3) 取出的球中至少有 1 个是红球；

(4) 取出的球是同一颜色；

(5) 取出的球颜色不同.

B 组

1. 掷一对不同颜色的均匀的骰子，计算：

(1) 所得的点数中一个恰是另一个的 2 倍的概率；

(2) 两粒骰子向上的点数相同的概率；

(3) 所得的点数中一个是偶数，另一个是奇数的概率.

2. 将一部共四卷的文集任意地排放在书架的同一层上，计算：

(1) 第二卷在第四卷左边的概率是多少？

(2) 第二卷在第三卷左边，并且第三卷在第四卷左边的概率是多少？

3. 假设你班上每个人的生日在一年 365 天中的任何一天的可能性是相同的，从你班上随机选取一位你的同学，计算他与你同一天出生的概率.

4. 一份测试题包括 20 道选择题，每题有 4 个选项，并且只有一个选项是正确的. 如果一个学生对每一个问题都随机猜测一个答案，试估计他至少答对 12 道题的概率(利用随机数表产生随机数进行模拟，至少模拟 15 次来估计这一概率).

C 组

1. 把一个正方体的表面涂上红色，在它的长、宽、高上等距离地各切三刀，则大正方体被分割成了 64 个大小相等的小正方体，将这些小正方体均匀地搅混在一起. 如果你从这些小正方体中随意地取出 1 个，这个小正方体各个面都没有涂红色的概率是多少？

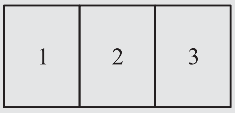
2. 一种电路控制器在出厂时每 4 件一等品装成一箱，工人在装箱时不小心把 2 件二等品和 2 件一等品装入了一箱，为了找出该箱中的二等品，对该箱中的产品逐件进行测试，计算：

（1）只测试 2 件就找到全部二等品的概率；

（2）测试的第 2 件产品是二等品的概率；

（3）恰好在前 3 次测试中测试到全部二等品的概率.

3. 如图，一面旗帜由 3 部分构成，这 3 部分必须分别涂上不同的颜色，现有红、黄、蓝、黑四种颜色可供选择，利用树状图列出所有可能结果，计算下列事件的概率：



（第 3 题）

（1）红色不被选中；

（2）红色和黑色被选中；

（3）第 1 部分是黑色并且第 2 部分是红色.

4. 一种桶装方便面正在搞促销活动，每一桶方便面里装有一份精美的纪念卡片，共有 6 种不同的纪念卡片. 假设每一种纪念卡片出现的机会是相等的，并且每一种纪念卡片的出现与其他卡片的出现是独立的，小明买了 10 桶方便面，他收集齐全部 6 种卡片的概率是多少？用随机数表产生随机数来模拟小明买 10 桶方便面收集卡片的过程，完成 20 次模拟. 根据模拟结果，对小明收集齐全部 6 种卡片的概率作出估计.

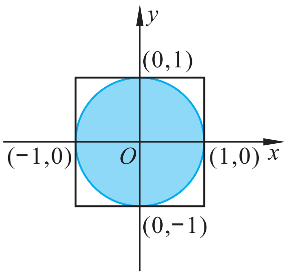
探究活动

用模拟方法估计圆周率 π 的值

一、问题情境和探究任务

问题提出

我国古代著名数学家祖冲之早在 1 500 多年前就算出圆周率 π 的值在 3.141 592 6 和 3.141 592 7 之间，这是我国古代数学的一大成就. 利用模拟方法，我们自己也可以对圆周率 π 的值作出估计.



如左图，一个单位圆内切于一个正方形，

$$\frac{\text{圆面积}}{\text{正方形面积}}=\frac{\pi}{4}.$$

任务 1 向正方形中随机地撒芝麻，数出落在圆内的芝麻数和落在正方形中的芝麻数，用芝麻落在圆内的频率来估计圆与正方形的面积之比(即$\frac{\pi}{4}$)，由此得出 π 的近似值.

