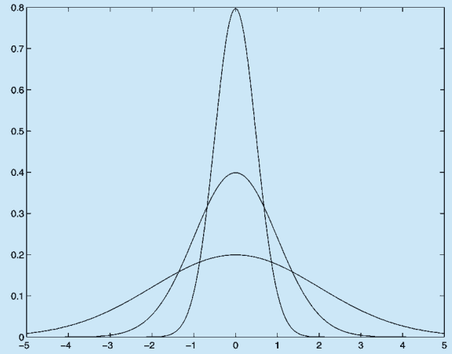


ISBN 978-7-5355-4601-2
9 787535 546012 >
定价：8.55元

绿色印刷产品

普通高中课程标准实验教科书

数学 第五册（必修）

湖南教育出版社

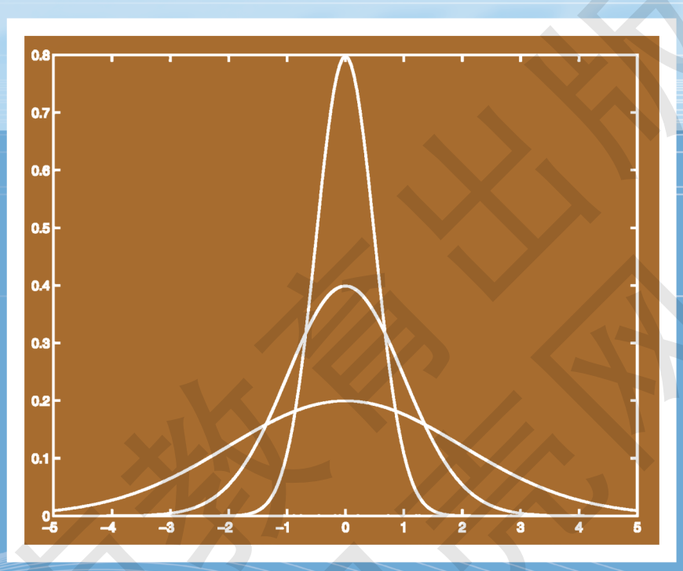
经全国中小学教材审定委员会 2005 年初审通过

Mathematics

普通高中课程标准实验教科书

数学

第五册（必修）



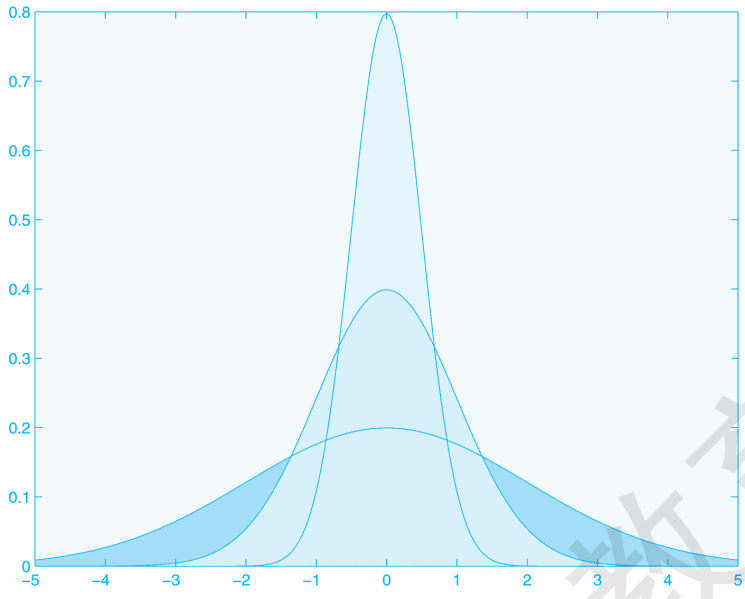
湖南教育出版社

Mathematics

普通高中课程标准实验教科书

数学

第五册（必修）



湖南教育出版社

主　　编　张景中　黄楚芳

执行主编　李尚志
编　　委　何书元　任宏硕　郑志明
　　文志英　王长平　李超贵

信息技术使数学更有力量

在这一册里，我们将学习有关算法、统计和概率的初步知识.

用各种各样的数学知识解决实际问题和各学科的理论问题，许多情形需要算出结果，需要有一套可以具体操作的办法，这就要有算法. 算法是数学及其应用的重要组成部分，是计算科学的重要基础. 在现代信息技术飞速发展的今天，算法在科学技术和社会发展中发挥着越来越大的作用，并且日益融入社会生活的许多方面.

在中国古代数学中，蕴含了丰富的算法思想. 针对社会生活中出现的数学问题，分门别类，提出有效的机械化的解答方法，即算法，正是中国古代数学的特色. 以几何学为主体的西方古代数学，主要特色虽然是公理化方法而不突出算法，但也出现了用算法解决问题的著名范例，即求最大公约数的欧几里得算法. 由于现代信息技术的发展，算法的研究和应用已经渗透到数学的各个分支. 算法思想已经成为现代人应当具备的一种数学素养.

同学们已经学过的许多解决数学问题的方法，像加减乘除，解方程或方程组，数学表达式求值，解三角形，等等，其实都是算法. 通过十年的学习，虽然还没有引进算法这个词，但可以初步体会和感受算法

的思想．在此基础之上，我们将结合对具体数学实例的分析，了解算法的概念和结构，体验算法的程序框图在解决问题中的作用；通过模仿、操作、探索，学习设计程序框图表达解决问题的过程，学习根据程序框图写出算法语句的基本方法；在模拟解决实际问题的过程中，体会算法的基本思想和有效性，发展有条理的思考和表达能力，提高逻辑思维的能力．

现代社会是信息化的社会，人们常常需要收集数据，从所获得的数据中提取有价值的信息，以作出合理的决策．如何从庞杂多变的社会万象中收集有用的数据？如何将大量的数据整理得井井有条？如何从整理过的数据中挖掘出宝贵的信息？这需要下一番去粗取精、去伪存真、由表及里、由此及彼的功夫，需要科学理论和方法的指导．统计正是研究如何合理地收集、整理和分析数据的学科，它可以为人们制定决策提供依据．

自然界和社会生活中，有些事情的发生和发展有着清楚的因果关系，但也有大量的现象表现出偶然性，即随机现象．尤其是在日常的社会生活中，随机现象几乎处处可见．人们收集到的许多数据，也不可避免地具有随机性．概率是研究随机现象规律的学科，它为我们认识客观世界提供了重要的思维模式和解决问题的方法，同时也为统计学的发展提供了理论基础．因此，统计与概率的基础知识已经成为现代社会公民的必备知识．

从小学到初中的数学课里，同学们已经知道一些

统计和概率的有关问题和方法．在这一阶段的学习中，我们将结合更多的实际问题情景，学习随机抽样、样本估计总体和线性回归的基本方法；体会用样本估计总体及其特征的思想．通过解决实际问题，较为系统地经历数据收集与处理的全过程，体会统计思维与确定性思维的差异．我们也将结合具体的实例，学习概率的某些基本性质和简单的概率模型，加深对随机现象的理解．我们还将动手动脑，通过有趣的实验、计算机或计算器的模拟，估计简单随机事件发生的概率．

用了计算机，可以更便捷地处理统计数据，模拟随机现象，以及执行形形色色的算法．希望大家尽可能地结合本册中的内容，学习信息技术有关的知识和操作．在实际操作中，能够更切实地感受算法、统计和概率的思想．

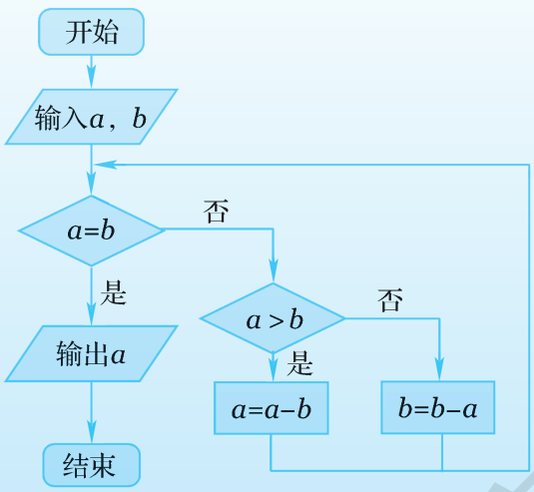
祝同学们在新的学期里学得好，玩得好！

目录

第 13 章　概率

第11章 算法初步

有规有矩成方圆，
步步为营释谜团.
亡羊歧路有判断，
戏马平川巧循环.
运筹帷幄操胜算，
袖里乾坤说大千.
欧公辗转堪垂范，
华夏九章更先鞭.




```
INPUT a,b
WHILE a<>b
    IF a>b THEN
        a=a-b
    ELSE
        b=b-a
    END IF
WEND
PRINT a
END
```

11.1 算法的概念

当今社会，计算机已广泛应用于生产、科研、生活的各个领域. 要让计算机做原来由人工做的繁复的重复性的工作，就必须使数学算法化，并用计算机能够接受的“语言”准确描述出来，这就凸显了古老的算法思想在当今信息时代的重要价值.

在数学中，算法（algorithm）通常是指由有限多个步骤组成的求解某一类问题的通用的方法，对于该类问题中的每个给定的具体问题，机械地执行这些步骤就可以得到问题的解答. 这就使得计算不仅可以由人完成，而且可以由计算机来代替.

其实算法对于我们并不陌生. 例如计算 $1+(5-3)\times 4$，第一步计算 $5-3=2$，第二步计算 $2\times 4=8$，第三步计算 $1+8=9$. 又例如解方程 $x^2+2x-3=0$，第一步计算 $b^2-4ac=2^2-4\times 1\times(-3)=16$，第二步计算 16 的算术平方根 $\sqrt{16}=4$，第三步计算 $x_1=\dfrac{-2+4}{2}=1$，第四步计算 $x_2=\dfrac{-2-4}{2}=-3$.

这就是说，对于这类问题，我们已经掌握了它的算法.

一般地，对一个问题的算法就是解决该问题的程序步骤的概要说明，它有以下三个特点：

这一程序步骤必须是确定的——各步骤的本质与次序被明确清楚地加以描述；

这一程序步骤必须是有效的——按此程序步骤最后必然能得到这一问题正确的解；

这一程序步骤必须是有限的——该程序在有限步之后终止.

例 1 设计一个算法，求 35 的大于 1 的最小约数.

根据最小约数的定义，依次用 2～35 去除 35，第一个能整除 35 的就是 35 的大于 1 的最小约数.

根据以上思路，可以写出以下算法：

S1：用 2 除 35，得到余数 1，2 不是 35 的约数；

S2：用 3 除 35，得到余数 2，3 不是 35 的约数；

S3：用 4 除 35，得到余数 3，4 不是 35 的约数；

S4：用 5 除 35，得到余数 0，5 是 35 的约数，因此，35 的大于 1 的最小约数是 5.

这里“S1”是“Step 1”的缩写，意即“第一步”.

一般地，你能写出“求正整数 n（$n>2$）的大于 1 的最小约数”的算法吗？

例 2 求 98 和 63 的最大公约数.

解 操作步骤如下：

步骤	a	b
1	98	63
2	35(=98−63)	63
3	35	28(=63−35)
4	7=(35−28)	28
5	7	21(=28−7)
6	7	14(=21−7)
7	7	7(=14−7)

通过上述 7 个步骤，就得到 7 是 98 和 63 的最大公约数.

虽然我们一眼能看出这两个数的最大公约数，但有许多情况是无法一眼看出的，因此，这里介绍在我国古代数学名著《九章算术》中介绍的“更相减损术”.

其操作方法可以归纳为：

大数减小数，

用差替大数；

依此反复做，

直到相等数.

你知道为什么到最后一步两个数相等时，就是所要求的最大公约数吗？

上面的步骤条理清晰、步骤明确、可操作性强，是一种机械化的程式，它体现出算法的特征. 利用它可以编制计算机程序，让计算机解决求两个正整数的最大公约数的这一类问题.

例 3 求 7 267 和 6 192 的最大公约数.

解 操作步骤如下：

步骤	a	b
1	7 267	6 192
2	1 075(=7 267−6 192)	6 192
3	1 075	5 117(=6 192−1 075)
4	1 075	4 042(=5 117−1 075)
5	1 075	2 967(=4 042−1 075)
6	1 075	1 892(=2 967−1 075)
7	1 075	817(=1 892−1 075)
8	258(=1 075−817)	817

用更相减损术求两个正整数的最大公约数的时候，过程大多数情况下会比较繁琐，为了解决这个问题，我们将在后面学习另一种求两个正整数的最大公约数的方法——辗转相除法.

步骤	a	b
9	258	559(＝817－258)
10	258	301(＝559－258)
11	258	43(＝301－258)
12	215(＝258－43)	43
13	172(＝215－43)	43
14	129(＝172－43)	43
15	86(＝129－43)	43
16	43(＝86－43)	43

1a41.eps

通过上述 16 个步骤，最后得到的 43 就是所要求的 7 267 和 6 192 的最大公约数.

习题 1

学而时习之

1. 设计一个算法，判断 7 是否为质数.

2. 利用“更相减损术”求 147 和 273 的最大公约数.

温故而知新

3. 下面给出一个问题的算法：

S1：输入非负实数 x；

S2：若 $x\geqslant 1$，则执行第三步，否则执行第四步；

S3：计算并输出 $2x$ 的值；

S4：计算并输出 x^2+1 的值.

当输入的 x 的值是多少时，输出的数值最小？

11.2 算法结构与程序框图

通过 11.1 节的学习我们知道，算法步骤有明确的顺序性. 算法可以用类似于 11.1 节例 1 和例 2 那样用自然语言描述，这样我们就知道哪些步骤在怎样的条件下执行，哪些步骤需要重复执行. 为了清晰、直观地描述设计好的算法，通常采用画图的办法，也就是用所谓程序框图（或流程图）的方法来表示.

如 11.1 节例 1 中“求 35 的大于 1 的最小约数”的算法可以用图 11－1 所示的程序框图表示.

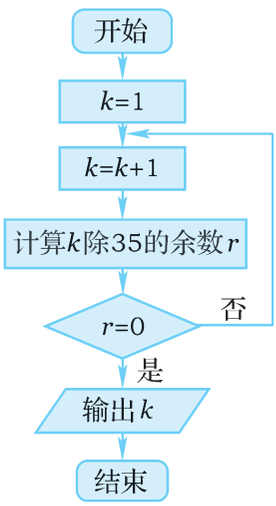


图 11－1

例 2 中利用更相减损术求两个正整数 a 和 b 的最大公约数的算法，就可以用图 11－2 所示的程序框图表示.

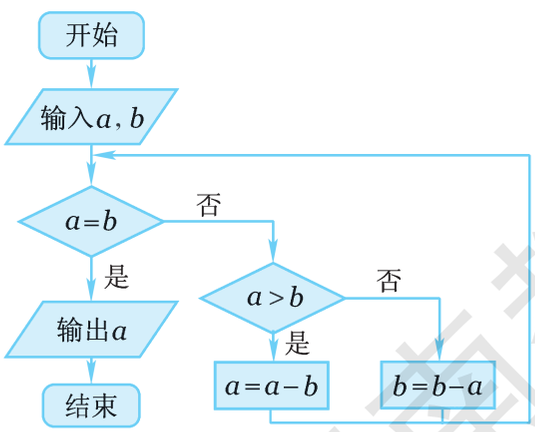


图 11－2

程序框图中的第一个和最后一个符号是终端框，它是任何程序框图都不可缺少的，分别表示一个算法的开始和结束. 大多数框图符号只有一个进入点和一个退出点，唯有判断框是有超出一个退出点的符号.

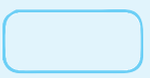
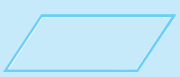
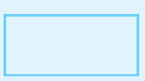
说一说，用这种程序框图描述算法有哪些优势？

你能对照 11.1 节例题中对算法的自然语言描述看懂这两张图吗？

在图 11－2 中，除了用圆角矩形表示“开始”和“结束”的终端框（起止框），还有输入、输出框，处理框（执行框），判断框. 一个或几个程序框的组合表示算法中的一个步骤，带有方向箭头的流程线将程序框连接起来，表示算法步骤的执行顺序.

由此可见，程序框图是一种用程序框、流程线以及文字符号说明等基本元件的组合来表示算法的图形.

下表列举出了程序框图中普遍采用的几个基本元件和它们表示的功能.

名　称	图　形	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		数据的输入或者结果的输出
处理框（执行框）		赋值、计算，传送结果
判断框（选择框）		根据给定条件判断，成立时出口为“是”，否则为“否”
流程线		连接程序框，表示流程方向
连接点		连接需分页的程序框图的两部分

用程序框图表示算法时，算法的逻辑结构呈现得非常清楚，尽管算法千差万别，但都可以由顺序结构、条件结构、循环结构这三种基本逻辑结构通过组合和嵌套表达出来.

11.2.1 顺序结构

依次进行多个处理步骤的结构称为顺序结构（sequence structure），它是一种最简单、任何算法都不可缺少的基本结构.

顺序结构的一般形式可以用程序框图图 11－3 表示，其中 A，B 是两个依次执行的步骤.

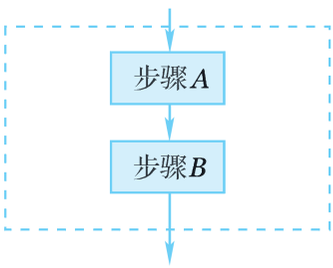


图 11－3

例 1 已知坐标平面内两点的坐标 $A(x_1, y_1)$，$B(x_2, y_2)$，利用中点坐标公式设计一个求 AB 的中点 P 的坐标的算法，并画出程序框图.

算法分析：

利用中点坐标公式 $x_0=\dfrac{x_1+x_2}{2}$，$y_0=\dfrac{y_1+y_2}{2}$ 分别计算出 x_0 和 y_0，输出结果即可，因此只需用顺序结构表达算法.

算法步骤如下：

S1：输入 A，B 两点的横、纵坐标 x_1，y_1 和 x_2，y_2；

S2：计算 $x_0=\dfrac{x_1+x_2}{2}$；

S3：计算 $y_0=\dfrac{y_1+y_2}{2}$；

S4：输出中点坐标 $P(x_0, y_0)$.

程序框图如图 11－4 所示.

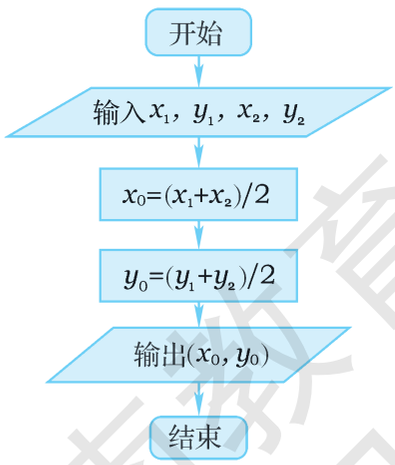


图 11－4

画程序框图时，为便于交流，必须遵守一些共同的规则：一是使用标准的框图符号；二是按从上到下、从左到右的方向画；三是图形符号内的描述语言要简练清楚.