任务 2 利用随机数表产生随机数来模拟向正方形中撒芝麻的试验. 用芝麻落在圆内的频率来估计圆与正方形的面积之比(即$\frac{\pi}{4}$)，由此得出 π 的近似值.

* **任务 3** 你还有没有其他的方法来进行模拟？若有，请完成模拟并估计出 π 的近似值.

任务 4 你还能用模拟方法解决其他的问题吗？提出你的问题，并给出模拟方案.

> **信息技术建议**
>
> 如果有条件的话，你可以利用计算器或计算机产生随机数来模拟向正方形中撒芝麻的试验.

二、实施建议

1. 可以组成小组进行探究、设计方案，自己独立试验，再与同学交流并汇总数据，给出对圆与正方形的面积之比(即$\frac{\pi}{4}$)的估计并求出 π 的近似值，完成每个人的“成果报告”.

2. 对完成任务 1 的建议

在一张白纸上画上一个正方形和它的内切圆，把纸铺在桌上，蒙住眼睛然后随机地向纸上撒芝麻，最后数出落在圆内的芝麻数和落

在正方形中的芝麻数.

3. 对完成任务 2 的建议

在随机数表中随机选择一个开始点，每次往后顺次选取 5 个数字，比如选取的是 2 5 6 8 8，则用它表示 0.256 88，产生两个这样的 0～1的随机数分别作为随机点的横坐标 x 和纵坐标 y，如果 $x^2+y^2\leqslant 1$，就表示随机点落在圆内，否则随机点就落在圆外，这样就完成了一次模拟. 大量重复进行模拟试验，算出芝麻落在圆内的频率，就可估计出圆与正方形的面积之比.

说 明

这里的做法相当于向上述正方形在第一象限的部分撒芝麻，而不是向该正方形撒芝麻. 这样做的结果是一样的.

4. “成果报告”的书写建议

成果报告可以用下表的形式呈现.

“用模拟方法估计圆周率 π 的值”探究学习成果报告表

______年级____班　　完成时间____________

<table>
<tr><td colspan="2">1.　课题组成员、分工、贡献</td></tr>
<tr><td>成员姓名</td><td>分工与主要工作或贡献</td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td colspan="2">2.　探究的过程和结果</td></tr>
<tr><td colspan="2">3.　参考文献</td></tr>
<tr><td colspan="2">4.　成果的自我评价：(请说明方法或原理的合理性、特色或创新点、不足之处等)</td></tr>
<tr><td colspan="2">5.　拓展(选做)：在解决问题的过程中发现和提出的新问题；可以延伸或拓展的内容；得到的新结果或猜想等</td></tr>
<tr><td colspan="2">6.　体会：描述在工作中的感受</td></tr>
</table>

5. 成果交流

建议以小组为单位，选出代表，在班级中报告研究成果，交流研究体会.

6. 评价建议

在评价中，采用自评、互评、教师评价相结合的形式，应善于发现别人工作中的特色，可主要考虑以下几个方面：

（1）求解过程和结果：合理、清楚、简捷、正确；

（2）独到的思考和发现；

（3）提出有价值的求解设计和有见地的新问题；

（4）发挥组员的特长，体现合作学习的效果.

附录 1

4 000 以下的素数表

2	179	419	661	947	1 229	1 523	1 823	2 131	2 437	2 749	3 083	3 433	3 733
3	181	421	673	953	1 231	1 531	1 831	2 137	2 441	2 753	3 089	3 449	3 739
5	191	431	677	967	1 237	1 543	1 847	2 141	2 447	2 767	3 109	3 457	3 761
7	193	433	683	971	1 249	1 549	1 861	2 143	2 459	2 777	3 119	3 461	3 767
11	197	439	691	977	1 259	1 553	1 867	2 153	2 467	2 789	3 121	3 463	3 769
13	199	443	701	983	1 277	1 559	1 871	2 161	2 473	2 791	3 137	3 467	3 779
17	211	449	709	991	1 279	1 567	1 873	2 179	2 477	2 797	3 163	3 469	3 793
19	223	457	719	997	1 283	1 571	1 877	2 203	2 503	2 801	3 167	3 491	3 797
23	227	461	727	1 009	1 289	1 579	1 879	2 207	2 521	2 803	3 169	3 499	3 803
29	229	463	733	1 013	1 291	1 583	1 889	2 213	2 531	2 819	3 181	3 511	3 821
31	233	467	739	1 019	1 297	1 597	1 901	2 221	2 539	2 833	3 187	3 517	3 823
37	239	479	743	1 021	1 301	1 601	1 907	2 237	2 543	2 837	3 191	3 527	3 833
41	241	487	751	1 031	1 303	1 607	1 913	2 239	2 549	2 843	3 203	3 529	3 847
43	251	491	757	1 033	1 307	1 609	1 931	2 243	2 551	2 851	3 209	3 533	3 851
47	257	499	761	1 039	1 319	1 613	1 933	2 251	2 557	2 857	3 217	3 539	3 853
53	263	503	769	1 049	1 321	1 619	1 949	2 267	2 579	2 861	3 221	3 541	3 863
59	269	509	773	1 051	1 327	1 621	1 951	2 269	2 591	2 879	3 229	3 547	3 877
61	271	521	787	1 061	1 361	1 627	1 973	2 273	2 593	2 887	3 251	3 557	3 881
67	277	523	797	1 063	1 367	1 637	1 979	2 281	2 609	2 897	3 253	3 559	3 889
71	281	541	809	1 069	1 373	1 657	1 987	2 287	2 617	2 903	3 257	3 571	3 907
73	283	547	811	1 087	1 381	1 663	1 993	2 293	2 621	2 909	3 259	3 581	3 911
79	293	557	821	1 091	1 399	1 667	1 997	2 297	2 633	2 917	3 271	3 583	3 917
83	307	563	823	1 093	1 409	1 669	1 999	2 309	2 647	2 927	3 299	3 593	3 919
89	311	569	827	1 097	1 423	1 693	2 003	2 311	2 657	2 939	3 301	3 607	3 923
97	313	571	829	1 103	1 427	1 697	2 011	2 333	2 659	2 953	3 307	3 613	3 929
101	317	577	839	1 109	1 429	1 699	2 017	2 339	2 663	2 957	3 313	3 617	3 931
103	331	587	853	1 117	1 433	1 709	2 027	2 341	2 671	2 963	3 319	3 623	3 943
107	337	593	857	1 123	1 439	1 721	2 029	2 347	2 677	2 969	3 323	3 631	3 947
109	347	599	859	1 129	1 447	1 723	2 039	2 351	2 683	2 971	3 329	3 637	3 967
113	349	601	863	1 151	1 451	1 733	2 053	2 357	2 687	2 999	3 331	3 643	3 989
127	353	607	877	1 153	1 453	1 741	2 063	2 371	2 689	3 001	3 343	3 659	
131	359	613	881	1 163	1 459	1 747	2 069	2 377	2 693	3 011	3 347	3 671	
137	367	617	883	1 171	1 471	1 753	2 081	2 381	2 699	3 019	3 359	3 673	
139	373	619	887	1 181	1 481	1 759	2 083	2 383	2 707	3 023	3 361	3 677	
149	379	631	907	1 187	1 483	1 777	2 087	2 389	2 711	3 037	3 371	3 691	
151	383	641	911	1 193	1 487	1 783	2 089	2 393	2 713	3 041	3 373	3 697	
157	389	643	919	1 201	1 489	1 787	2 099	2 399	2 719	3 049	3 389	3 701	
163	397	647	929	1 213	1 493	1 789	2 111	2 411	2 729	3 061	3 391	3 709	
167	401	653	937	1 217	1 499	1 801	2 113	2 417	2 731	3 067	3 407	3 719	
173	409	659	941	1 223	1 511	1 811	2 129	2 423	2 741	3 079	3 413	3 727	