例 2 “鸡兔同笼”是我国魏晋南北朝时期的数学著作《孙子算经》中的一个有趣而富有深远影响的问题：“今有雉兔同笼，上有三十五头，下有九十四足，问雉兔各几何？”你能设计一个算法解决这个问题吗？

算法分析：

用方程组的思想不难解决这个问题. 设有 x 只鸡，y 只兔，则有

$$\begin{cases} x+y=35, & ① \\ 2x+4y=94, & ② \end{cases}$$

算法步骤如下：

S1：②－①×2 得 $2y=24$；

S2：解 $2y=24$ 得 $y=12$；

S3：将 $y=12$ 代入①，解得 $x=23$.

程序框图如图 11－5 所示.

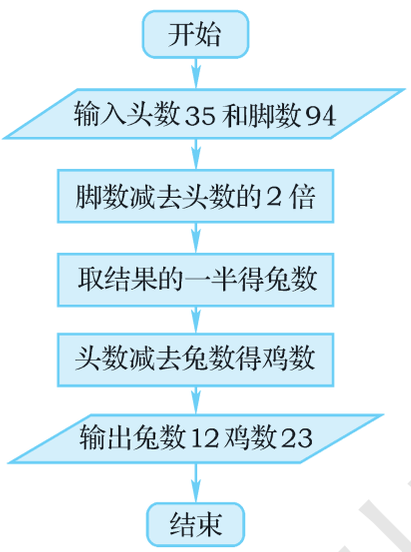


图 11－5

这也是一个顺序结构，解题的过程是自然地由上而下依次执行，虽然是一种最简单的算法过程，但是它对所有的鸡兔同笼问题都适用.

从上面的例子可以看到，与用自然语言描述算法相比，用框图表示算法更直观、形象，逻辑结构展现得非常清楚，容易理解.

仿照上面的步骤可以构建解二元一次方程组 $\begin{cases} a_1x+b_1y=c_1, \\ a_2x+b_2y=c_2 \end{cases}$

“鸡兔同笼”问题都可以用下面的算法解决：

S1：计算脚数减去头数的 2 倍，再取其一半得兔数；

S2：将头数减去兔数得鸡数.

想一想，为什么可以这样做？

（$a_1b_2-a_2b_1\neq0$）的算法，转换成计算机程序后，只需输入相应未知数的系数和常数项，就能计算出方程组的解.

练 习

1. 阅读下面的程序框图（图 11－6），若输入 $x=3$，则输出结果是多少？

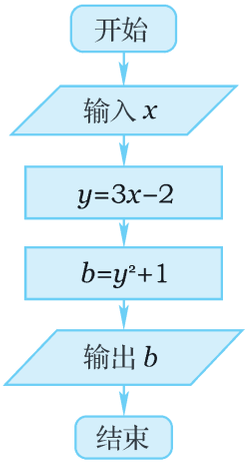


图 11－6

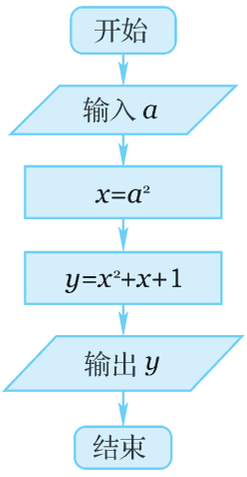


图 11－7

2. 阅读分析如图 11－7 所示的程序框图，回答下列问题：

（1）当输入 $a=2$ 时，输出值 y 是多少？

（2）当输出值 $y=3$ 时，则输入值 a 是多少？

3. 写出求解二元一次方程组

$$\begin{cases} x-2y=1, & ① \\ 2x+y=1 & ② \end{cases}$$

的算法步骤，并画出程序框图.

4. 已知一个三角形的三边长分别为 a，b，c，则它的面积可以用公式

$$S=\sqrt{p(p-a)(p-b)(p-c)}$$

来计算，其中 $p=\dfrac{a+b+c}{2}$. 请你设计一个用该公式计算三角形面积的算法，并画出程序框图.

这个公式称为“海伦-秦九韶公式”.

11.2.2 条件结构

在一个算法中，先根据条件是否成立作出判断，再决定执行哪一种操作，从而使算法流程产生不同流向的结构称为条件结构（conditional structure）.

条件结构的一般形式可以用程序框图表示为如下两种形式（图 11－8 和图 11－9）：

条件结构中都有一个判断框，框内注明判断的条件 p，条件 p 成立时，执行步骤 A，条件 p 不成立时，执行步骤 B 或退出条件结构执行后面的步骤.

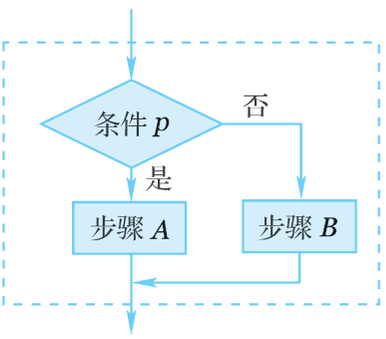


图 11－8

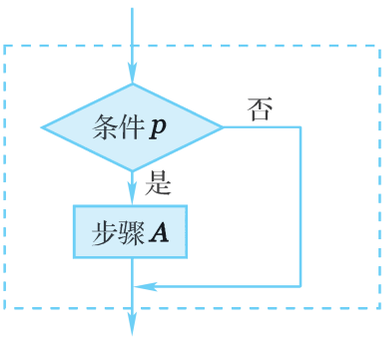


图 11－9

例 1 给定三条线段，其长度分别为 a，b，c，设计一个算法，判断此三条线段能否成为一个三角形的三边，并用程序框图表示出来.

算法分析：

验证是否满足“三角形任意两边之和大于第三边”，如果满足条件，则能构成三角形，否则不能. 这个验证需要用到条件结构.

算法步骤如下：

S1：输入 a，b，c；

S2：判断 $a+b>c$，$b+c>a$，$c+a>b$ 是否同时成立，若成立，则输出“能构成三角形”；否则，输出“不能构成三角形”.

程序框图如图 11－10 所示.

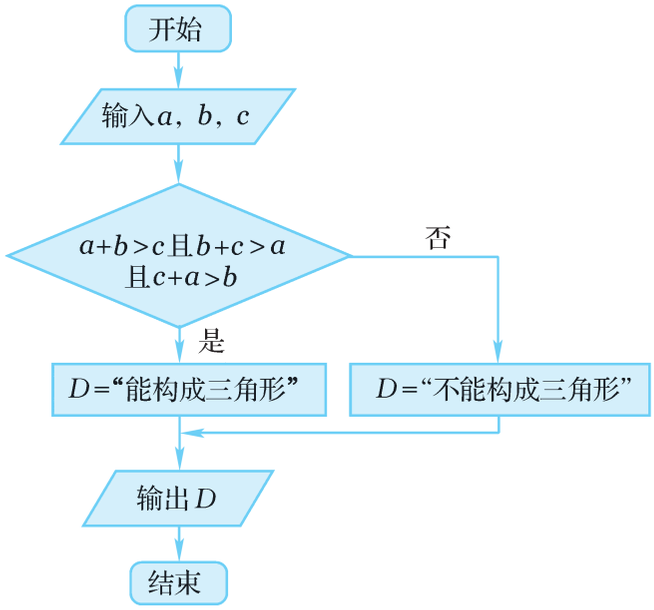


图 11－10

判断框中的条件是与流程走向相关联的，你能画出不同的程序框图吗？

例 2 设计一个判断一元二次方程 $ax^2+bx+c=0$（$a\neq0$）在实数范围内根的情况的算法，并用程序框图表示出来.

算法分析：

我们知道一元二次方程 $ax^2+bx+c=0$（$a\neq0$）在实数范围内根的情况可以由判别式 $\Delta=b^2-4ac$ 的值来判断. 若 $\Delta>0$，方程有两个不相等的实数根；若 $\Delta=0$，方程有两个相等的实数根；若 $\Delta<0$，方程没有实数根. 根据判别式的符号输出不同的结果，这个过程可以由条件结构来实现.

算法步骤如下：

S1：输入三个系数 a，b，c 的值；

S2：计算 $\Delta=b^2-4ac$；

S3：判断 $\Delta<0$ 是否成立，若是，则输出“方程没有实数根”，结束算法；若不是，则执行 S4；

S4：判断 $\Delta=0$ 是否成立，若是，则输出“方程有两个相等的实数根”；若不是，则输出“方程有两个不相等的实数根”.

程序框图如图 11－11 所示.

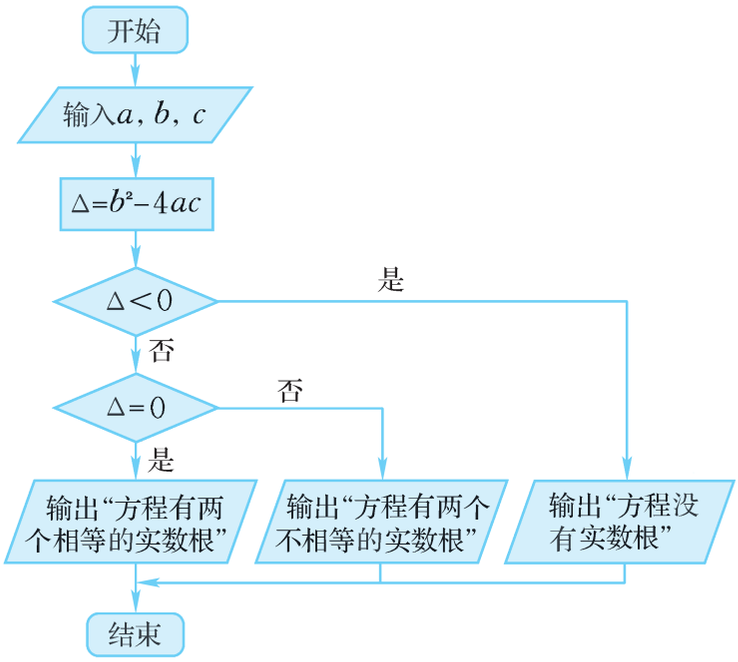


图 11－11

例 3 已知函数 $f(x)=\begin{cases}x^2-1 & (x\geqslant 0),\\ 2x-1 & (x<0),\end{cases}$

设计一个算法，求该函数的函数值，并画出程序框图.

算法分析：

这是一个分段函数，利用条件结构可以实现求不同条件下的函数值.

算法步骤如下：

S1：输入一个实数 x；

S2：判断 x 的符号，若 $x\geqslant 0$，则输出 x^2-1；否则，输出 $2x-1$.

程序框图如图 11－12 所示.

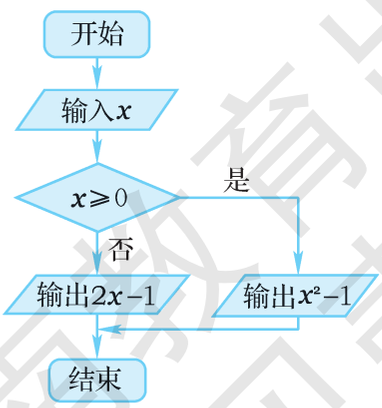


图 11－12

练 习

1. 阅读程序框图（图 11－13），若输入 $x=5$，则输出结果是多少？若输入 $x=2$，则输出结果是多少？

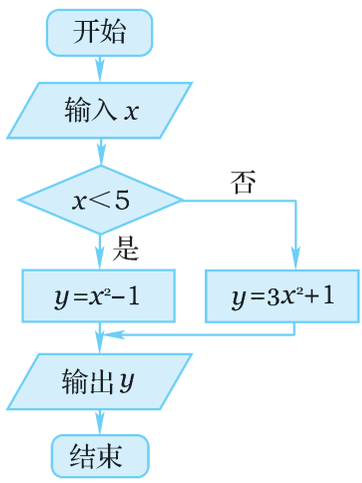


图 11－13

2. 一个商贩有 9 个外观相同的玉镯，其中有 1 个略重的假货．小明在“寻找假玉镯”算法问题中，用程序框图（图 11－14）表示了他的算法，请你阅读框图，并说出他的算法．

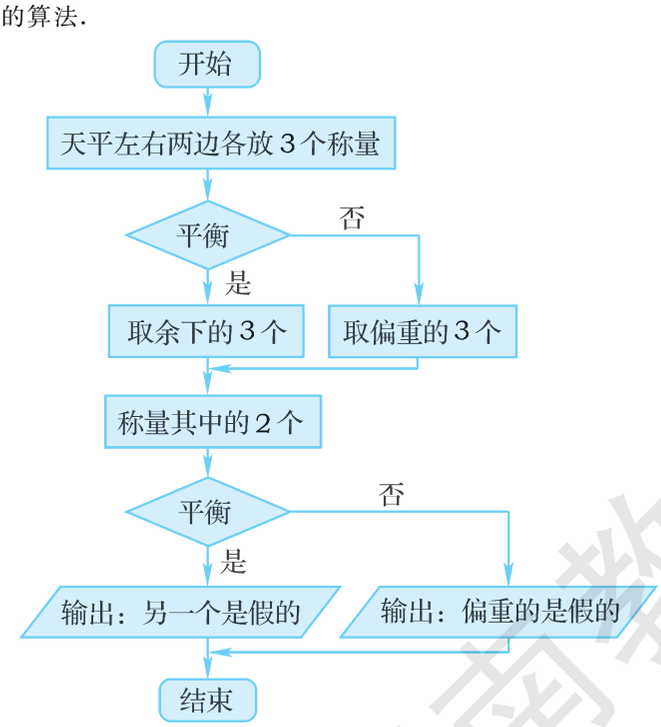


图 11－14

3. 已知函数

$$f(x)=\begin{cases}1 & (x\geqslant 0),\\ -1 & (x<0),\end{cases}$$

设计一个算法，求该函数的函数值，并画出程序框图.

4. 在学业水平考试中，规定 60 分以上（含 60 分）为“合格”，60 分以下为“不合格”. 请设计一个将考生的原始分数转换成等级的算法，并用程序框图表示出来.

11.2.3 循环结构

在算法中，从某处开始按照一定的条件重复执行某些步骤的结构称为循环结构（cycle structure），其中反复执行的步骤形成循环体.

循环结构的一般形式通常有两种，用程序框图分别表示如下（图 11－15 和图 11－16）：

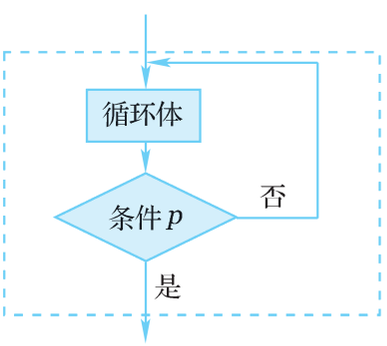


图 11－15

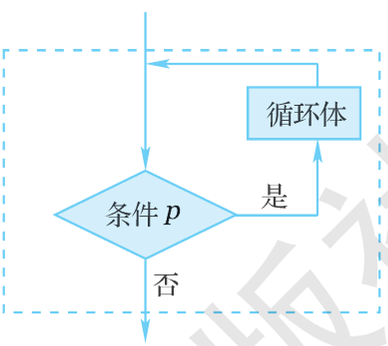


图 11－16

循环结构中一定包含有条件结构，用于确定何时终止执行循环体.

图 11－15 所示的循环结构是先执行循环体，再判断条件 p 是否成立，若条件 p 不成立，就继续执行循环体，直到条件 p 成立时才终止循环，我们把这种循环结构称为直到型循环结构.

你能说出直到型循环结构和当型循环结构的重要区别在哪里吗？

图 11－16 所示的循环结构是在每次执行循环体前都先判断条件 p 是否成立，当条件 p 成立，就执行循环体，当条件 p 不成立时才终止循环，我们把这种循环结构称为当型循环结构.

例 1 设计一个算法，计算 $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{100}$，并用程序框图表示出来.

算法分析：

按照加法法则，我们可以从左往右先计算 $1+\frac{1}{2}=\frac{3}{2}$，再把所得结果与$\frac{1}{3}$相加，依次进行下去，直到 99 次加法后得出结果. 显然这个过程包含重复操作的步骤，可以利用循环结构来实现.

算法步骤如下：

S1：赋初始值 $k=1$，$S=0$；

S2：若 $k\leqslant 100$ 成立，则执行 S3；否则，输出 S，结束算法；

S3：赋值 $S=S+\frac{1}{k}$，$k=k+1$.

程序框图如图 11－17 所示.

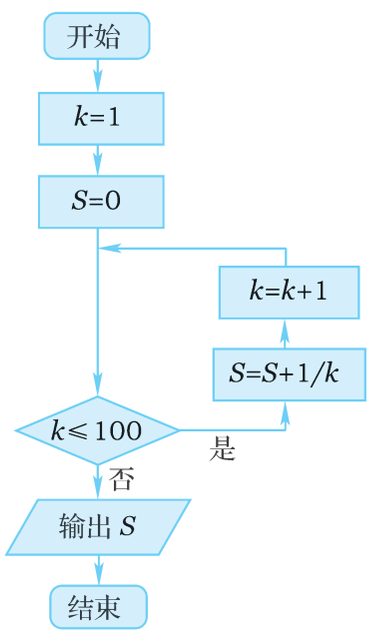


图 11－17

图 11－17 中的“k”一般被称为计数变量（或循环变量），被用来计算循环的次数；图 11－17 中的“S”一般被称为累加变量，被用来求和. 理解一个算法的作用，关键在于分析清楚算法中的计数变量和累加变量的作用.

引进变量 k 的目的是用来控制循环的次数，称为循环变量.

其中 S 一般称为“累加变量”，被用来求和.

判断框中的条件“$k\leqslant 100$”能改成“$k>100$”吗？

思考：在程序框图图 11－17 中使用了哪种循环结构？你能用另一种循环结构设计一个算法解决同样问题吗？

例 2 某公司购买了一辆价值 25 万元的商务车，汽车将以每年 20％的速度折旧，请用算法描述汽车的价值变化，输出几年后汽车的价值跌破 10 万元，并用程序框图表示出来.

算法分析：

按照题意，设汽车当年价值为 P，汽车每年折旧 20％，说明一年后汽车价值是上一年的 80％，即 $0.8P$，反映汽车价值变化即是这样一个循环过程，可以用循环结构来实现.

算法步骤如下：

S1：初始化变量 $P=25$，$k=0$；

S2：计算下一年汽车价值 P；

S3：判断汽车价值 P 是否小于 10，若是，输出 k 值结束算法；否则返回 S2.

程序框图如图 11－18 所示.

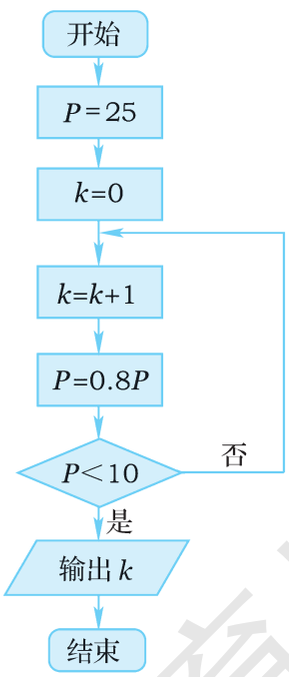


图 11－18

在这个程序框图中使用了哪种循环结构？你能用另一种循环结构设计算法来解决同样问题吗？

如果要求分别输出购车 1～5 年后汽车的价值，你能画出相应的程序框图吗？

通过上面的例题分析，我们发现要构造循环结构，需要确定循环体、初始化变量、设定循环控制条件等步骤.

思考：在循环结构中，循环控制条件的作用是什么？

练 习

1. 设计一个计算 $1\times3\times5\times\cdots\times99$ 的算法，并画出程序框图.

2. 设计一个计算 $1^2+2^2+3^2+\cdots+100^2$ 的算法，并画出程序框图.

3. 给出下面的程序框图（图 11－19），如果最后运行输出的结果是 2 070，那么图中“?”处应是多少?

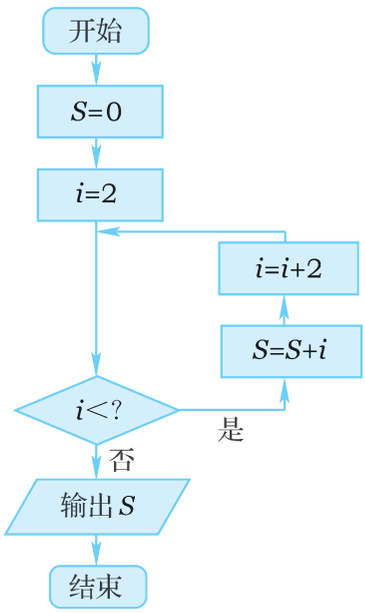


图 11－19

4. 阅读分析如图 11－20 所示的程序框图，说明它的功能.

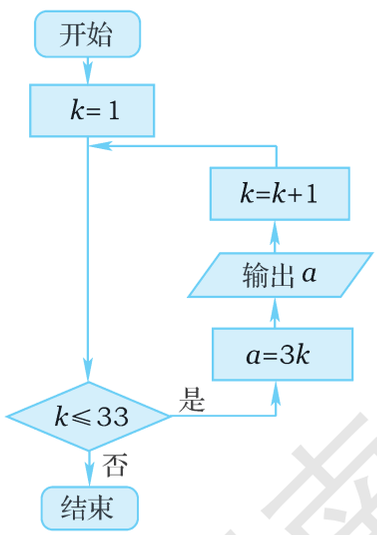


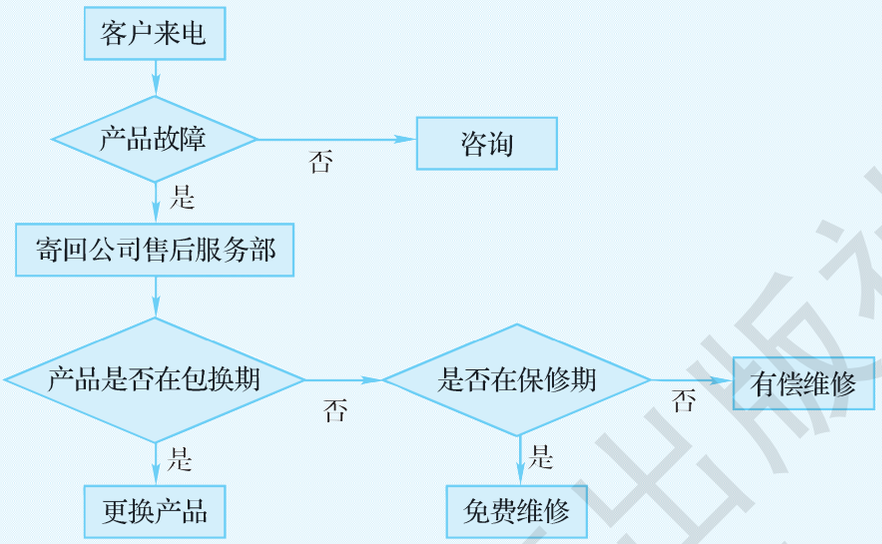
图 11－20

阅读与思考

生活中的流程图

流程图是流经一个系统的信息流、观点流或部件流的图形代表．它是一种用尽可能少、尽可能简单的方法来描述某个系统过程的方法，由于它的符号简单明了，所以非常易于阅读和理解．在我们的生活中，流程图随处可见，如在派出所我们可以看到“身份证申领流程图”，在医院我们可以看到“就诊流程图”，等等．

下面是一个产品售后服务流程图的例子：



在企业中，流程图主要用来说明某一过程．这种过程既可以是生产线上的工艺流程，也可以是完成一项任务必需的管理过程．例如，一张流程图能够成为解释某个零件的制造工序，甚至组织决策制定程序的方式之一．这些过程的各个阶段均用图形块表示，不同图形块之间以箭头相连，代表它们在系统内的流动方向．下一步何

去何从，要取决于上一步的结果，典型做法是用“是”或“否”的逻辑分支加以判断.

流程图是揭示和掌握封闭系统运动状况的有效方式. 作为诊断工具，它能够辅助决策制定，让管理者清楚地知道，问题可能出在什么地方，从而确定出可供选择的行动方案.

你能结合你的生活实际，设计一个学校“新生报到注册流程图”吗？

习题 2

学而时习之

1. 下面的 4 句话中不是解决问题的算法的是（　　）

   A. 从济南到北京旅游，先坐火车，再坐飞机抵达

   B. 解一元一次方程的步骤是去分母、去括号、移项、合并同类项、系数化为 1

   C. 方程 $x^2-1=0$ 的解是两个实根

   D. 求 $1+2+3+4+5$ 的值，先计算 $1+2=3$，再由 $3+3=6$，$6+4=10$，$10+5=15$，得最终结果为 15

2. 任何一个算法都必须有的基本结构是（　　）

   A. 顺序结构　　B. 条件结构　　C. 循环结构　　D. 以上都有

3. 设计一个算法，输出 1 000 以内能被 3 和 5 整除的所有正整数，并用程序框图表示出来.

4. 已知函数

$$y=\begin{cases} x & (x\leqslant 1), \\ 2x-1 & (1<x<10), \\ 3x-11 & (x\geqslant 10), \end{cases}$$

   设计一个算法，求该函数的函数值，并画出程序框图.

5. 某小区物业公司每月按楼层向居民收取电梯运行费，计费方法是：三楼及三楼以下的住户，每户收取 10 元，超过三楼的住户，每上一层楼，每户加收 2 元. 设计一个算法，根据输入的楼层，计算应收取的电梯运行费，并画出程序框图.

温故而知新

6. 某程序框图如图 11－21 所示，若输出的 $S=57$，则判断框内的条件是什么？

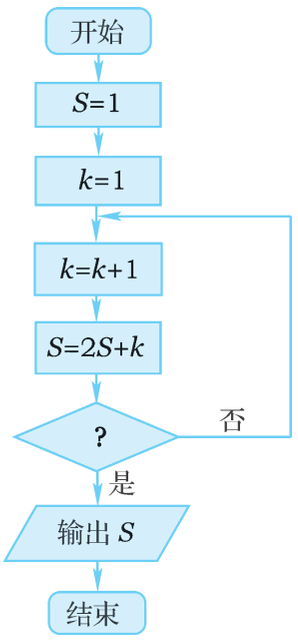


图 11－21

7. 阅读分析如图 11－22 所示的程序框图，说明该算法的功能.

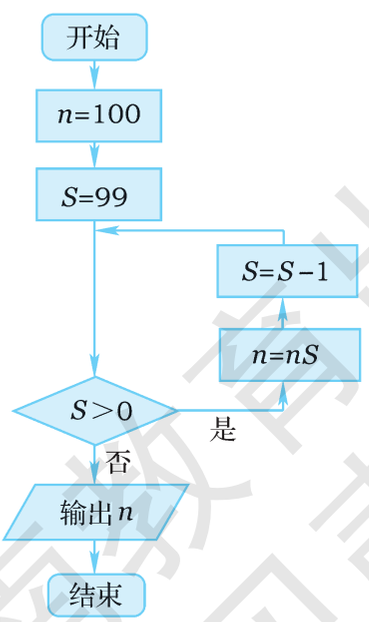


图 11－22

11.3 基本算法语句

前面我们学习了用自然语言和程序框图描述算法，但要使算法成为计算机解决问题的工具，还得借助计算机程序语言将算法编制成计算机程序. 尽管计算机程序语言种类很多，语法规则不尽相同，但都包含一些重要的基本语句，如输入语句、输出语句、赋值语句、条件语句和循环语句. 下面我们将学习把算法转换成算法语言（或称伪代码），伪代码稍加改造，就可以很方便地转换成各种计算机程序语言了.

11.3.1 输入、输出语句和赋值语句

在 11.2.1 节中，图 11－4 所示的算法就可以转换成下面的伪代码：

```
INPUT “x1，y1＝”；x1，y1
INPUT “x2，y2＝”；x2，y2
x0＝(x1＋x2)/2
y0＝(y1＋y2)/2
PRINT “x0，y0＝”；x0，y0
END
```

一个算法的伪代码由若干个语句行组成，按照从上到下的行排列的顺序执行其中的语句. 最后一行的结束语句是每个算法伪代码必不可少的.

其中第一行、第二行是输入语句，第五行是输出语句，第三行、第四行是赋值语句，最后一行“END”是表示算法结束的语句.

输入语句（input statement）、输出语句（output statement）分别与程序框图中的输入、输出框相对应，通常表示输入的数据和输出的结果.

输入语句的一般格式是：INPUT “提示信息”；变量

输入、输出语句中“提示信息”与变量或表达式之间用“；”隔开，输入语句一次可以给多个变量赋值，变量与变量之间用“，”分开.

其中“提示信息”一般是指提示用户输入什么样的内容，并非必需的部分，可以在语句中省略. 输入语句在计算机中会等待用户人工

干预，接受变量的值.

输出语句的一般格式是：PRINT“提示信息”；表达式

和输入语句一样，输出语句中也可以有“提示信息”，同样也不是必需的部分，而表达式则可以是常量、变量的值或者系统信息.

赋值语句（assignment statement）对应于程序框图中的处理框，负责将表达式的值赋给变量或者给变量提供初始值. 赋值语句的一般格式是：

变量＝表达式

其中的“＝”是赋值号，它的左右两边是同类型的变量或表达式，它和数学中的等号不完全一样. 赋值语句被执行时，将右边表达式的值赋给左边的变量.

例 1 某次数学竞赛的赛制规定，预赛成绩的 30％和复赛成绩之和构成选手的总成绩. 请设计一个算法，求参赛选手的总成绩.

算法分析：

要设计好算法，必要时，先将算法用程序框图表示出来，以呈现算法的逻辑结构，再转换成伪代码.

算法步骤如下：

S1：输入选手预赛成绩 x 和复赛成绩 y；

S2：计算总成绩 $S=0.3x+y$；

S3：输出 S.

程序框图如图 11－23 所示.

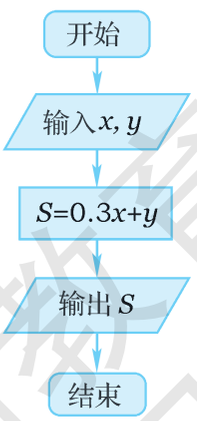


图 11－23

伪代码：

```
INPUT “x=”; x
INPUT “y=”; y
S=0.3 * x+y
PRINT “S=”; S
END
```

编写伪代码的一般步骤：

1. 根据提供的问题，利用相关知识设计解决问题的算法；

2. 依据算法分析，画出程序框图；

3. 根据程序框图中的算法步骤，逐步把算法用相应的算法语句表达出来.

熟练之后前两步可以省略.

例 2 阅读下面的伪代码，PRINT 语句输出的值是多少？

```
a=1
a=a^3+2
PRINT  a
END
```

在伪代码语句中，$a\wedge 3$ 表示 a^3，$a\wedge b$ 是乘幂运算，即表示 a^b. “×”用“*”表示，“÷”用“/”表示；“<>”表示“≠”，“>=”表示“≥”，“<=”表示“≤”等.

解 因为赋值语句是将右边表达式的值赋给左边的变量，所以在程序的第二行中，右边 a 的值是 1，a^3+2 的值是 3，“a=a^3+2”的作用是将 a^3+2 的值仍然赋给 a，所以第三行的输出值是 3.

例 3 阅读下面的伪代码，说明它的功能是什么.

```
INPUT  x, y
PRINT  x, y
a=x
x=y
y=a
PRINT  x, y
END
```

你知道变量 a 在这里的作用是什么？

解 第三行“a=x”将变量 x 的值赋给了 a，变量 x 空出；第四行“x=y”将变量 y 的值赋给了变量 x，变量 y 又空出；紧接着第五行“y=a”又将变量 a 的值赋给了变量 y，实现了两个变量 x，y 的

值的交换. 可见第二行与第六行虽然语句完全相同，但输出的结果不一样，分别是变量 x 和变量 y 交换前后的值.

练习

1. 设计一个算法，要求输入一个圆的半径，能输出圆的周长和面积（π 取 3.14），并分别用程序框图和伪代码表示你的算法.
2. 设计一个算法，计算并输出某次考试中学生的语文、数学、英语三科的平均成绩，并分别用程序框图和伪代码表示你的算法.
3. 阅读下面的伪代码：

```
INPUT "a=";a
b=a+3
a=a+2
b=a+b
PRINT  a,b
END
```

若输入的值是 5，则最后输出 a，b 的值分别是多少？

4. 已知函数 $y=x^2+2x-3$，试设计一个算法，任意输入一个自变量 x 的值，计算该函数的函数值，并分别用程序框图和伪代码表示你的算法.

11.3.2 条件语句

条件语句（conditional statement）是用来表达程序框图中条件结构的常用语句. 与图 11－8 和图 11－9 所示的条件结构的一般形式相对应，条件语句的一般格式也有两种：

```
IF  条件  THEN
    语句体 A
ELSE
    语句体 B
END IF
```

```
IF  条件  THEN
    语句体 A
END IF
```

这里的语句体是指计算机按顺序执行的一组语句. 在书写时将它们缩进，是为了体现语句的逻辑结构，便于阅读和交流.

上述语句被执行时，首先对 IF 后面的条件进行判断，如果（IF）条件符合，那么（THEN）就执行后面的语句体 A，否则（ELSE）就执行语句体 B 或者直接退出条件结构，执行 END IF 后面的语句.

图 11－10 所示的程序框图就可以转换成下面的伪代码：

```
INPUT “a，b，c=”；a，b，c
IF  a+b>c  AND  b+c>a  AND  c+a>b  THEN
    D=“能构成三角形”
ELSE
    D=“不能构成三角形”
END IF
PRINT  D
END
```

这里 D 是文本变量，用来储存文本信息. “AND”用于连接几个同时满足的条件，对应于逻辑连接词“且”. 与之相应的还有“OR”，相当于“或”.

计算机在执行这个程序时，当由输入语句输入 a，b，c 后，首先判断 $a+b>c$，$b+c>a$，$c+a>b$ 是否同时成立，若是，则给文本变量 D 赋值“能构成三角形”；否则给文本变量 D 赋值“不能构成三角形”，最后输出文本信息.

例 1 设计一个算法，比较两个数的大小，并输出较大的数.

算法分析：

设要比较的两个数是变量 a 和 b，为节约变量，仍用变量 a 表示较大的数，即最后总是输出 a.

算法步骤如下：

S1：输入 a，b；

S2：判断 $a<b$ 是否成立，若不是，执行 S3；若是，则把 b 赋给 a；

S3：输出 a.

程序框图如图 11－24 所示.

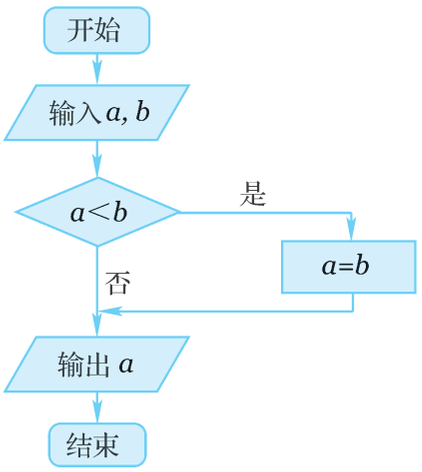


图 11－24

伪代码：

```
INPUT  a，b
IF  a<b  THEN
    a=b
END IF
PRINT  a
END
```

例 2 自来水公司对用户用水收费规定：每月用水量在 3 t 以内者，每吨收 2.1 元；每月用水量超过 3 t 且在 5 t 以内者，超过的部分，每吨收 2.6 元；每月用水量超过 5 t 者，超过 3 t 的部分，每吨收 3.2 元. 请为自来水公司编写一个计费算法的伪代码.

算法分析：

用变量 x 表示用户的用水量（t），用变量 y 表示用户应缴纳的水费（元）.

算法步骤如下：

S1：输入 x；

S2：若 $x \leqslant 3$，则 $y=2.1x$；若 $3<x \leqslant 5$，则 $y=6.3+2.6(x-3)$；若 $x>5$，则 $y=6.3+3.2(x-3)$；

S3：输出 y.

程序框图如图 11－25 所示.

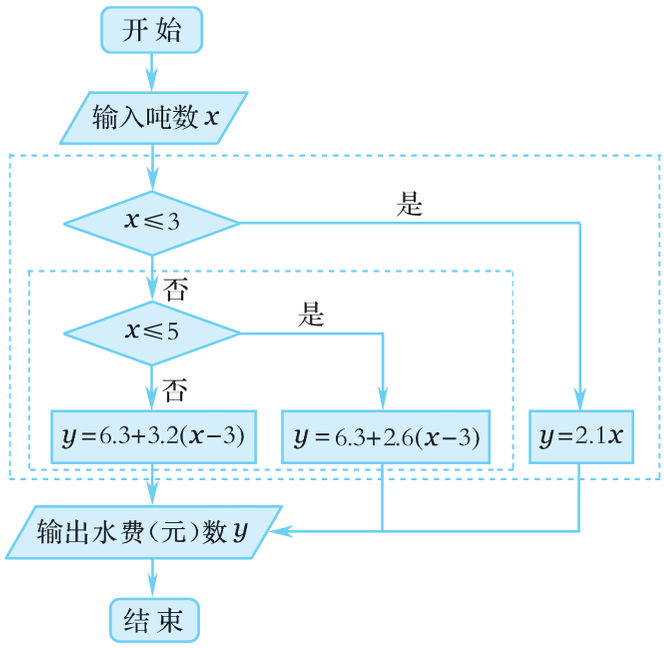


图 11－25

伪代码：

```
INPUT "x="; x
IF  x<=3  THEN
    y=2.1*x
ELSE
    IF  x<=5  THEN
        y=6.3+2.6*(x-3)
    ELSE
        y=6.3+3.2*(x-3)
    END IF
END IF
PRINT  "y="; y
END
```

是否可以将嵌套的条件交换位置？试着改写程序框图和相应的伪代码.