附录 2

随机数表

10	09	78	25	33	76	52	01	35	86	34	67	35	48	76	80	95	90	91	17	39	29	27	49	45
37	54	20	48	05	64	89	47	42	96	24	80	52	40	37	20	63	61	04	02	00	82	29	16	65
08	42	26	89	53	19	64	50	93	03	23	20	90	25	60	15	95	33	47	64	35	08	03	36	06
99	01	90	25	29	09	37	67	07	15	38	31	13	11	65	88	67	67	43	97	04	43	52	76	59
02	80	79	99	70	80	15	73	61	47	64	03	23	66	53	98	95	11	68	77	12	17	17	68	33
66	06	57	47	17	34	07	27	68	50	36	69	73	61	70	65	81	33	98	85	11	19	92	91	70
81	05	01	08	05	45	57	18	24	05	35	30	34	28	14	88	79	90	74	39	23	40	30	97	32
83	26	97	76	02	02	05	16	56	92	68	55	57	48	18	73	05	38	52	47	18	62	38	85	79
63	57	33	21	35	05	32	54	70	48	90	55	85	75	18	28	46	82	87	09	83	40	12	56	24
73	79	64	57	53	03	52	96	47	78	35	80	83	42	82	60	93	52	03	44	35	27	38	84	35
08	52	01	77	67	14	00	56	86	07	22	10	94	05	58	60	97	00	34	33	50	50	07	36	98
11	60	50	54	31	39	80	82	77	32	50	72	36	82	48	29	40	52	42	01	52	77	56	78	51
83	45	39	96	34	06	28	89	80	83	13	74	57	00	78	18	47	54	06	10	68	71	17	78	17
88	68	54	02	00	86	50	75	84	01	36	76	55	79	51	90	36	47	64	93	29	60	91	10	62
99	59	46	73	48	87	51	76	49	69	91	82	50	89	28	93	78	55	13	68	23	47	83	41	13
65	48	11	76	74	17	46	85	09	50	58	04	77	69	74	73	03	95	71	86	40	21	81	65	44
80	12	43	56	35	17	72	70	80	15	45	31	82	23	74	21	11	57	82	63	14	38	55	37	63
74	35	00	98	17	77	40	27	72	14	43	23	60	02	10	45	52	16	42	37	95	28	60	26	55
69	91	62	68	03	66	25	22	91	48	35	93	68	72	03	76	62	11	39	90	94	40	05	64	18
09	89	32	05	06	14	22	56	85	14	46	42	75	67	88	96	29	77	83	22	54	38	21	45	98
91	49	91	45	23	68	47	92	76	86	46	16	28	35	54	94	75	08	99	23	37	08	92	00	48
80	33	69	45	98	26	94	03	68	58	70	29	73	41	35	53	14	03	38	40	42	05	08	23	41
44	10	48	19	49	85	15	74	79	64	32	97	92	65	75	57	60	04	08	81	22	22	20	64	13
12	55	07	37	42	11	10	00	20	40	12	56	07	46	97	96	64	48	94	39	28	70	72	58	15
63	60	64	93	29	16	50	53	44	84	40	21	95	85	68	43	65	17	70	82	07	20	73	17	90
61	19	69	04	46	26	45	74	77	74	51	92	43	37	29	65	30	45	95	93	42	58	26	05	27
15	47	44	52	66	95	27	07	99	53	59	36	78	38	48	82	39	61	01	18	33	21	15	94	66
94	55	72	85	73	67	89	75	43	87	54	62	24	44	31	91	19	04	25	92	92	92	74	59	73
42	48	11	62	13	97	34	40	87	21	16	86	84	87	67	03	07	11	80	59	25	70	14	66	70
23	52	37	83	17	73	80	88	98	37	68	93	59	14	16	26	25	22	96	63	05	52	28	25	62
04	49	35	24	94	75	24	63	38	24	45	86	25	10	25	64	96	27	93	35	65	33	71	24	72
00	54	99	76	54	54	05	18	81	59	96	11	96	38	96	54	69	28	23	91	23	28	72	95	29
35	96	31	53	07	26	89	80	93	54	33	35	13	54	62	77	97	45	00	24	90	10	33	93	33
59	80	80	83	91	45	42	72	68	42	83	60	94	97	00	13	02	12	48	92	78	56	62	01	06
46	05	88	52	35	01	39	09	22	86	77	28	14	40	77	93	91	08	36	47	70	61	74	29	41
32	17	90	05	97	87	37	92	52	41	05	56	70	70	07	86	74	31	71	57	85	39	41	18	38
69	23	46	14	06	20	11	74	52	04	16	96	66	00	00	18	74	39	84	28	97	11	89	63	38
19	56	54	14	30	01	75	87	53	79	40	41	92	15	85	55	67	43	68	05	84	96	28	52	07
45	15	51	49	38	19	47	60	72	46	43	66	79	45	43	59	04	79	00	33	20	82	66	95	41
94	86	43	19	94	36	16	81	08	51	34	88	88	15	53	01	54	03	54	56	05	01	45	11	76