为什么在伪代码中会连续出现两行“END IF”？

在这段伪代码中用到了条件语句的嵌套（即复合 IF 语句），对应于复合条件结构，在书写伪代码时，将虚线框内的条件语句模块整体缩进，是为了体现算法的层次性，便于阅读与交流.

思考：先将图 11－11 所示的判断一元二次方程 $ax^2+bx+c=0$（$a\neq0$）在实数范围内根的情况的算法用伪代码表示出来，你能总结出此类条件语句的嵌套的一般形式吗？

例 3 阅读下面的伪代码，说明它的功能是什么.

```
INPUT "a，b，c="；a，b，c
IF  b>a  THEN
    t=a
    a=b
    b=t
END IF
IF  c>a  THEN
    t=a
    a=c
    c=t
END IF
IF  c>b  THEN
    t=b
    b=c
    c=t
END IF
PRINT  a，b，c
END
```

条件语句嵌套的一般形式：

```
IF 条件 THEN
        语句体 A
ELSE
    IF 条件 THEN
        语句体 B
    ELSE
        语句体 C
    END IF
END IF
```

解 程序中包含三个条件语句模块，通过 11.3.1 节中例 3 的学习我们知道，下面的三行语句

```
t=a
a=b
b=t
```

实现了两个变量 a 和 b 的值的交换，结合执行这三行语句的条件“$b>a$”，那么在“$a\geqslant b$”时不交换变量的值，因而保证了最后输出的总是“$a\geqslant b$”. 同理后面的两组也保证了 $a\geqslant c$ 和 $b\geqslant c$，即 $a\geqslant b\geqslant c$. 因此该伪代码可以实现对任意输入的三个数 a，b，c 按从大到小排列.

练 习

1. 设计一个算法，求实数 x 的绝对值，并用程序框图和伪代码表示.

2. 设计一个算法，对于函数

$$y=\begin{cases} x+1 & (x<1), \\ 2x & (1\leqslant x<10), \\ 3x-10 & (x\geqslant 10), \end{cases}$$

输入 x 的值，输出相应的函数值，并用程序框图和伪代码表示.

3. 到某商业银行办理个人异地汇款时，银行要收取一定的手续费，汇款额不超过 100 元，收取 1 元手续费；超过 100 元但不超过 5 000 元，按汇款额的 1% 收取；超过 5 000 元，一律收取 50 元手续费. 试编写伪代码，描述汇款额为 x（元）时，银行收取的手续费 y（元）的算法过程.

4. 阅读下面的伪代码：

```
INPUT  a，b，c
IF  a>b  THEN
    a=b
END IF
IF  a>c  THEN
    a=c
END IF
PRINT  a
END
```

如果输入的三个数是 -2，-13，7，那么输出的结果是什么？说明该算法的功能.

11.3.3 循环语句

循环语句（cycle statement）与程序框图中的循环结构相对应，用来控制算法中在一定条件下需要重复执行的步骤.

与图 11－15 所示的直到型循环结构相对应的是直到型循环语句（UNTIL 语句），它的一般格式是：

```
DO
    循环体
LOOP UNTIL  条件
```

这里循环体是指计算机反复执行的一组程序语句.

当计算机执行上面的语句时，先执行循环体中的语句，然后对 UNTIL 之后的条件进行判断，当条件不满足时，继续执行 DO 和 UNTIL 之间的循环体语句，判断条件是否满足，直到条件满足时退出循环体，执行 UNTIL 之后的其他语句.

利用直到型循环语句可以将程序框图图 11－18 所示算法转换成以下伪代码：

```
P＝25
k＝0
DO
    k＝k＋1
    P＝0.8＊P
LOOP UNTIL  P＜10
PRINT  k
END
```

与图 11－16 所示的当型循环结构相对应的是当型循环语句(WHILE 语句)，它的一般格式是：

```
WHILE  条件
    循环体
WEND
```

当计算机执行上面的语句时，先判断 WHILE 之后的条件是否满足，当条件满足时，执行 WHILE 和 WEND 之间的循环体语句，再判断条件是否满足，直到条件不满足时退出循环体，执行 WEND 之后的其他语句.

利用当型循环语句可以将程序框图图 11－17 所示算法转换成以下伪代码：

```
k＝1
S＝0
WHILE  k＜＝100
    S＝S＋1/k
    k＝k＋1
WEND
PRINT  S
END
```

UNTIL 语句和 WHILE 语句虽然语法格式不一样，但都可以实现算法中循环结构的转换. 如上面的伪代码就可以改用 UNTIL 语句来表达：

```
k=1
S=0
DO
    S=S+1/k
    k=k+1
LOOP UNTIL  k>100
PRINT  S
END
```

可见，用两种不同循环语句编写的伪代码可以相互转化，实现相同的功能.

例 1 已知函数 $y=x^3-25x+7$，编写伪代码，要求连续输入自变量的 10 个值，计算并输出相应的函数值.

算法分析：

算法的重点在于如何控制输入的自变量的个数，可以用循环语句来实现.

具体步骤如下：

S1：输入自变量的值 x；

S2：计算 y；

S3：输出 y；

S4：记录输入次数；

S5：判断输入次数是否大于 10，若是，则结束算法；否则返回 S1.

程序框图如图 11－26 所示.

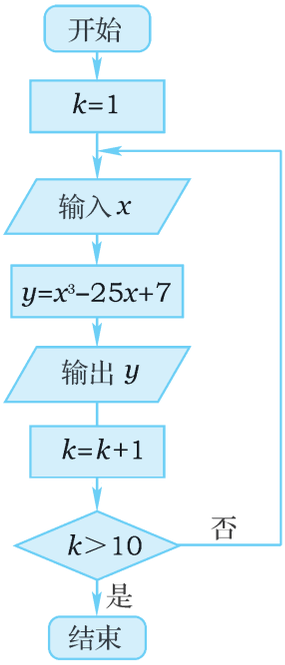


图 11－26

伪代码：

```
k＝1
DO
    INPUT   x
    y＝x^3－25＊x＋7
    PRINT   y
    k＝k＋1
LOOP UNTIL   k＞10
END
```

你能用另一种循环语句表达吗？

这里是用到哪种循环结构？

例 2 数列 0，1，1，2，3，5，8，…叫作斐波那契数列，它的后一项等于前两项的和. 阅读下面的伪代码，说明它的功能.

```
A1＝0
A2＝1
k＝3
PRINT　A1，A2
WHILE　k＜61
　　A3＝A1＋A2
　　PRINT　A3
　　A1＝A2
　　A2＝A3
　　k＝k＋1
WEND
END
```

这种及时释放变量的做法，是算法设计的一个重要原则，因为它可以减少变量，为计算机节约存储空间，提高计算机工作效率.

算法分析：

前四行很容易理解，输出了斐波那契数列的前两项，执行到第七行时，输出的是斐波那契数列的第三项. 后面两行即对变量 $A1$，$A2$ 重新赋值，将前面已经输出的第二项和第三项赋给 $A1$ 和 $A2$，当 k 的值增加到 4 时，继续执行循环体，此时 $A3＝A1＋A2$ 实际上就计算出了斐波那契数列的第四项. 如此反复，当 $k＝61$ 时退出循环，因此，该伪代码的功能是依次输出斐波那契数列的前 60 项.

例 3　设计一个算法，将一个班 40 个同学的考试成绩中不及格（少于 60 分）的分数找出来.

算法分析：

从输入第一个分数开始，就逐一与 60 比较，若小于 60 就输出，否则就继续与下一个输入的数做比较，需要反复比较 40 次，可以用一个计数变量来控制输入的数的个数.

具体步骤如下：

S1：输入分数 x；

S2：判断是否小于 60，若是，输出分数，否则执行下一步；

S3：判断输入的分数是否达到 40 个，若是，结束算法；若没有，则返回 S1.

程序框图如图 11－27 所示.

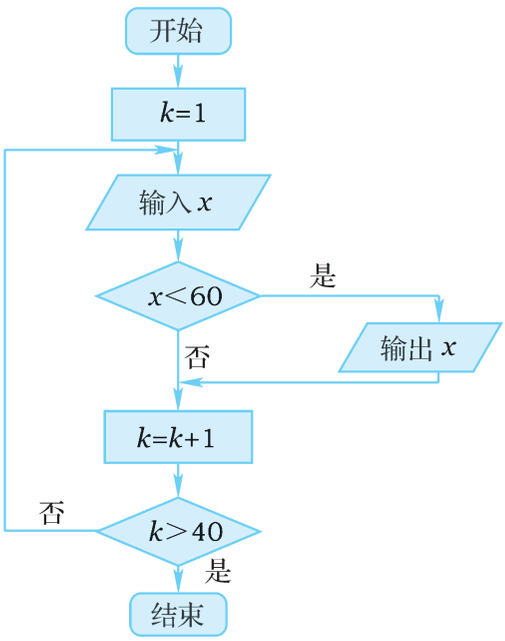


图 11－27

伪代码：

```
k＝1
DO
    INPUT   x
    IF   x＜60   THEN
        PRINT   x
    END IF
    k＝k＋1
LOOP UNTIL   k＞40
END
```

练 习

1. 分别用 UNTIL 语句和 WHILE 语句设计算法，求 $100\times99\times\cdots\times2\times1$ 的值.

2. 阅读下面的伪代码，说明该算法的处理功能.

```
S=0
T=1
k=1
WHILE  k<21
    k=k+1
    S=S+k
    T=T*k
WEND
PRINT  S，T
END
```

3. 设计一个算法，求 $\frac{1}{1\times2}+\frac{1}{2\times3}+\cdots+\frac{1}{99\times100}$ 的值.

4. 某班共有 40 个学生，每次考试后，老师总要统计成绩在 80 分以上、60～80 分和 60 分以下的各分数段人数，请你帮助老师设计一个算法.

习题 3

学而时习之

1. 阅读下面的伪代码，写出程序表示的函数.

```
INPUT x
IF x<0 THEN
    y=0
ELSE
    IF x<1 THEN
        y=1
    ELSE
        y=x
    END IF
END IF
PRINT y
END
```

2. 已知摄氏温度与华氏温度的转换公式是：

$$\text{摄氏温度}=(\text{华氏温度}-32)\times\frac{5}{9}$$

编写伪代码，输入一个华氏温度，输出其相应的摄氏温度.

3. 编写伪代码，判断任意输入的整数的奇偶性.

4. 编写伪代码，判断大于 2 的整数是否是质数.

5. 某停车场对小型汽车临时停车（不超过 24 h）的收费标准是：停车时间 1 h（含 1 h）内收费 5 元，停车时间 1 h 以上，超出 1 h 的部分每小时加收 5 元，临时停车最高限额为 100 元. 请据此设计一个计费算法.

温故而知新

6. 闰年是指能被 4 整除但不能被 100 整除，或者能被 400 整除的年份，编写伪代码，判断输入的年份是否为闰年.

7. 请设计一个算法，求 $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{1}{99}-\frac{1}{100}$ 的值.

11.4 算法案例

本节我们介绍几个引人入胜的数学问题的算法，进一步体会算法思想，若能举一反三，用心探索，则可以提高逻辑思维水平和算法设计能力.

案例 1：辗转相除法

在本章开始，我们介绍了用更相减损术求两个正整数的最大公约数的算法，下面我们介绍另一种古老而有效的算法——辗转相除法，利用它同样可以求两个正整数 a，b（$a>b$）的最大公约数. 这种方法的核心是要求出这样一列数：

$$a，b，r_1，r_2，\cdots，r_n，0$$

这列数从第三项开始起，每一项都是前两项相除所得的余数，余数为 0 的前一项 r_n 就是 a，b（$a>b$）的最大公约数. 下面举例说明.

辗转相除法是公元前 3 世纪，欧几里得首先提出来的，又叫欧几里得算法.

例 1 求 8 251 和 6 105 的最大公约数.

解 因为 $8\ 251=6\ 105\times1+2\ 146$，

$6\ 105=2\ 146\times2+1\ 813$，

$2\ 146=1\ 813\times1+333$，

$1\ 813=333\times5+148$，

$333=148\times2+37$，

$148=37\times4$，

所以 8 251 和 6 105 的最大公约数是 37.

算法分析：

通过上面的计算过程，我们相当于找到了这样一列数 8 251，6 105，2 146，1 813，333，148，37，0，因为 37 是 148 和 37 的最大公约数，从“$333=148\times2+37$”可知，37 也是 333 和 148 的最大公约数，因而也是 1 813 和 333 的最大公约数……依此类推，37 就是 8 251 和 6 105 的最大公约数.

由此我们可以归纳出用辗转相除法求两个正整数 a，b（$a>b$）

的最大公约数的算法步骤：

S1：输入两个正整数 a，b；

S2：计算 a 除以 b 所得的余数 r，即 $r=a$ MOD b；

S3：$a=b$，$b=r$；

S4：判断 $r=0$ 是否成立，若成立，输出最大公约数 a；否则返回 S2.

程序框图如图 11－28 所示.

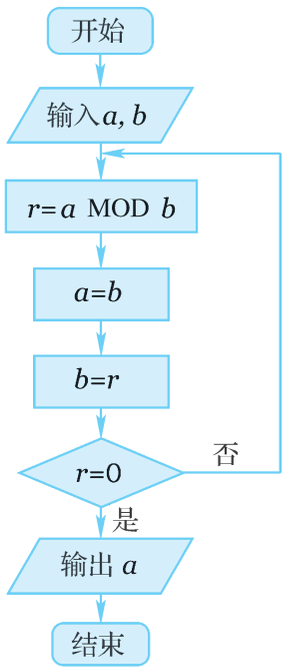


图 11－28

伪代码：

```
INPUT   a，b
DO
      r＝a MOD b
      a＝b
      b＝r
LOOP UNTIL   r＝0
PRINT   a
END
```

思考：你能将上面的算法、程序框图和伪代码改用当型循环结构

这里 a MOD b 是一个函数，表示 a 除以 b 所得的余数，$a \backslash b$ 表示 a 除以 b 所得的商.

常用到的函数还有：ABS(x)用来求 x 的绝对值，即 ABS(x)＝$|x|$；SQR(x)用来求某个非负数 x 的算术平方根，即 SQR(x)＝$\sqrt{x}$.

来描述吗?

案例 2：中国剩余定理

例 2 “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，问物几何?”即一个整数除以 3 余 2，除以 5 余 3，除以 7 余 2，求符合条件的最小正整数.

算法分析：

先列出除以 3 余 2 的数：2，5，8，11，14，17，20，23，26，…

再列出除以 5 余 3 的数：3，8，13，18，23，28，…

这两列数中，首先出现的公共数是 8.

3 与 5 的最小公倍数是 15. 两个条件合并成一个就是 8＋15×整数，列出这一串数是 8，23，38，…

再列出除以 7 余 2 的数 2，9，16，23，30，…

就得出符合题目条件的最小数是 23.

事实上，我们已把题目中的三个条件合并成一个：被 105 除余 23.

自从《孙子算经》中提出这个“物不知其数”的问题后，便引起了人们很大的兴趣，提出了这类问题的不同解法.

《孙子算经》约成书于公元 400 年前后，作者生平和编写年代都不清楚. 现在传承的《孙子算经》共三卷，本题出现在下卷的第 26 题.

到了明代，数学家程大位用诗歌概括了这一算法，可谓绝伦美妙，他写道：

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.

这首诗的意思是：用 3 除所得的余数乘上 70，加上用 5 除所得余数乘以 21，再加上用 7 除所得的余数乘上 15，结果大于 105 就减去 105 的倍数，这样就知道所求的数了.

运用这种方法，可以很快得出答案. 因为 2×70＋3×21＋2×15＝233，233－105×2＝23，所以符合题目条件的最小数是 23.

这种问题的通用解法相当于求一个不定方程组

$$\begin{cases} m=3x+2, \\ m=5y+3, \\ m=7z+2 \end{cases}$$

的正整数解.

解决这个问题的算法思路如下：假定所求的数为 m，分别计算 m 被 3，5，7 除所得的余数 a，b，c，即 $a=m$ MOD 3，$b=m$ MOD 5，$c=m$ MOD 7，可以从 $m=3$ 开始检验三个条件 $a=2$，$b=3$，$c=2$ 是否同时成立，若三个条件中有任何一个不成立，则 m 递增 1，当三个条件同时成立时，则输出 m 的值.

程序框图如图 11－29 所示.

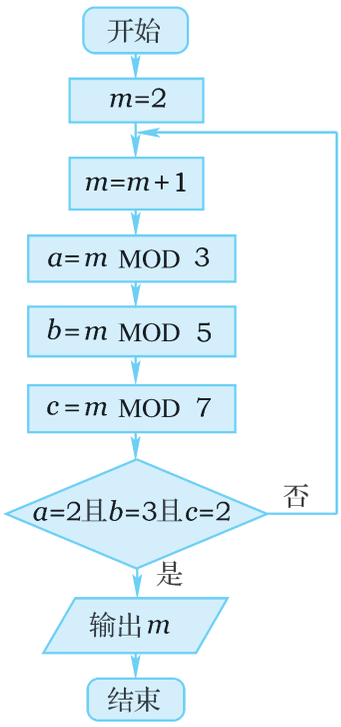


图 11－29

伪代码：

```
m＝2
DO
    m＝m＋1
    a＝m MOD 3
    b＝m MOD 5
    c＝m MOD 7
LOOP UNTIL  a＝2 AND b＝3 AND c＝2
PRINT  m
END
```

案例 3：二分法

前面我们学过求函数零点（方程的根）近似值的二分法，其基本思想是，将方程的有解区间平均分成两个小区间，然后判断解在哪个小区间内；继续把有解区间一分为二进行判断，如此反复进行，直到求出满足精度要求的近似解.

下面结合具体实例，体会二分法的实现过程.

例 3 求方程 $f(x)=x^5+x-1=0$ 在 $[0, 1]$ 上的根的近似值，精确到 0.05.