98	08	62	48	26	45	24	02	84	04	44	99	90	88	96	39	09	47	34	07	35	44	13	18	80
33	18	51	62	32	41	94	15	09	49	89	43	54	85	81	88	69	54	19	94	37	54	87	30	48
80	95	10	04	06	96	38	27	07	74	20	15	12	33	87	25	01	62	52	98	94	62	46	11	71
79	75	24	91	40	71	96	12	82	96	69	86	10	25	91	74	85	22	05	39	00	38	75	95	79
18	63	33	25	37	98	14	50	65	71	31	01	02	46	74	05	45	56	14	27	77	93	80	19	36
74	02	94	39	02	77	55	73	22	70	97	79	01	71	19	52	52	75	80	21	80	81	45	17	48
54	17	84	56	11	80	99	33	71	43	05	33	51	29	69	56	12	71	92	56	36	04	09	08	24
11	66	44	98	83	52	07	98	48	27	59	38	17	15	39	09	97	33	84	40	88	46	12	33	56
48	32	47	79	28	31	24	96	47	10	02	29	53	68	70	32	30	75	75	46	15	02	00	99	94
69	07	49	41	38	87	68	79	19	76	35	58	40	44	01	10	51	82	16	15	01	84	87	69	38
09	18	82	00	97	32	82	53	95	27	04	22	08	63	04	83	38	98	73	74	64	27	85	80	44
90	04	58	54	97	51	98	15	06	54	94	98	88	19	97	91	87	07	61	50	68	47	66	46	59
73	18	95	02	07	47	67	72	52	69	62	29	06	44	64	27	12	46	70	18	41	36	18	27	60
75	76	87	64	90	20	97	18	17	49	90	42	91	22	72	95	37	50	58	71	93	82	34	31	78
54	01	64	40	56	65	28	13	10	03	00	55	22	73	98	20	71	45	32	95	07	70	61	78	13
08	35	85	99	10	78	64	24	27	85	13	65	15	88	73	04	61	89	75	53	31	22	30	84	80
28	30	60	82	64	81	33	31	05	91	40	51	00	78	93	32	60	46	04	75	94	11	90	18	40
53	84	08	62	33	81	59	41	36	28	51	21	59	02	90	28	46	66	87	96	77	76	22	07	91
91	75	75	37	41	61	61	35	22	69	50	26	39	02	12	55	78	17	65	14	83	48	34	70	55
89	41	59	26	94	00	39	75	83	91	12	60	71	75	46	48	94	97	23	06	94	54	13	74	08
77	51	30	38	20	85	83	42	99	01	58	41	48	27	74	51	90	81	39	80	72	89	35	55	07
19	50	23	71	74	69	97	92	02	88	55	21	02	97	73	74	28	77	52	51	65	34	46	74	15
21	81	85	93	13	93	27	88	17	57	05	58	67	31	56	07	08	28	50	46	31	85	33	84	52
51	47	46	54	99	68	10	72	36	21	94	04	99	18	45	42	83	60	91	91	08	00	74	54	49
99	55	96	83	31	62	53	52	41	70	69	77	71	28	30	74	81	97	81	42	43	86	07	28	34
83	71	34	80	07	93	58	47	28	69	51	92	66	47	21	58	30	32	98	22	93	17	49	39	72
85	27	48	68	93	11	30	32	92	70	28	83	43	41	37	73	51	59	04	00	71	14	84	36	43
84	13	38	96	40	44	03	55	21	66	73	85	27	00	91	61	22	26	05	61	62	32	71	84	23
56	73	21	62	34	17	39	59	61	31	10	12	39	16	22	85	49	65	75	60	81	60	41	88	80
65	13	85	68	06	87	64	88	52	61	34	31	38	58	61	45	87	52	10	69	85	64	44	72	77
38	00	10	21	76	81	71	91	17	11	71	60	29	29	37	74	21	96	40	49	55	58	44	96	98
37	40	29	63	97	01	30	47	75	86	56	27	11	00	86	47	32	46	26	03	40	03	03	74	38
97	12	54	03	48	87	08	33	14	17	21	81	53	92	50	75	23	76	20	47	15	50	12	95	78
21	82	84	11	34	47	14	33	40	72	64	63	88	59	02	49	13	90	64	41	03	85	65	45	52
73	13	54	27	42	95	71	90	90	35	85	79	47	42	96	08	78	98	81	56	64	59	11	92	02
07	63	87	79	29	03	06	11	80	72	96	20	74	41	56	23	82	19	95	38	04	74	36	69	91
60	52	88	34	41	07	95	41	98	14	59	17	52	06	95	05	53	35	21	39	61	21	20	64	55
83	59	63	56	55	06	95	89	29	83	05	12	80	97	19	77	43	35	37	83	92	30	15	04	98
10	85	06	27	46	98	59	91	05	07	13	49	90	63	19	53	07	57	18	39	06	41	01	93	62
89	82	09	89	52	48	52	26	31	47	64	42	18	08	14	48	80	00	93	51	31	02	47	31	67