解 ∵$f(0)=-1$，$f(1)=1$，$f(0)\times f(1)<0$，

∴区间 $[0, 1]$ 为方程的有解区间，精度为 $1-0=1>0.05$；

取区间 $[0, 1]$ 的中点 0.5，计算 $f(0.5)\approx -0.468\ 8$；由 $f(0.5)\times f(1)<0$ 可得新的有解区间为 $[0.5, 1]$，精度为 $1-0.5=0.5>0.05$；

取区间 $[0.5, 1]$ 的中点 0.75，计算 $f(0.75)\approx -0.012\ 7$；由 $f(0.75)\times f(1)<0$ 可得新的有解区间为 $[0.75, 1]$，精度为 $1-0.75=0.25>0.05$；

取区间 $[0.75, 1]$ 的中点 0.875，计算 $f(0.875)\approx 0.387\ 9$；由 $f(0.75)\times f(0.875)<0$ 可得新的有解区间为 $[0.75, 0.875]$，精度为 $0.875-0.75=0.125>0.05$；

取区间 $[0.75, 0.875]$ 的中点 0.812 5，计算 $f(0.812\ 5)\approx 0.166\ 6$；由 $f(0.75)\times f(0.812\ 5)<0$ 可得新的有解区间为 $[0.75, 0.812\ 5]$，精度为 $0.812\ 5-0.75=0.062\ 5>0.05$；

取区间 $[0.75, 0.812\ 5]$ 的中点 0.781 25，计算 $f(0.781\ 25)\approx 0.072\ 29$；由 $f(0.75)\times f(0.781\ 25)<0$ 可得新的有解区间为 $[0.75, 0.781\ 25]$，精度为 $0.781\ 25-0.75=0.031\ 25<0.05$；

至此已满足精度要求，一般取区间 $[0.75, 0.781\ 25]$ 的中点 0.765 625，它就是方程的一个近似解.

实际上，在区间 $[0.75, 0.781\ 25]$ 中的实数都是当精度为 0.05 时原方程的近似解.

算法分析：

从上面的解答过程我们发现，一些步骤需要重复进行，因而我们可以使用循环结构来实现这一算法功能.

下面是估计方程 $f(x)=0$ 在某有解区间 $[a,b]$ 上符合误差限制 c 的近似解的一般算法：

S1：确定有解区间 $[a,b]$ 和精度 c；

S2：取 $[a,b]$ 的中点 $x_0=\dfrac{a+b}{2}$；

S3：若 $|a-b|\geqslant c$，则进入 S4；否则输出 x_0 结束算法；

S4：若 $f(x_0)\neq 0$，则进入 S5；否则 $x=x_0$ 就是方程的根，输出 x_0，结束算法；

S5：若 $f(a)f(x_0)>0$，则解在 $[x_0,b]$，用 x_0 替换 a；若 $f(a)f(x_0)<0$，则解在 $[a,x_0]$，用 x_0 替换 b；返回 S2.

程序框图如图 11－30 所示.

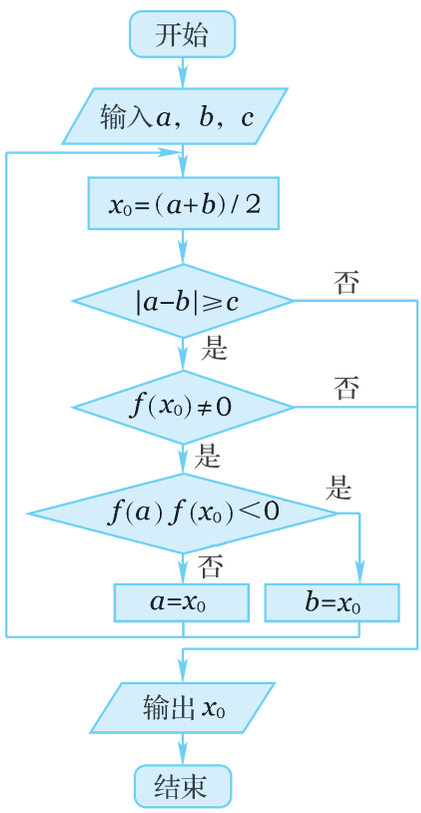


图 11－30

伪代码：

```
INPUT  a，b，c
x0＝(a＋b)/2
WHILE f(x0)<>0 AND ABS(a－b)>＝c
    IF f(a)＊f(x0)<0 THEN
        b＝x0
    ELSE
        a＝x0
    END IF
    x0＝(a＋b)/2
WEND
PRINT  x0
END
```

案例 4：秦九韶算法

怎样求多项式 $f(x)=x^5+x^4+x^3+x^2+x+1$ 当 $x=5$ 时的值呢？

一种自然的做法是，因为 $f(x)=x^5+x^4+x^3+x^2+x+1$，

所以 $f(5)=5^5+5^4+5^3+5^2+5+1$

$=3\ 125+625+125+25+5+1$

$=3\ 906.$

这时，共做了 $1+2+3+4=10$（次）乘法运算，5 次加法运算.

也可以用另一种方法来计算：

$f(5)=5^5+5^4+5^3+5^2+5+1$

$=5\times(5^4+5^3+5^2+5+1)+1$

$=5\times(5\times(5^3+5^2+5+1)+1)+1$

$=5\times(5\times(5\times(5^2+5+1)+1)+1)+1$

$=5\times(5\times(5\times(5\times(5+1)+1)+1)+1)+1$

这样只需做 4 次乘法运算，5 次加法运算就可以了. 相比之下，这种做法减少了运算次数，因而能够提高运算效率.

一般地，把一个 n 次多项式 $f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$ 改写成如下形式：

$$\begin{aligned}f(x)&=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0\\&=(a_nx^{n-1}+a_{n-1}x^{n-2}+\cdots+a_1)x+a_0\\&=((a_nx^{n-2}+a_{n-1}x^{n-3}+\cdots+a_2)x+a_1)x+a_0\\&=\cdots\cdots\\&=(\cdots((a_nx+a_{n-1})x+a_{n-2})x+\cdots+a_1)x+a_0.\end{aligned}$$

求多项式的值时，首先计算最内层括号内一次多项式的值，即

$$v_1=a_nx+a_{n-1},$$

然后由内向外逐层计算一次多项式的值，即

$$v_2=v_1x+a_{n-2},$$

$$v_3=v_2x+a_{n-3},$$

$$\cdots\cdots$$

$$v_n=v_{n-1}x+a_0.$$

这样，求 n 次多项式 $f(x)$ 的值就转化为求 n 个一次多项式的值.

像这样将一元 n 次多项式的求值问题转化为求 n 个一次多项式的值的算法称为秦九韶算法，其大大简化了计算过程，即使在现代，利用计算机解决多项式的求值问题时，秦九韶算法依然是最优的算法.

例 4 已知一个五次多项式 $f(x)=5x^5+2x^4+3.5x^3-2.6x^2+1.7x-0.8$，用秦九韶算法求这个多项式当 $x=5$ 时的值.

解 将多项式变形：

$$f(x)=((((5x+2)x+3.5)x-2.6)x+1.7)x-0.8,$$

按由里到外的顺序，依次计算一次多项式当 $x=5$ 时的值：

$$v_0=5,$$

$$v_1=5\times5+2=27,$$

$$v_2=27\times5+3.5=138.5,$$

$$v_3=138.5\times5-2.6=689.9,$$

$$v_4=689.9\times5+1.7=3\ 451.2,$$

$$v_5=3\ 451.2\times5-0.8=17\ 255.2.$$

所以，当 $x=5$ 时，多项式的值等于 17 255.2.

“秦九韶算法”在西方被称作霍纳算法，是以英国数学家霍纳命名的.

秦九韶（约公元 1208—约 1261 年），字道古，南宋末年人，出生于鲁郡（今山东曲阜一带人）. 早年曾从隐君子学算术，后因其父往四川做官，即随父迁徙，后也认为是普州安岳（今四川安岳县）人. 秦九韶与李冶、杨辉、朱世杰并称宋元数学四大家.

秦九韶

思考：你能把用秦九韶算法求 n 次多项式 $f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$ 当 $x=x_0$ 时的值的算法表示出来吗？

算法分析：

从上面的解答过程我们发现，计算 v_k 时要用到 v_{k-1} 的值，若令 $v_0=a_n$，那么

$$v_k=v_{k-1}x+a_{n-k}\quad(k=1,2,\cdots,n)$$

是一个反复执行的步骤，可以用循环结构来实现.

具体算法步骤如下：

S1：输入多项式次数 n、最高次项的系数 a_n 和 x 的值；

S2：将 v 的值初始化为 a_n，将 k 的值初始化为 $n-1$；

S3：输入 k 次项的系数 a_k；

S4：$v=vx+a_k$，$k=k-1$；

S5：判断 k 是否大于或等于 0，若是，则返回 S3；否则输出多项式的值 v.

程序框图如图 11－31 所示.

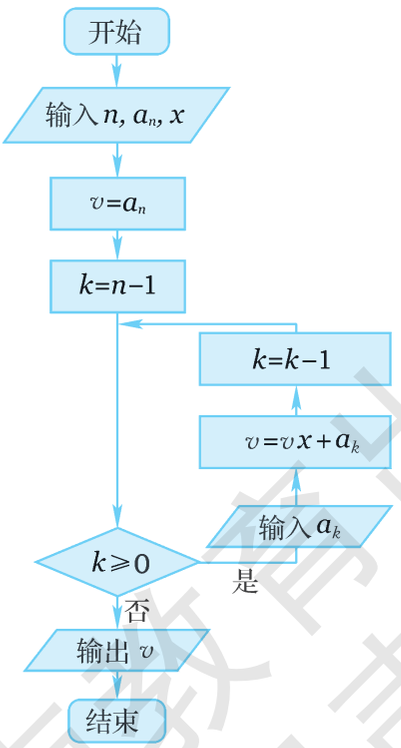


图 11－31

伪代码：

```
INPUT "n="; n
INPUT "an="; a
INPUT "x="; x
v=a
k=n-1
WHILE  k>=0
    PRINT "k="; k
    INPUT "ak="; a
    v=v*x+a
    k=k-1
WEND
PRINT  v
END
```

习题 4

学而时习之

1. 用辗转相除法求 204 和 85 的最大公约数.
2. 有一筐苹果，三个三个地数，多出两个；五个五个地数，多出三个；七个七个地数，多出四个. 你知道这筐苹果至少有多少个吗？
3. 用二分法求方程 $x^3-x-1=0$ 在区间 $[1, 1.5]$ 内的一个近似解（误差不超过 0.001).
4. 已知多项式 $f(x)=x^5+5x^4+10x^3+10x^2+5x+1$，用秦九韶算法求这个多项式当 $x=-2$ 时的值.

温故而知新

5. 分别用辗转相除法和更相减损术求 5 280 和 12 155 的最大公约数.

6. 已知多项式 $f(x)=2x^4-6x^3-5x^2+4x-6$，用秦九韶算法求这个多项式当 $x=5$时的值.

阅读与思考

进位制

进位制也称进制，是人们规定的一种进位方法. 我们通常使用的是十进制，它的特点有两个：一是由 0，1，2，…，9 十个基本数字组成；二是十进制数运算是按“逢十进一”的规则进行的. 一般地，对于任何一种进制——k 进制，就表示某一位置上的数运算时是“逢 k 进一”，其中 k 叫作基数. 十六进制是逢十六进一，二进制就是逢二进一，它们的基数分别是 16 和 2.

我们知道，一个十进制数 110，其中百位上的 1 表示 1 个 10^2，即 100，十位的 1 表示 1 个 10^1，即 10，个位的 0 表示 0 个 10^0，即 0，于是：

$$110=1\times10^2+1\times10^1+0\times10^0.$$

类似地，一个二进制数 110，其中高位的 1 表示 1 个 2^2，即 4，低位的 1 表示 1 个 2^1，即 2，最低位的 0 表示 0 个 2^0，即 0，于是：

$$110_{(2)}=1\times2^2+1\times2^1+0\times2^0.$$

一个十六进制数 110，其中高位的 1 表示 1 个 16^2，即 256，低位的 1 表示 1 个 16^1，即 16，最低位的 0 表示 0 个 16^0，即 0，于是：

$$110_{(16)}=1\times16^2+1\times16^1+0\times16^0.$$

可见，在数制中，各数位上的数字所表示值的大小不仅与该数字本身的大小有关，还与该数字所在的位置有关，我们称此关系为数的位权.

十进制数的位权是以 10 为底的幂，二进制数的位权是以 2 为底的幂，十六进制数的位权是以 16 为底的幂. 数位由高向低，以

通常我们在 k 进制数的右下角注明它的基数以示区别，如 $7\ 342_{(8)}$ 表示八进制数，十进制数一般不需注明.

降幂的方式排列.

其他进制数转换为十进制数的规律是相同的. 如把二进制数按位权形式展开成多项式和的形式，求出其最后的和，就是其对应的十进制数，这种方法简称“按权求和法”.

例 1 把 $100\ 101_{(2)}$ 转换为十进制数.

解 $100\ 101_{(2)}=1\times2^5+0\times2^4+0\times2^3+1\times2^2+0\times2^1+1\times2^0$

$=32+0+0+4+0+1$

$=37.$

思考：一般地，如何将 k 进制数转换成十进制数？

算法分析：

设一个 k 进制数 a 共有 n 位，转换为十进制数 b，按位权形式展开成多项式和的形式就是 k 进制数的右数第 i 位数值 a_i 与 k^{i-1}（$k\in\mathbf{N}$，$2\leqslant k\leqslant9$）的乘积 $a_i\cdot k^{i-1}$（$k\in\mathbf{N}$，$2\leqslant k\leqslant9$）累加的结果，可用循环结构构造算法，具体步骤如下：

S1：输入 a，k，n；

S2：$b=0$，$i=1$；

S3：$b=b+a_i\cdot k^{i-1}$，$i=i+1$；

S4：判断 $i>n$ 是否成立，若是，输出 b 的值；否则返回 S3.

程序框图如图 11－32 所示.

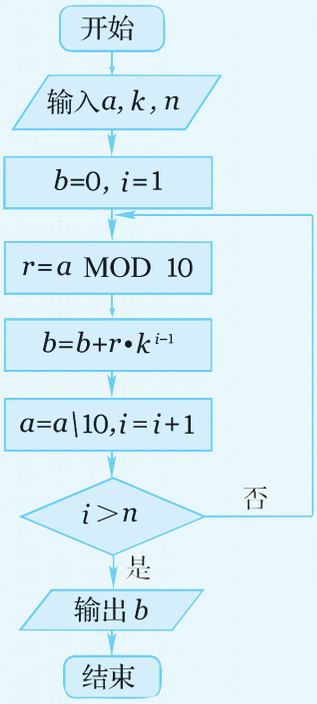


图 11－32

伪代码：

```
INPUT "a, k, n="; a, k, n
b=0
i=1
DO
    r=a MOD 10
    b=b+r*k^(i-1)
    a=a\10
    i=i+1
LOOP UNTIL  i>n
PRINT  b
END
```

例 2 把 115 转换成二进制数.

解 因为 $115=2\times57+1$,

$57=2\times28+1$,

$28=2\times14+0$,

$14=2\times7+0$,

$7=2\times3+1$,

$3=2\times1+1$,

$1=2\times0+1$,

所以 $115=2\times(2\times(2\times(2\times(2\times(2\times(2\times0+1)+1)+1)+0)+0)+1)+1$

$=1\times2^6+1\times2^5+1\times2^4+0\times2^3+0\times2^2+1\times2^1+1\times2^0$.

所以 $115=1\ 110\ 011_{(2)}$.

通过解题过程，你发现这几个算式中的余数与最后结果中的数字排列有什么关系？

思考：一般地，如何将十进制整数转换成 k（$k\in\mathbf{N}$，$2\leqslant k\leqslant9$）进制数？

将上面方法进行推广，就可以得到将十进制整数转换成 k（$k\in\mathbf{N}$，$2\leqslant k\leqslant9$）进制数的一般方法，称为“除 k 取余法”. 其规则是：

(1) 用 k 去除给出的十进制数，取其余数作为转换后的 k 进制数据的最低位数字；

(2) 用 k 去除所得的商，取其余数作为转换后的 k 进制数据的高一位数字；

(3) 重复执行(2)操作，一直到商为 0 结束.

算法分析：

显然可以用循环结构实现上面的算法. 将十进制数 a 转换成 k 进制数 b 的算法具体步骤是：

S1：输入十进制整数 a，基数 k；

S2：$s=a\backslash k$，$r=a$ MOD k；

S3：把得到的余数从右往左排列；

S4：若 $s\neq0$，则 $a=s$，并返回 S2；否则，输出余数的排列得到 k 进制数 b.

程序框图如图 11－33 所示.

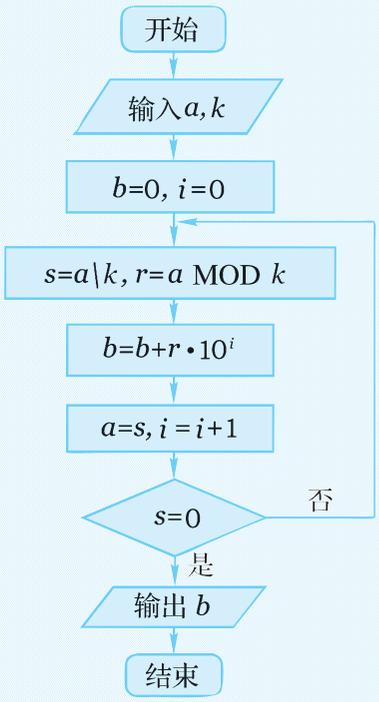


图 11－33

伪代码：

```
INPUT   a，k
b＝0
i＝0
DO
      s＝a \ k
      r＝a MOD k
      b＝b＋r * 10^i
      i＝i＋1
      a＝s
LOOP UNTIL   s＝0
PRINT   b
END
```

掌握了 k 进制数转换成十进制数的按权求和法和十进制数转换成 k 进制数的除 k 取余法，要实现各数制间数的互化就不是难事了.

小结与复习

一、内容提要

算法是数学及其应用的重要组成部分，是计算科学的重要基础. 随着现代信息技术的飞速发展，算法在科学技术、社会发展中发挥着越来越大的作用，并日益融入社会生活的许多方面，算法思想已经成为现代人应具备的一种数学素养. 需要特别指出的是，中国古代数学中蕴涵了丰富的算法思想.

本章第一节了解算法的概念，初步感受算法思想；第二节结合对具体数学实例的分析，会用自然语言表达解决问题的算法，体验程序框图在表达算法中的作用，理解程序框图的三种基本逻辑结构；第三节通过模仿、操作、探索，学习算法的程序语言表达，理解几种基本算法语句——输入语句、输出语句、赋值语句、条件语句、循环语句；第四节通过几个典型算法的分析，体会算法的基本思想以及算法的重要性和有效性，发展有条理地思考与表达的能力，提高逻辑思维的能力.

二、学习要求

1. 算法的含义、程序框图

(1) 通过对解决具体问题过程与步骤的分析，体会算法的思想，了解算法的含义.