59	58	00	64	78	75	56	97	88	00	88	83	55	44	86	23	76	80	61	56	04	11	10	84	08
38	50	80	73	41	23	79	34	87	63	80	82	29	70	22	17	71	90	42	07	95	95	44	99	53
30	69	27	06	68	94	68	81	61	27	56	19	68	00	91	82	06	76	34	00	05	45	26	92	00
65	44	39	56	59	18	28	82	74	37	49	63	22	40	41	08	33	76	56	76	96	29	99	08	36
27	26	75	02	64	13	19	27	22	94	07	47	74	46	06	17	98	54	89	11	97	34	13	03	58
91	30	70	89	91	19	07	22	42	10	36	69	95	37	28	28	82	53	57	93	28	97	66	62	52
68	43	49	46	88	84	47	31	35	22	62	12	69	84	08	12	84	38	25	90	09	81	59	31	46
48	90	81	58	77	54	74	52	45	91	35	70	00	47	51	83	82	45	26	92	54	13	05	51	60
06	91	34	51	97	42	67	27	85	01	11	88	00	95	28	63	01	19	89	01	14	97	44	03	44
10	45	51	60	19	14	21	03	37	12	91	34	23	78	21	88	32	58	08	51	43	66	77	08	83
12	88	39	73	43	65	02	76	11	84	04	28	50	13	92	17	97	41	50	77	90	71	22	67	69
21	77	83	00	76	38	80	73	69	61	31	64	94	20	96	63	28	10	20	23	08	81	64	74	49
19	52	35	95	15	65	12	25	96	59	86	28	36	82	58	59	57	21	37	98	16	43	59	15	29
67	24	55	26	70	35	58	31	65	63	79	24	68	66	86	76	46	33	42	22	26	65	59	08	02
60	58	44	73	77	07	50	03	79	92	45	13	42	65	29	26	76	08	36	37	41	32	64	43	44
53	85	34	13	77	36	06	69	48	50	58	83	87	38	59	49	36	47	33	31	96	24	04	36	42
24	63	73	87	86	74	38	48	93	42	52	62	30	79	92	12	36	91	86	01	03	74	28	38	73
83	08	01	24	51	38	99	22	28	15	07	75	95	17	77	97	37	72	75	85	51	97	23	78	67
16	44	42	43	34	36	15	19	90	73	27	49	37	09	39	85	13	03	25	52	54	84	65	47	59
60	79	01	81	57	57	17	86	57	62	11	16	17	85	76	45	81	95	29	79	85	13	00	48	60
03	99	11	04	61	93	71	61	68	91	56	08	32	46	53	84	60	95	82	32	88	61	81	91	61
38	55	59	55	54	32	88	65	97	80	08	35	56	08	50	29	73	54	77	62	71	20	92	38	53
17	54	67	37	04	92	05	24	52	15	55	12	12	92	81	59	07	60	79	36	27	95	45	89	09
32	64	35	28	61	95	81	90	68	31	00	51	19	89	36	76	35	59	37	79	80	86	30	05	14
69	57	26	87	77	39	51	03	59	05	14	06	04	06	19	29	54	96	96	18	33	56	46	07	80
24	12	26	65	91	27	69	90	64	94	14	84	54	66	72	61	95	87	71	00	90	89	97	57	54
61	15	63	02	31	92	96	26	17	33	41	83	95	53	82	17	26	77	09	43	78	03	87	02	67
30	53	22	17	04	10	27	41	22	02	39	68	52	83	09	10	06	16	88	29	55	98	56	64	85
03	78	89	35	99	75	86	72	07	17	74	41	65	31	56	35	20	83	33	74	87	53	90	88	23
48	22	86	33	79	85	78	34	76	19	53	15	26	74	33	35	66	35	29	72	16	81	86	03	11
60	36	59	46	53	35	07	53	39	49	42	61	42	92	97	01	91	82	83	16	98	95	37	32	31
83	79	94	24	02	56	62	33	44	42	34	99	44	13	74	70	07	11	47	36	09	95	81	80	65
32	96	00	74	05	36	40	98	32	32	99	38	54	16	00	11	13	30	75	86	15	91	70	62	53
19	32	25	38	45	57	62	05	26	05	66	49	76	86	46	78	13	85	65	59	19	64	09	94	13
11	22	09	47	47	07	39	93	74	08	48	50	92	39	29	27	48	24	54	76	85	24	43	51	59
31	75	15	72	60	68	98	00	53	39	15	47	04	83	55	88	65	12	25	96	03	15	21	52	21
88	49	29	93	82	14	45	40	45	04	20	09	49	89	77	74	84	39	34	13	22	10	97	85	08
30	93	44	77	41	07	48	18	38	28	73	78	80	65	33	28	59	72	04	05	94	20	52	03	80
22	88	84	88	93	27	49	99	87	48	60	53	04	51	28	74	02	28	46	17	82	03	71	02	68
78	21	21	69	93	35	90	29	13	86	44	37	21	54	86	65	74	11	40	14	87	48	13	72	80