（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，理解程序框图的三种基本逻辑结构：顺序结构、条件结构、循环结构．

2．基本算法语句

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——输入语句、输出语句、赋值语句、条件语句、循环语句，体会算法的基本思想．

3．通过阅读典型的算法案例，进一步体会算法思想，体会中国古代数学对世界数学发展的贡献，增强民族自豪感．

三、需要注意的问题

1．算法一方面具有具体化、程序化、机械化的特点，同时又具有高度抽象性、概括性和精确性．对于一个具体算法而言，从算法分析到算法语言的实现，任何一个疏漏或错误都将导致算法的失败．算法是思维的条理化、逻辑化．

2．算法既重视“算则”，更重视“算理”．对于算法而言，一步一步的程序化步骤，即算则固然重要，但这些步骤的依据，即算理有着更基本的作用，算理是算则的基础，算则是算理的表现．

3．算法的操作性很强，因此应当强调动手实践．应充分应用教科书中提供的实例，在解决具体问题的过程中学习一些基本逻辑结构和算法语句．算法内容是将数学中的算法与计算机技术建立联系，形式化地表示算法．为了有条理地、清晰地表达算法，往往需要将解决问题的过程整理成程序框图；为了能在计算机上实现，又要将自然语言或程序框图翻译成计算机语言．因此，如果能上机，算法设计的整个过程就可以得到完整的体现，可以及时看到自己设计的算法的可行性、有效性，这不但可以很好地激发兴趣，而且还

能提高学习效果. 因此，有条件的应尽可能上机尝试.

4. 算法的思想方法应渗透在高中数学课程其他有关内容中. 学习时，要尽可能地运用算法解决相关问题，要体现数学与算法的有机结合，体会算法思想，看到数学在算法设计中的作用，以及掌握算法思想对于提高数学能力的重要性.

复习题十一

学而时习之

1. 阅读下面的程序框图（图 11－34），其运行结果是_____.

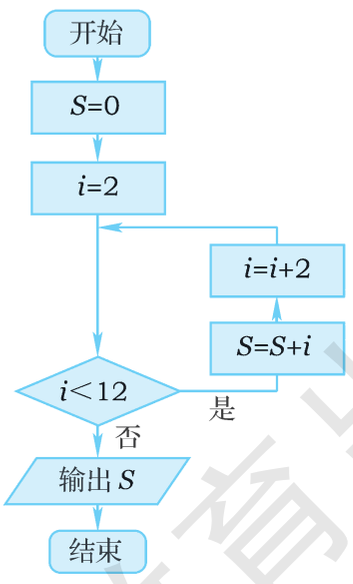


图 11－34

2. 下面的伪代码运行后，输出的结果是_____.

```
x=5
y=-20
IF  x<0  THEN
    x=y-3
ELSE
    y=y+3
END IF
PRINT  x-y, y-x
END
```

3. 已知多项式 $f(x)=0.83x^5+0.41x^4+0.16x^3+0.33x^2+0.5x+1$，用秦九韶算法求这个多项式当 $x=5$ 时的值.

4. 对于任意的正整数 n，画出求 $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}$ 的值的程序框图，并写出伪代码.

5. 设计一个算法，求 $1+2+4+\cdots+2^{49}$ 的值，并画出程序框图.

6. 这是中国古代的一个著名算法案例：一群小兔一群鸡，两群合到一群里，腿数 48，头数 17，多少小兔多少鸡？

温故而知新

7. 某企业 2014 年的年生产总值为 200 万元，经企业改制后预计以后年生产总值将以 5% 的速度递增. 设计一个算法，输出预计年生产总值超过 300 万元的最早年份.

8. 画出程序框图，输入自变量 x 的值，输出函数

$$y=\begin{cases}(x+2)^2 & (x<0),\\ 4 & (x=0),\\ (x-2)^2 & (x>0)\end{cases}$$

的函数值.

9. 用辗转相除法求 228 和 1 995 的最大公约数，并用更相减损术检验你的结果.

上下而求索

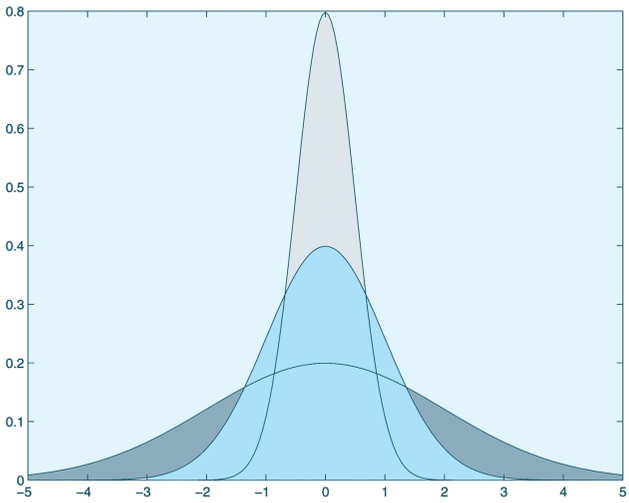
10. 用二分法求$\sqrt{2}$的近似值（误差不超过 0.005）.

11. 中国网通规定：拨打市内电话时，如果不超过 3 min，则收取话费 0.22 元；如果通话时间超过 3 min，则超出部分按每分钟 0.1 元收取通话费，不足1 min 按 1 min 计算. 设通话时间为 t（min），通话费用为 y（元），请设计一个算法，计算通话的费用.

12. 通过网络查阅中国古代数学的优秀算法案例和同学们交流.

第 12 章
统计学初步

数据纷繁沙一盘，　管窥蠡测理当然.
总体抽样图良策，　均值方差求指南.
辨明真假成功路，　分清主次艳阳天.
国运民生关统计，　知风知浪好行船.



高　斯（Gauss）研究测量误差时发现了正态分布曲线，这是对统计学的重要贡献.

现代生活是建立在数据之上的. 没有数据，一切很难想象. 统计学就是研究如何从数据中提取有用信息的科学，内容包括如何收集、整理、描述和分析数据. 基于统计学的数据处理方法称为统计方法.

统计学初步仅仅帮助你了解统计学的一些基本语言，知道一些统计学的基本概念. 学习了统计学初步后你也许会觉得知道一些统计学的初步知识是有用的. 我们认为统计学还可以激发你的智力，给你的生活带来更多的乐趣.

12.1 总体和个体

日常生活中我们总是自觉或不自觉地和总体与样本打交道. 夏天买西瓜时，先要看看这批西瓜甜不甜. 如果瓜甜又不很贵，你可能买一个或两个.

我们可以称这批西瓜是一个总体，单个的西瓜是个体，但是这样就不能强调我们关心的是西瓜的甜度. 因为西瓜的好坏还有其他的指标，例如个的大小，是否新上市的，等等.

在关心这批西瓜的甜度时，我们称单个西瓜的甜度是“个体”，称所有的西瓜的甜度为“总体”. 这样就把西瓜的甜不甜数量化了.

要了解一批西瓜的甜度情况，你不可能品尝每个西瓜. 你只能买一两个尝一尝，然后通过这一两个西瓜的甜度判断这批西瓜的甜度. 这就是用少数个体推断总体. 我们把买的西瓜的甜度称为“样本”，于是你已经可以用样本推断总体.

12.1.1 总体、个体和总体均值

总体、个体和均值是统计学的最基本概念.

要调查全校期中考试的数学成绩时，称全校同学的期中数学成绩是总体，称单个同学的数学成绩是个体.

要调查全校同学期中考试的语文成绩时，称全校同学的语文成绩是总体，称单个同学的语文成绩是个体.

在统计学中，我们把所要调查对象的全体叫作**总体**（population），把总体中的每个成员叫作**个体**（individual）.

总体中个体的某一特征总可以用数量表示. 为了叙述的简单和明确，我们把个体看成数量，把总体看成数量的集合.

调查全校同学期中考试成绩时，指出数学或语文是为了明确总体. 不同的总体不能混为一谈.

总体中个体的数目有时是确定的，有时较难确定．调查全校同学期中考试的数学成绩时，参加考试的人数是明确的，相应总体的个数也就明确了．在调查全国人口的年龄分布时，总体是全国人口的年龄，是明确的，但是个体总数很难精确下来．

在判断一批西瓜甜不甜时，你没有必要知道一共有多少个西瓜．

全校同学期中数学考试成绩的平均值是总体平均，全校同学期中语文考试成绩的平均值也是总体平均．总体平均是总体的指标之一，是我们所关心的指标．

总体平均是总体的平均值，也称为总体**均值**（mean）．

在统计学中，常用 μ（音 miu）表示总体均值．当总体含有 N 个个体，第 i 个个体是 y_i 时，总体均值

$$\mu=\frac{y_1+y_2+\cdots+y_N}{N}.$$

练　习

用 $\bar{x}$ 表示数据 x_1，x_2，…，x_n 的均值，用 b 表示常数．对于数据

$$y_1=x_1+b,\quad y_2=x_2+b,\quad \cdots,\quad y_n=x_n+b$$

的均值 $\hat{y}$，验证：

$$\hat{y}=\bar{x}+b.$$

练习的结论表明：每个数据增加相同的量，数据的均值也增加相同的量．

习题 1

学而时习之

1．简述总体平均的含义．

2．用 $\bar{x}$ 表示观测数据 x_1，x_2，…，x_n 的均值，用 a 表示常数．用 $\hat{y}$ 表示观测数据 $y_1=ax_1$，$y_2=ax_2$，…，$y_n=ax_n$ 的均值时，证明：

$$\hat{y}=a\bar{x}.$$

第 2 题的结论表明：数据同时扩大 a 倍，均值也扩大 a 倍．

3. 对于数据

65	60	59	60	53	60	62	63
70	59	65	66	67	56	63	63
55	57	68	61	56	66	65	57

用 μ_1，μ_2，μ_3 分别表示第 1，第 2，第 3 行的平均值，用 μ 表示全体 24 个数的平均值.

(1) 计算 μ_1，μ_2，μ_3 和 μ；

(2) 是否有 $\mu=(\mu_1+\mu_2+\mu_3)/3$?

12.1.2 样本与样本均值

要了解一盘菜炒得好吃不好吃，你品尝一下就可以下结论了，没有必要等到把菜吃完再做结论. 你品尝的菜就是样本，你的品尝就是把样本进行平均，然后你用样本的平均推断总体的平均.

考察 A 中学高一年级 500 个同学某时间的平均身高 μ. 要得到这 500 个同学的平均身高不是一件很困难的事情，只要了解了每个同学的身高就可以利用公式

$$\mu=\frac{\text{这 500 个同学身高之和}}{500}$$

计算得到.

同一天对每个同学进行一次身高测量可以得到均值 μ 的准确值，但是要花费老师和同学们较多的时间和精力. 统计上解决这类问题的最好方法是进行抽样调查，例如在 500 个同学中只具体测量 50 个同学的身高，用这 50 个同学的平均身高作为总体平均身高 μ 的近似. 这时我们称这 50 个同学的身高为总体的一个样本，称 50 为样本量.

从总体中抽取一部分个体，称这些个体为**样本**（sample).

样本也叫作**观测数据**（observed data).

称构成样本的个体数目为**样本容量**，简称为**样本量**（sample size).

称从总体抽取样本的工作为**抽样**（sampling).

按照上面的定义，总体也是一个样本，称为**全样本**，但是样本一般不是总体.

在考虑身高问题时，对于前述被选中的 50 个同学，用 x_1，x_2，…，x_{50} 分别表示第 1，第 2，…，第 50 个同学在调查日的身高，则这 50 个同学的身高

$$x_1，x_2，\cdots，x_{50}$$

是样本. 用 n 表示样本量，则 $n=50$.

样本均值是样本的平均值，用 $\bar{x}$ 表示.

如果样本是 y_1，y_2，…，y_n，就用 $\bar{y}$ 表示样本均值.

总体均值是总体的指标，是一个固定的量. 但是样本均值依赖于样本的选择，从不同的样本会计算出不同的样本均值. 所以我们说样本均值带有随机性.

和总体均值 μ 做比较后知道，只要抽样合理，对于较大的样本量 n，样本均值 $\bar{x}$ 会接近 μ. 于是，$\bar{x}$ 是总体均值 μ 的近似，所以称为 μ 的**估计**（estimator）.

问题 在考察 A 中学高一年级 500 个同学的平均身高时，决定调查 50 个同学，用这 50 个同学的平均身高作为全年级平均身高的估计. 有 5 名女同学主动承担了这次调查任务，她们每人负责选择了 10 个同学，在 9 月的第一周测量出了所选择的 50 个同学的身高如下（单位：cm）：

156	166	165	157	160	164	162	158	158	164
155	165	165	172	165	158	164	155	161	161
162	160	168	150	164	167	166	165	162	166
165	160	159	160	153	160	162	163	170	159
165	166	167	156	163	163	155	157	168	161

其中 $x_1=156$，$x_2=166$，$x_3=165$，…，$x_{50}=161$ 分别是第 1，第 2，…，第 50 个被选中的同学的身高，样本量 $n=50$.

对上述 50 个测量数据进行平均后得到

$$\bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}=161.86\ (\text{cm}).$$

于是，对全年级平均身高 μ 的估计是 161.86 cm.

上述调查结果公布后，引起了同学们的议论，普遍认为

161.86 cm偏小了. 问题出在哪里呢? 我们在学习抽样调查方法时再解答这个问题.

练 习

用 $\bar{x}$ 表示观测数据 x_1，x_2，…，x_n 的均值，用 a，b 表示常数. 用 $\bar{y}$ 表示观测数据 $y_1=ax_1+b$，$y_2=ax_2+b$，…，$y_n=ax_n+b$ 的均值时，证明 $\bar{y}=a\bar{x}+b$.

习题 2

学而时习之

1. 简述样本均值和总体均值的关系.

2. 当样本 x_1，x_2，…，x_n 中有 n_1 个 y_1，n_2 个 y_2，…，n_k 个 y_k 时，验证：

$$\bar{x}=\frac{n_1y_1+n_2y_2+\cdots+n_ky_k}{n}.$$

3. 将某调查公司得到的 20 个数据从小到大排列后得到如下数据：

680　680　680　680　685　685　685　685　685　690
690　690　690　690　690　696　696　696　700　700

计算样本均值 $\bar{x}$.

4. 将一个总体中的 $5n$ 个个体平均分成 n 份，每份 5 个个体. 先计算每份的均值，得到 n 个均值. 这 n 个均值的平均值是否等于总体均值? 证明你的结论.

12.1.3 方差和标准差

拔河比赛是一项有益于身体健康和增进团结的体育活动，某居民区的 2 号楼和 6 号楼决定进行拔河比赛. 2 号楼组成 2 号队，6 号楼组

成 6 号队，每队 15 人. 参加比赛时，2 号队的年龄（单位：岁）组成是

8，8，9，9，9，10，10，10，

10，12，57，61，62，65，65.

6 号队的年龄（单位：岁）组成是

26，26，26，26，26，27，27，27，

27，27，28，28，28，28，28.

这两个队的平均年龄都是 27 岁，但是各队一出场，大家就基本能够判断出比赛的结果了. 2 号队的年龄相差悬殊，是老爷爷带小朋友；6 号队的年龄整齐，都是中青年. 看来只靠平均年龄无法判定拔河队的实力，还需要有一个能衡量年龄的整齐程度的量. 这个量就是要学习的方差.

1. 总体方差.

当 y_1，y_2，…，y_N 是总体的全部个体，μ 是总体均值时，称

$$\sigma^2=\frac{(y_1-\mu)^2+(y_2-\mu)^2+\cdots+(y_N-\mu)^2}{N}$$

是总体的平均平方误差，简称为**总体方差**或**方差**（variance）.

总体方差 σ^2 描述了总体中的个体向总体均值 μ 的集中程度：方差越小，表示个体与 μ 的距离越近，个体向 μ 集中得越好.

总体方差 σ^2 也描述了总体中个体的整齐程度或波动幅度，方差越小，表示个体越整齐，波动越小.

例 1 同一年级的甲班有 35 个同学，乙班有 37 个同学. 期中考试后，数学的平均成绩分别是 79.4 分和 82.7 分. 方差分别是 68.6 和 148.8. 如何就这次考试的结果评价这两个班的数学课的学习情况？

解 从平均分上看，乙班的数学平均成绩好于甲班，但是从成绩的整齐程度方面看，甲班好于乙班. 甲班的分数比乙班的分数更集中.

下面是这两个班数学考试成绩从低到高的排列结果：

甲班成绩

60	65	65	66	70	71	71	72	72	75	76	78
79	79	80	80	81	81	81	82	82	83	83	83
84	84	85	85	85	85	86	87	91	95	98	

σ 音 sigma，σ^2 读作 sigma 方.

可以计算 2 号拔河队年龄的方差是 $\sigma_2^2\approx 616.3$

6 号拔河队年龄的方差是 $\sigma_6^2\approx 0.7$. 相差太悬殊了.

乙班成绩

52	56	64	67	67	67	69	73	74	74	77	77	80
82	83	83	83	83	85	85	86	87	88	88	88	88
89	90	90	92	97	98	99	99	100	100	100		

从中看出，甲班没有同学不及格，也没有同学得满分；乙班有同学得满分，但是也有同学不及格．甲班的数学成绩更整齐．

2．样本方差．

知道总体方差后，可以作出更好的比较结果．

给定 n 个观测数据 x_1，x_2，…，x_n，用 $\bar{x}$ 表示这 n 个数据的均值．称

$$s^2=\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]$$

为这 n 个数据的**样本方差**，也简称为**方差**．

样本方差 s^2 是描述观测数据关于样本均值 $\bar{x}$ 发散程度的指标，也是描述数据的发散程度或波动幅度的指标．

样本方差依赖于样本的选取，也带有随机性．样本方差是总体方差的估计．

例 2 一箱内有 50 个苹果，净重 10 kg，则苹果的平均质量是

$$\frac{10\,000}{50}=200\ (\mathrm{g}).$$

要了解这箱苹果的整齐程度，就需要估计这箱苹果质量的方差 σ^2．

解 从中抽出 10 个，测得这 10 个苹果的质量（单位：g）是

201，218，187，192，193，198，202，194，176，291．

样本均值是

$$\bar{x}=\frac{201+218+\cdots+291}{10}=205.2\ (\mathrm{g}).$$

样本方差是

$$s^2=\frac{1}{10}[(201-205.2)^2+(218-205.2)^2+\cdots+(291-205.2)^2]$$

$$=923.76\ (\mathrm{g}^2).$$

我们可以用样本方差 $s^2=923.76$（g^2）作为总体方差 σ^2 的估计．

方差也可以通过下列公式来计算.

$$s^2=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-\bar{x}^2.$$

将方差定义中的每个$(x_i-\bar{x})^2$展开,再利用

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

得到

$$\begin{aligned}s^2&=\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]\\&=\frac{1}{n}[(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2(x_1+x_2+\cdots+x_n)\bar{x}+n\bar{x}^2]\\&=\frac{1}{n}[(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-2n\bar{x}\bar{x}+n\bar{x}^2]\\&=\frac{1}{n}[(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-n\bar{x}^2]\\&=\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-\bar{x}^2.\end{aligned}$$

3. 标准差.

在例 2 中，数据的单位是 g，样本方差 s^2 的单位是 g^2，和数据的单位不一致. 为了使描述数据的波动幅度的量和数据的单位一致，我们再引入标准差.

标准差（standard deviation）是方差的算术平方根；

如果 s^2 是样本方差，就称 $s=\sqrt{s^2}$ 是**样本标准差**；

如果 σ^2 是总体方差，就称 $\sigma=\sqrt{\sigma^2}$ 是**总体标准差**.