附录 3

上机实现参考程序

1. 二分法求方程 $x^3-x-1=0$ 的近似解,精度为 0.000 01

```
Private Function f(ByVal x As Single)As Single
    f=x*x*x-x-1
End Function

Private Sub Command1_Click()
    a=1
    b=1.5
    p=0.00001
    Do
        Midl=(a+b)/2
        If f(Midl)=0 Then
            Exit Do
        Else
            If f(Midl)>0 Then
                b=Midl
            Else
                a=Midl
            End If
        End If
    Loop While b-a>p
    Labell.Caption="方程的根是"&(a+b)/2
End Sub
```

2. 求 20 个数中的最大数

```
max1=Val(InputBox("请输入第 1 个数"))
For i=2 To 20
    n=Val(InputBox("请输入第"&i&"个数"))
    If n>max1 Then max1=n
Next
Labell.Caption="你刚才输入的 20 个数中最大的一个是"& max1
```

3. 判断给定的数是否是素数

```
p=Int(Val(InputBox("请输入一个整数")))
pri=Abs(p)
For i=2 To Sqr(pri)
    If pri Mod i=0 Then Exit For
Next
If i<=Sqr(pri)  Then
    Labell.Caption=p&"不是素数"
Else
    Labell.Caption=p&"是素数"
End If
```

4. 输出 1～200 的所有素数

```
Private Function prime(ByVal p As Integer) As Boolean
    pri=Abs(p)
    For i=2 To Sqr(pri)
        If pri Mod i=0 Then Exit For
    Next
    If i<=Sqr(pri)  Then
        prime=False
    Else
        prime=True
    End If
End Function

Private Sub Command1_Click()
    Label1.Caption="1 到 200 的素数为："& vbCrLf
    For i=2 To 200
        If prime(i)Then Label1.Caption=Label1.Caption& ""& i
    Next
End Sub
```


5. 百鸡问题算法

```
Label1.Caption=""
For a=0 To 20
    For b=0 To 33
        c=100-a-b
        If 15*a+9*b+c=300 Then
            Label1.Caption=Label1.Caption&"公鸡母鸡小鸡各"&a&","&b&","
                        &c&"只"&vbCrLf
        End If
    Next
Next
```

6. 辗转相除求最大公因数

```
a=InputBox("请输入一个整数")
b=InputBox("请输入另一个整数")
n1=Abs(a)
n2=Abs(b)
If n1>n2 Then
    tt=n1
    n1=n2
    n2=tt
End If
Do While n2 Mod n1<>0
    tt=n2 Mod n1
    n2=n1
    n1=tt
Loop
Labell.Caption=a&"和"&b&"的最大公因数为"&n1
```

附录 4

部分数学专业词汇中英文对照表

中文	英文
决策	decision
调查问卷	questionnaire
数学期望	mathematical expectation
统计	statistics
概率	probability
数据	data
随机现象	random phenomenon
抽样	sampling
估计	estimation
线性回归	linear regression
随机抽样	random sampling
分层抽样	layering sampling
系统抽样	system sampling
总体分布	population distribution
频率	frequency
频率分布	frequency distribution
频率直方图	frequency histogram
频率折线图	frequency polygon
茎叶图	stem and leaf diagram
散点图	scatter plot diagram
线性回归方程	linear regression equation
最小二乘法	method of least squares
统计推断	statistical inference
总体	population
样本	sample
算法	algorithm
程序	program
顺序结构	proper order construction
条件结构	conditional construction
输入语句	input sentence
输出语句	output sentence
循环语句	for statement

中文	英文
复杂性	complexity
标准差	standard deviation
方差	variance
平均数	mean
互斥事件	exclusive events
古典概率	classical probability
随机数	random number
模拟方法	simulated method

附录 5

信息检索网址导引

基础教育教材网

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

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

后　记

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

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

本套教材的编委会组成如下(按姓氏笔画排序)：

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

参加本册教材编写的还有(按姓氏笔画排序)：

王松浦、白永潇、张丹、张敏、张饴慈.

由于时间仓促，教材中的错误在所难免，恳请广大使用者批评指正.

北京师范大学出版社