在例 2 中，10 个苹果质量的标准差是 $s=\sqrt{923.76}\approx30.39$（g）. 其单位和数据的单位一致.

当数据带有单位时，标准差的单位是和数据的单位一致的. 标准差也是描述数据发散程度或波动幅度的指标. 样本标准差是总体标准差的估计.

给定数据 x_1，x_2，…，x_n 和均值 $\bar{x}$. 由方差计算公式知道，标准差 s 可以由下面的公式之一计算.

$$s=\sqrt{\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]},$$

$$s=\sqrt{\frac{1}{n}(x_1{}^2+x_2{}^2+\cdots+x_n{}^2)-\bar{x}^2}.$$

例 3 比赛中，甲乙两位射击运动员分别进行了 10 次射击，成绩（单位：环）分别如下：

甲：9.5　9.9　9.9　9.9　9.8　9.7　9.5　9.3　9.6　9.6

乙：9.4　9.3　9.5　9.0　9.1　9.8　9.7　9.5　9.3　9.4

问：哪个运动员平均水平高？哪个运动员水平更稳定？

解 用 $\bar{x}$，s_x 和 $\bar{y}$，s_y 分别表示甲和乙成绩的均值和标准差．经过计算得到

$$\bar{x}=9.67,\quad s_x=0.195\ 2,\quad \bar{y}=9.4,\quad s_y=0.232\ 4.$$

因此，甲的平均水平和稳定性都比乙好．

练 习

练习的结论表明：观测数据同时加上相同的常数后发散程度不变，方差也不变．

用 s_x^2 表示 x_1，x_2，…，x_n 的方差，用 b 表示常数，用 s_y^2 表示 y_1，y_2，…，y_n 的方差．当 $y_1=x_1+b$，$y_2=x_2+b$，…，$y_n=x_n+b$ 时，验证 $s_y^2=s_x^2$．

习题 3

学而时习之

1. 下面的数据是 1900—1936 年奥林匹克男子跳高比赛金牌获得者的跳跃高度（单位：cm）．计算均值、方差和标准差（精确到小数点后两位）．

年　份	高　度	年　份	高　度
1900	190.0	1904	190.3
1908	190.5	1912	193.0
1920	193.5	1924	198.1
1928	194.1	1932	197.1
1936	202.9		

2. 某连锁超市销售部收到甲乙两厂家送来的质地相同的白糖各 10 包，测量后得到甲乙两厂家白糖的质量（单位：g）分别是：

甲厂	501	500	499	500	502	500	500	501	499	498
乙厂	497	501	500	502	499	501	503	500	500	497

销售部应当销售哪家的白糖？

3. 某公司希望能为飞机制造公司提供零部件，在向飞机制造公司推荐自己的生产能力时，应当重点明确以下哪些内容（　　）

（A）所生产部件的平均规格符合标准

（B）所生产部件的规格的方差不小于某个数

（C）所生产部件的规格的方差不大于某个数

（D）能够按时供货

4. 对一本书进行校稿前，抽查了其中的 21 页. 将排版时输入错误的情况总结如下：

输入错误数	1	7	4	0	11	6	2
出现总页数	3	6	3	2	2	1	4

（1）计算每页的平均输入错误数；

（2）计算样本方差（精确到 0.01）；

（3）计算样本标准差（精确到 0.01）.

温故而知新

5. 当观测数据 x_1，x_2，…，x_n 的样本方差 $s^2=0$ 时，证明所有的 x_i 相同.

6. 当数据 x_1，x_2，…，x_n 同时增加到原来的 a 倍时，证明方差增加到原来的 a^2 倍.

7. 当数据 x_1，x_2，…，x_n 同时增加到原来的 a 倍时，证明标准差增加到原来的 $|a|$ 倍.

12.2 抽样调查方法

在日常生活中人们总是自觉或不自觉地应用抽样方法，例如在市场上买花生或瓜子时总要先抓几颗看看是否饱满，干燥．在厨房做饭的过程中经常要取一点尝尝咸淡．

在考察锅里汤的味道时，没有必要把汤喝完，只要把汤搅拌均匀，从中品尝一勺就可以了．注意无论这锅汤有多少，只要一勺就够了．这就是窥一斑而知全豹．

记住上面的例子是大有好处的，因为它提供了抽样调查方法的最重要信息．

在刚加盐的地方舀出的汤做样本，你会作出汤太咸了的错误结论．

第一，“把汤搅拌均匀”是说明抽样的随机性，没有抽样的随机性，样本就不能很好地反映总体的情况．

第二，“品尝一勺”指出了选取的样本量不能太少，也不必太大．太少了不足以品出味道，品尝一大碗也没有必要．

抽样调查是相对于普查而言的．其含义是从总体中按一定的方式抽出样本进行考察，然后用样本的情况来推断总体的情况．

第三，“无论这锅汤有多少，只要一勺就够了”．这里体现出抽样调查的如下基本性质：总体个数增大时，样本量不必跟着增大．

抽样调查的必要性

在评价 1 000 台同型号的微波炉的平均工作寿命 μ 时，预备从中抽取 n 台进行工作寿命的测量试验，用这 n 台微波炉的平均工作寿命估计总体的平均工作寿命 μ．

这里，总体是 1 000 台微波炉的工作寿命，样本量是 n，被选中的微波炉的工作寿命构成样本．样本平均 $\bar{x}$ 是总体均值 μ 的估计．

根据喝汤的经验，没有必要调查很多微波炉的工作寿命．

在正确抽样的前提下，样本量越大，$\bar{x}$ 越接近总体均值 μ．但是，较大的样本量造成的损失很大．因为这 n 台微波炉做完寿命试验后就报废了．在本问题中要想得到真正的总体均值 μ 是不可能的，除非把这 1 000 台微波炉都拿来做工作寿命试验，报废掉这 1 000 台微波炉．

窥一斑而知全豹．

在很多实际问题中，采用抽样的方法来确定总体性质不仅是必要的，也是必须的．

12.2.1 随机抽样

在 12.1.2 节的问题中，通过选取和测量 50 个同学的身高，得到了总体平均身高 μ 的估计 $\bar{x}=161.86$（cm）. 结果公布后，同学们普遍反映估计值偏低. 其原因是什么呢？

造成估计值偏低的原因是抽样方案设计不合理.

原因在于女同学在选择调查对象时更倾向于，或更方便选择到女同学，所以 50 个同学的样本中女生身高占了绝大多数. 这样就解释了估计值偏低的原因.

如何设计抽样方案才能得到满意的估计值呢？

如果请 5 名男同学做抽样调查，就有可能得出估计值偏高的结果.

在对总体的情况不清楚的时候，最好的抽样方案应当将总体中的个体一视同仁：每个个体被抽中的机会相同.

如果总体中的每个个体都有相同的机会被抽中，就称这样的抽样方法为**随机抽样**方法.

人们经常用“**任取**”，“**随机抽取**”或“**等可能抽取**”等来表示随机抽样.

例 口袋中有质地相同的小球 10 个，分 3 种颜色. 从中无放回地随机抽取 1 个，共抽取 n（$\leqslant 10$）个，这种抽样的方法被称为无放回地随机抽样. 从袋中每次随机抽取一球记录颜色后放回，共抽取 n 次，这样的抽样方法被称为有放回地随机抽样. 这两种随机抽样有什么区别吗？

解 无放回随机抽样下，同一个小球不会被抽中两次. 而有放回地随机抽样下，同一个小球可能被抽中多次. 当样本量 $n=10$，采用无放回随机抽样就可以完全了解袋中小球的颜色分布情况，采用有放回地随机抽样还不能对袋中小球的颜色分布作出准确判断.

随机抽样又分为无放回地随机抽样和有放回地随机抽样. 无放回地随机抽样指在总体中抽出一个个体后，下次在余下的个体中再进行随机抽样. 有放回地随机抽样指抽出一个个体，记录下抽到的结果后放回，摇匀后再进行下一次随机抽样.

一般地，设一个总体含有 N 个个体，从中逐个不放回地抽取 n

($n \leqslant N$)个个体为样本，如果每次抽取时总体内的各个个体被抽到的机会都相等，则把这样的抽样方法称为**简单随机抽样**.

简单随机样本指简单随机抽样得到的样本.

在没有特殊声明时，**所有的随机抽样都是指简单随机抽样**.

试验和理论都证明：在随机抽样下，样本均值 $\bar{x}$ 是总体均值 μ 很好的估计，样本标准差 s 是总体标准差 σ 很好的估计. 在样本量不大时，增加样本量可以比较好地提高估计的精确度.

在相同的总体中和相同的样本量下，简单随机抽样得到的结果比有放回的随机抽样得到的结果要好. 但是当总体的数量很大，样本量相对总体的数量又很小时，这两种抽样方法得到的结果是相近的.

在 12.1.2 节的问题中，实现简单随机抽样的方法是先将 500 个同学从 1 到 500 进行编号，然后将 500 张由 1 到 500 编号的小纸片放入一个大纸箱充分地摇匀，最后从纸箱中无放回地抽取 50 张纸片. 纸片上的号码就是被选中的同学的号码. 纸片上的这 50 个数被称为**随机数**（random number）.

摇匀后，一次取出 50 个和无放回地依次抽取 50 个的效果是相同的：都是没有重复的随机抽样.

随机数可以利用计算机产生. 下面是用计算机在 1 至 500 中随机抽取的 50 个随机数.

50 个随机数

476	116	304	243	446	382	229	17	411	223
308	396	461	370	89	203	468	459	206	447
29	177	407	5	95	70	102	109	302	137
100	8	374	224	466	233	210	424	263	106
337	420	10	341	190	416	252	355	215	153

现在我们继续解决 12.1.2 节中的问题. 将 A 中学高一年级的 500 个同学从 1 到 500 进行编号，按照上述随机数表中的号码选取出 50 个同学，在 9 月的某一天测量他们的身高（单位：cm）如下：

165	165	169	162	165	163	165	177	162	156
169	159	166	161	170	170	164	151	167	177
173	155	165	161	159	157	168	163	166	163
163	168	171	181	156	159	173	163	156	168
177	171	180	169	179	163	158	164	174	158

用这 50 个观测数据计算出的样本均值是

$$\bar{x} = 165.68 \text{ (cm)}.$$

于是，高一年级同学平均身高的估计是 165.68 cm.

由于这次抽样是通过随机抽样完成的，因而避免了有偏的结果．这次抽样调查的结果得到了同学们的认可．

关于历史上采取不正确的抽样方案而导致调查结论严重失真的教训，可参阅本小节的“阅读与思考”．

练 习

1. 调查 1 000 支日光灯管的平均使用寿命时，随机抽样方法是有放回地随机抽样吗？
2. 调查某市出租车司机的月平均收入时，在街面上进行随机抽样调查，得到的样本是简单随机样本吗？

习题 4

学而时习之

1. 简述什么是简单随机抽样．
2. 简述什么是有放回地随机抽样．
3. 举例说明随机抽样方法中“随机”的必要性．
4. 在某一个海域对海豚的体重进行抽样调查时，应采用有放回地还是无放回地捕获抽样？
5. 在调查某个城市的家庭年平均收入时，能否只在该市的娱乐场所（如电影院、歌剧院、游乐场、健身馆等）进行随机抽样？原因是什么？能否只在该市的公共汽车站进行随机抽样？原因是什么？

阅读与思考

《文学摘要》的破产

1936 年是美国总统选举年. 这年**罗斯福**（Roosevelt）任美国总统期满，参加第二届的连任竞选，对手是堪萨斯州州长**兰登**（Landon）. 当时美国刚从经济大萧条中恢复过来，失业人数仍高达 900 多万，人们的经济收入下降了 1/3 后开始逐步回升. 当时，观察家们普遍认为罗斯福会当选. 而美国的《文学摘要》杂志的调查却预测兰登会以 57%对 43%的压倒性优势获胜.《文学摘要》的预测是基于对 240 万选民的民意调查得出的. 自 1916 年以来，在历届美国总统的选举中《文学摘要》都作了正确的预测.《文学摘要》的威信有力地支持着它的这次预测.

但是选举的结果却是罗斯福以 62%对 38%的压倒性优势获胜. 此后不久《文学摘要》杂志就破产了.

要了解《文学摘要》预测失败的原因就必须检查他们的抽样调查方案.《文学摘要》是将问卷寄给了 1 000 万选民，这些选民的地址是在诸如电话簿、俱乐部会员名单等上面查到的.

只收回了 240 万张问卷.

分析 1936 年只有大约 1/4 的家庭安装了电话. 由于有钱人才更有可能安装家庭电话和参加俱乐部，所以《文学摘要》的调查方案漏掉了那些不属于俱乐部的穷人和没有安装电话的穷人，这就导致了调查结果有排除穷人的偏向.

在 1936 年，由于经济开始好转，穷人普遍有赞同罗斯福当选的倾向，富人有赞同兰登当选的倾向.《文学摘要》的调查结果更多地代表了富人的意愿，导致了预测的失败.

评论 抽样的方案应当公平地对待每一位选民和每一个群体，以便得到选民的真实情况. 将哪一个群体排除在外的抽样方案都会

导致有偏的样本，从而导致错误的结论.

同一年，刚刚成立的盖洛普调查公司正确地预测了罗斯福获胜. 以后，盖洛普公司做过多次美国总统大选的民意调查. 由于采取了正确的抽样设计方案，在调查人数不是很多的情况下，预测的结果都是成功的.

12.2.2 调查问卷的设计

在随机抽样的前提下，具体抽样调查的实施方式也会影响调查的结果.

例 （敏感问题调查）某一个住宅区有2 060个家庭，调查人员已经拿到了所有住户的门牌号码. 在调查本住宅区存在家庭暴力的家庭比例时，因为经费的原因，只能调查 200 个家庭. 现在已经用随机抽样的方式抽出了 200 个门牌号码，问：以下哪种调查方案较好？

(A) 派调查员用询问记录的方式登门调查这 200 个家庭.

(B) 登门调查这 200 个家庭前先准备 200 张问卷和若干支笔. 只需被调查者在下面的匿名问卷上打勾，然后请被调查者自己将问卷放入调查员的书包.

> 问卷：请选择［有家庭暴力］ ［无家庭暴力］
> 我们承诺没有人知道你的回答是什么.

(C) 采用和(B)相同的调查方案，但是告诉被调查者要调查 200 个家庭，而且让他们将答卷折叠后投入随身携带的封闭投票箱，全部调查完毕后再开箱统计.

(D) 将(B)中的 200 张问卷投入抽中的 200 个家庭的信箱，请他们在规定的时间内将答卷放入指定的投票箱.

解 由于家庭暴力是不光彩的家庭隐私，所以调查时应当让被调查者知道他的回答是得到严格保密的. 只有这样，调查才有可能获得事实真相. 按照这个原则，方案(C)是其中较好的方案.

方案(A)忽略了被调查者的感受，容易得到有偏的结果.

方案(B)记得带笔方便了调查，但是忽略了被调查者因不信任而不回答事实真相.

方案(D)的缺点是只能收回少量的问卷.

在抽样调查中，调查的方式方法也是非常重要的. 无论是当面调查还是问卷调查都应当做到以下两点：

(1) 提问的内容要简单明确. 提问太长，会给回答带来困难和引

起被调查者的反感，不利于得到正确的回答.

（2）用词要确切，通俗易懂，有礼貌和不用引导词语.

例如：

“您用什么牌子的牙膏?”时间范围不明确，应改为“您现在用什么牌子的牙膏?”

“很多人都要买汽车，您呢?”带有引导性，应改为“您最近打算买汽车吗?”

练 习

某校高中一年级有 6 个班，每班有 43 个同学，其中一个班是特长班. 该年级的全体同学已经将各自的学号写在规格相同的小纸条上，放入一个大纸箱中摇匀. 现在校长要通过抽样的方法调查该年级学生参加课外体育锻炼的情况，规定只对 36 个同学进行详细了解. 以下抽样方法正确的是（　　）

（A）从纸箱中无放回地随机取出 36 个学号，选取这 36 个学号的同学

（B）各班的班主任在自己的班上点出 6 个同学

（C）从纸箱中有放回地随机抽取 36 个学号，选取这 36 个学号的同学

（D）从特长班中用随机抽样的方法选取 36 个同学

习题 5

学而时习之

1. 用随机抽样的方法，在你的语文书中抽查 10 页，回答以下问题：

（1）你的抽样是如何进行的，采用的是简单随机抽样还是有放回地随机抽样?

（2）这 10 页中，平均每页有多少个句号?

（3）你对语文书平均每页句号个数的估计是多少?

（4）如果另外的同学抽查了 20 页，你认为谁的估计更准确?

2.（数学实践）在抽样调查本校同学的手机个人拥有率、家庭汽车拥有率和每天完成作业所用的时间时，规定样本量为 $n=100$. 请同学们设计一个合理的调查方案和一份调查问卷（参考 P.76 例题中的问卷），并具体实施一次抽样调查工作.

12.2.3 分层抽样和系统抽样

分层抽样

要了解一盘菜炒得好吃不好吃，一般只要随机品尝几口就可以下结论了，没有必要等到把菜吃完再作出结论.

但是如果品尝的是西红柿炒鸡蛋，你进行随机抽样品尝就容易只品尝到西红柿或只品尝到鸡蛋，这对于你作出正确的判断是不利的. 你应当随机品尝一下西红柿，再随机品尝一下鸡蛋，然后进行综合评价. 这种品尝方法就是分层抽样方法.

例 1 某市进行家庭年收入调查时，分别对城镇家庭和农村家庭进行调查. 在全部城镇的 85 679 户中无放回地随机抽取了 350 户，在全部农村的 275 692 户中无放回地随机抽取了 360 户. 调查结果为：城镇家庭年平均收入是 35 612 元，农村家庭年平均收入是 5 623 元. 试计算该市家庭年平均收入.

解 这里遇到了两个分总体 A_1 和 A_2，第一个分总体 A_1 是所有城镇家庭的年收入，第二个分总体 A_2 是所有农村家庭的年收入. 用 A 表示该市所有家庭的年收入时，总体 A 是两个分总体 A_1 和 A_2 的并.

用 $\bar{x}_1$ 表示来自总体 A_1 的样本均值，用 $\bar{x}_2$ 表示来自总体 A_2 的样本均值，则 $\bar{x}_1=35\ 612$，$\bar{x}_2=5\ 623$.

A_1 在 A 中所占的比例是

$$W_1=\frac{85\ 679}{85\ 679+275\ 692}\approx 0.237\ 1.$$

A_2 在 A 中所占的比例是

$$W_2=\frac{275\ 692}{85\ 679+275\ 692}\approx 0.762\ 9.$$

A 的总体均值 μ 的估计是

$$\begin{aligned}\overline{X} &= W_1\overline{x}_1 + W_2\overline{x}_2\\ &= 0.237\ 1\times 35\ 612 + 0.762\ 9\times 5\ 623\\ &\approx 12\ 733\ (\text{元}).\end{aligned}$$

于是该市平均年家庭收入的估计是 12 733 元.

把总体 A 分成 L 个互不相交的子总体：

$$A = A_1 + A_2 + \cdots + A_L,$$

称这些子总体为层，称 A_i 为第 i 层. 然后按照一定的比例，对各层独立地进行简单随机抽样，然后将各层抽样出来的个体合在一起作为样本. 这种抽样方法称为分层抽样.

用 N 表示总体 A 的个体总数，用 N_i 表示第 i 层的个体总数时，有

$$N = N_1 + N_2 + \cdots + N_L.$$

我们称

$$W_i = \frac{N_i}{N}\ (i = 1,\ 2,\ \cdots,\ L)$$

为第 i 层的**层权**（weight）.

用 μ 表示 A 的总体均值. 对 $i = 1,\ 2,\ \cdots,\ L$，用 $\overline{x}_i$ 表示从第 i 层抽出样本的样本均值. 我们称

$$\overline{X} = W_1\overline{x}_1 + W_2\overline{x}_2 + \cdots + W_L\overline{x}_L$$

是总体均值 μ 的**简单估计**.

分层抽样是一种常用的抽样方法，有如下的特点：

（1）分层抽样在获得总体均值估计的同时，也得到各层的均值估计. 在例 1 中，不但得到了 A 的均值估计，还得到了 A_1 和 A_2 的均值估计.

（2）将差别不大的个体分在同一层，使得分层抽样得到的样本更具有代表性，从而提高估计的准确度.

（3）抽样调查的实施更加方便，调查数据的收集、处理也更加方便.

系统抽样方法

例 2 在调查某居民住宅区的 999 户住户对住宅区的环境满意程度时，是按照 1∶14 的比例进行抽样调查，试计算样本均值.

解 先将这 999 户按门牌号码的顺序依次编号，每个号对应一户的门牌号码.

1	2	3	4	5	6	7	…	13	14
15	16	17	18	19	20	21	…	27	28
29	30	31	32	33	34	35	…	41	42
…	…	…	…	…	…	…	…	…	…
981	982	983	984	985	986	987	…	993	994
995	996	997	998	999					

在 1～14 中随机抽取一个数字，如果抽到 7，就调查排在第 7 列的所有家庭，请这些家庭对小区环境的满意程度打分. 分数分为 1，2，3，4，5 级. 第 7 列有 71 户，所以样本量 $n=71$. 这 71 户的平均分是样本均值. 用样本均值作为全体住户对小区环境的平均分的估计.

用 x_i 表示这 71 户中第 i 户的打分，样本均值是

$$\bar{x}=\frac{x_1+x_2+\cdots+x_{71}}{71}.$$

我们称上面的抽样方法为系统抽样法.

如果总体中的个体按一定的方式排列，在规定的范围内随机抽取一个个体，然后按照制定好的规则确定其他个体的抽样方法称为**系统抽样方法**（systematic sampling method）.

最简单的系统抽样方法是取得一个个体后，按相同的间隔抽取其他个体.

系统抽样方法的主要优点是实施简单，只需先随机抽取第一个个体，以后按规定抽取就可以了. 系统抽样方法不像随机抽样方法，随机抽样方法每次都要随机抽取个体.

练 习

A 中学高一年级的 500 名同学中有 218 名女生，在调查全年级同学的平均身高时，预备抽样调查 50 个同学. 请你做以下工作，并回答以下问题.

(1) 设计一个合理的分层抽样方案.

(2) 你的设计中，第 1 和第 2 层分别是什么？

(3) 分层抽样是否在得到全年级同学平均身高的估计时，还分别得到了男生和女生的平均身高的估计？

习题 6

学而时习之

1. 调查你使用的语文书每页平均有多少个“。”，调查的比例是全书页数的 1/10.

(1) 设计一个系统抽样方法；

(2) 具体实施你的系统抽样方法，写出调查的样本，样本量；

(3) 计算样本均值；

(4) 你估计全书每页平均有多少个“。”？

(5) 把你的结果和其他同学的结果进行比较，对比较的结果给出简单的分析.

2. 调查全班 49 个同学的平均身高时，决定采用系统抽样方法抽取 14 个样本作平均. 请 49 个同学按高矮顺序排列，身高（单位：cm）情况如下：

150	153	155	155	155	156	156
157	157	158	158	158	159	159
160	160	160	160	160	161	161
161	162	162	162	162	163	163
163	164	164	164	164	165	165
165	165	165	165	166	166	166
166	167	167	168	168	170	172

应当怎样抽样，才能避免抽样偏差（　　）

(A) 前两行的样本平均

(B) 后两行的样本平均

(C) 前两列的样本平均

(D) 两个对角线上数据的样本平均

3. 计算问题 2 中的总体平均和你所选用方法的样本平均（精确到小数点后 1 位）.

大量的原始数据如果不经过有效的分析、整理，并通过适当的形式表示出来，就好比一堆没有经过冶炼的矿物，没有什么用途.

分析整理数据的方法之一是用图表把它们表达出来. 图表中包含的信息极多，因为数据中的大量信息都可以概括在图表内. 图表使人一目了然. 一幅图或一张表有时候往往胜过语言的表述.

12.3 用样本分布估计总体分布

无论是从抽样调查，还是从科学实验，工农业生产中得到的数据，在统计学中都被称为观测数据或样本. 观测数据也简称为**数据**，数据的个数被称为样本量.

在实际问题中，样本量往往是比较大的，这时数据中的主要信息隐藏在背后. 要从数据中得到这些信息，必须对观测数据进行整理. 下面是几种常用的数据整理方法.

12.3.1 频率分布表

当样本量是 n 的观测数据中有 n_i 个 y_i 时，我们称

$$f_i=\frac{n_i}{n}$$

是 y_i 出现的**频率**（frequency），简称为 y_i 的频率. 例如数据

2，2，2，2，3，3，3，5，5，5

中，2 的频率是 4/10＝0.4，3 的频率是3/10＝0.3，5 的频率是 3/10＝0.3. 频率也可以用百分数表示. 在上面的例子中，2 的频率是 40%，3 的频率是 30%，5 的频率是 30%.

案例 自 1500 年至 1931 年的 $n=432$ 年间，比较重要的战争（简称为战争）在全世界共发生了 299 次. 以每年为一个时间段的记录如下：

表 12.1

爆发的战争数 i	爆发 i 次战争的年数 n_i	频率 $f_i=n_i/n$
0	223	51.6%
1	142	32.9%
2	48	11.1%
3	15	3.5%
4＋	4	0.9%
总　计	432	100%

其中第一行的 0，223，51.6%表示在 432 年中有 223 年发生战争的次数是 0，发生的频率是

$$f_1=\frac{223}{432}\approx 51.6\%;$$

第二行的 1，142，32.9%表示在 432 年中有 142 年发生战争的次数是 1，发生的频率是

$$f_2=\frac{142}{432}\approx 32.9\%;$$

……

第五行表示在 432 年中有 4 年发生战争的次数是大于等于 4 次的，发生的频率是

$$f_5=\frac{4}{432}\approx 0.9\%.$$

我们称表 12.1 是观测数据的**频率分布表**（frequency distribution table）. 它简化了 432 年中有关战争爆发的 432 个观测数据，帮助我们更清楚地看到战争爆发的特征和规律.

制作频率分布表时，先将数据从小到大排列，然后将排列后的数据进行分段，相等的数据分在同一段内. 每段中的数据被称为一组数据，所以我们又把分段称为**分组**. 一般来讲，当样本量是 n，可以参照下面的经验公式将数据分成大约

$$K=1+4\lg n$$

段. 但是这里的经验公式只对分段起参考作用. 实际应用时，应当根据样本量的大小和数据的特点以及分析的要求灵活确定.

让我们通过例子学习频率表的制作方法.

例 下面是某城市公共图书馆在一年中通过随机抽样调查得到的 60 天的读者借书数，数据已经从小到大排列，请制作频率分布表.

213	230	239	289	291	301	308	310	311	312
318	318	337	343	344	348	349	351	360	362
368	372	374	379	383	385	390	393	396	399
400	404	406	425	429	430	436	438	440	441
444	446	450	453	456	458	471	473	475	483

484　495　498　498　521　524　549　556　568　584

解　数据中的最小值是 213，最大值是 584．这 60 个数据就散布在闭区间［213，584］中．取一个略大的区间［200，600］，它的端点都是整数．用经验公式计算出

$$K=1+4\lg n=1+4\lg 60\approx 8.$$

将［200，600］八等分，排在表的第一列．计算出数据落入各段的个数 n_i，填入第二列．计算出数据落入各段的频率

$$f_1=\frac{3}{60}=5\%,\ f_2=\frac{2}{60}\approx 3.3\%,\ \cdots,\ f_8=\frac{3}{60}=5\%,$$

依次填入第三列．最后将各列之和填入最后一行，得到频率分布表 12.2.

表 12.2

借出书数 i	发生次数 n_i	f_i＝发生频率
［200，250］	3	5%
（250，300］	2	3.3%
（300，350］	12	20%
（350，400］	14	23.3%
（400，450］	12	20%
（450，500］	11	18.3%
（500，550］	3	5%
（550，600］	3	5%
总　　计	60	99.9%

由于计算频率时四舍五入引起计算误差，频率之和可能是 1 的近似．

从上述频率分布表可以方便地分析出以下结果：

有 8.3%的工作日借出的图书少于等于 300 册；

有 63.3%的工作日借出图书的数量在 301 至 450 册之间；

有 48.3%的工作日借出的图书在 400 册以上；

只有 10%的工作日借出的图书多于 500 册．

当总体是全年每个工作日的借书数量时，上述结果可以作为对总体的推测．

从上例可以总结出制作频率分布表的一般步骤如下：

第一步：将数据从小到大排列，将排列后的数据进行分段，相等的数据必须分在同一段内．每段中的数据被称为一组数据，所以我们

又把分段称为**分组**.

分段的多少应当适中. 分段过多，数据过于分散，不利于看出数据的特征和规律；分段过少也不利于看到数据的特征和规律. 当样本量是 n，可以参照经验公式将数据分成大约 $K=1+4\lg n$ 段.

也可以用 $K=\sqrt{n}$ 作经验公式.

第二步：决定各段的长短. 在许多情况下，为了方便，除去第一和最后的两段，可以把其他各段的长度取作相同. 还应当把各段的端点确定在便于记忆的数值上. 为了达到以上目的，第一段的左端点可以比数据的最小值小一些，最后一段的右端点可以比数据的最大值大一些.

第三步：绘制频率分布表的第一列（参考上例）.

第四部：计算每段内数据的个数 n_i，填入表格的第二列.

第五步：计算数据落在第一段内的频率 f_i，填入表格的第三列.

第六步：将第二、第三列之和填入最后一行.

说明：由于频率分布表的制作没有统一的数据分段方法，所以对相同的数据，同学们可以作出不同的频率分布表. 但是好的频率分布表应当是简单明了的.

练 习

制作习题 6 第 2 题中 49 个同学身高的频率分布表.

习题 7

学而时习之

用随机抽样方法调查了某城市 50 辆公交车的营业额（单位：元），数据已从小到大排列.

259 294 295 297 300 300 300 301 301 302

303 306 308 309 311 314 315 315 321 323

327 328 331 334 336 339 339 339 347 348

350 350 352 355 359 359 361 363 370 376

377 383 388 389 390 396 404 410 410 411

（1）制作频率分布表；

（2）对频率分布表进行简单的分析（参考 P. 84 的例题分析）.

12.3.2 频率分布直方图

从文献记载上看，直方图在 1895 年由著名的英国统计学家**皮尔逊**（Pearson）做了描述，这可能是直方图的第一次使用. 他在为伦敦皇家协会发表的讲话中，当谈及 1885 年至 1886 年英格兰房地产估价的时候使用了直方图.

数据的频率分布图初步展示了数据分布的一些规律. 如果用图形来表示频率分布就会更加形象和直观. 显示数据频率分布的图形有频率分布直方图和茎叶图.

有了数据的频率分布图，很容易作出频率分布的直方图.

将观测数据按照制作频率分布表的方法进行分段，计算出数据落入各段的频率 f_i，将各段的端点画在直角坐标系中的横坐标上，用 f_i/组距作为纵坐标的高，就得到了由相连接长方形构成的图形. 我们把所得到的图形称为数据的频率分布直方图，简称为**直方图**(histogram).

例 绘制 P. 83 例题中图书馆借出图书数据的频率分布直方图.

解 在横坐标上标出所有的数据分段的端点，

200，250，…，550，600.

在区间［200，250］上绘制以 0.05/50＝0.001 为高的矩形；

在区间［250，300］上绘制以频率 0.000 66 为高的矩形；

……

在区间［550，600］上绘制以频率 0.001 为高的矩形.

就得到了需要的频率分布直方图. 如图 12－1.

从频率分布直方图可以更直观地看到图书馆每日借出图书册数的分布情况.

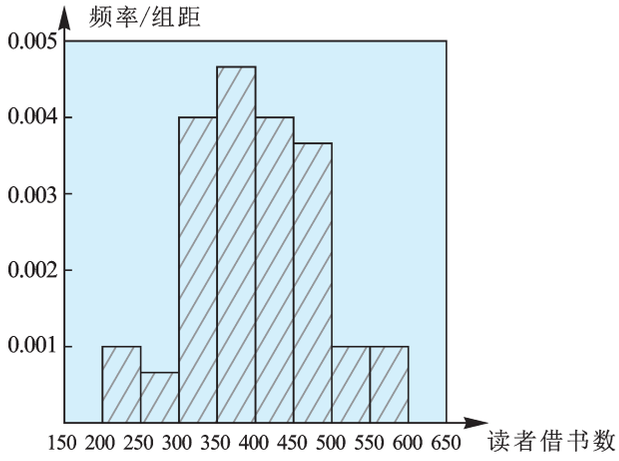


图 12－1　频率表 12.2 的分布直方图

练　习

制作习题 6 第 2 题中 49 个同学身高的频率分布直方图.

习题 8

学而时习之

下面是 1997 年至 2000 年广州的月降水量（单位：mm，数据摘自《中国气象年鉴》），请绘制频率分布直方图.

1997 年	66	106	60	197	166	469	248	282	204	143	10	49
1998 年	73	112	41	245	306	370	223	121	165	48	22	10
1999 年	34	0	62	117	152	152	176	496	273	40	26	54
2000 年	11	30	28	419	203	197	288.9	185	57	304	34	43

12.3.3 频率折线图

用 d_1，d_2，…，d_k 分别表示频率分布直方图中各矩形上边的中点，在直方图的左边延长出一个分段，分段的中点用 d_0 表示．在直方图的右边也延长出一个分段，分段的中点用 d_{k+1} 表示．

用直线连接 d_0，d_1，…，d_{k+1} 就得到了一条折线，这条折线叫作频率折线图．频率折线图也反映出数据频率分布的规律．

图 12－2 是 P.83 的例题中图书馆借出图书数目的频率折线图．

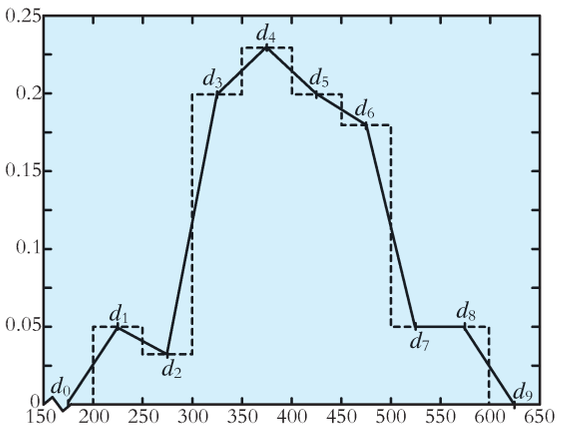


图 12－2　频率表 12.2 的频率折线图

案例　（选择性繁殖问题）为了研究老鼠的智力能否遗传，伯克利（Berkeley）大学教授做了以下的试验．分别让 142 只老鼠走相同的迷宫，每只老鼠走 19 次．老鼠犯错误的次数就是走不出迷宫的次数．我们把犯错误少的老鼠称为伶俐老鼠．记录每只老鼠犯错误的次数，得到 142 个数据．这 142 个数据的频率折线图如图 12－3．

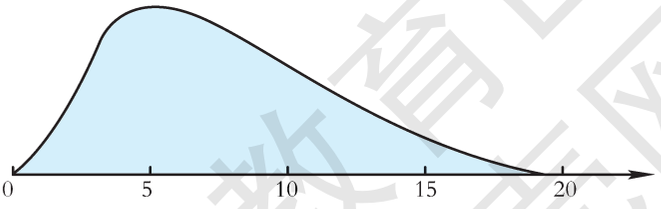


图 12－3

试验后把伶俐的老鼠放在一起，让它们进行繁殖，把不伶俐的老鼠放在一起进行繁殖．繁殖 7 代之后，得到伶俐组的后代 85 只，非

伶俐组的后代 68 只. 让这两组老鼠再走相同的迷宫，每只走 19 次. 得到各组老鼠犯错误的次数后，为两组数据作出的频率折线图如图 12－4.

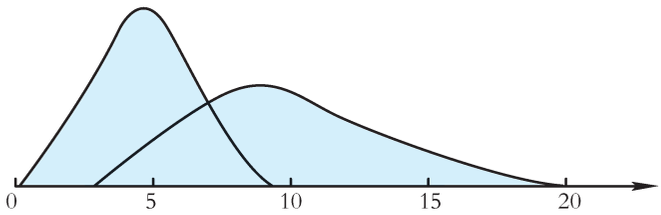


图 12－4

左面的折线图是伶俐组后代的频率折线图，右面的折线图是非伶俐组后代的频率折线图. 这两条折线图有明显的差异. 伶俐组的后代犯错误的次数明显地少，说明老鼠走迷宫的能力是具有遗传性的.

练　习

制作习题 6 第 2 题中 49 个同学身高的频率折线图.

习题 9

学而时习之

根据习题 8 中 1997 年至 2000 年的广州月降水量数据，绘制频率折线图.

12.3.4　数据茎叶图

直方图主要用于展示分段数据的频率分布，对于没有分段的观测

数据还可以用数据的茎叶图展示它的特性.

数据的**茎叶图**（stemplot）由“茎”和“叶”两部分组成，在制作茎叶图的时候要先确定数据的“茎”和“叶”. 从数据的茎叶图可以看出数据的分布形状及数据是否对称，是否集中等分布特性.

我们通过举例说明茎叶图的制作方法.

例 1 下面是上海市 2004 年 7 月 11 日至 2004 年 8 月 1 日空气中可吸入颗粒物的监测数据. 请为这批数据制作茎叶图.

85 85 66 71 62 52 55 59 52 62 59

70 80 96 97 94 62 51 57 67 96 93

解 将数据从小到大排列得到：

51 52 52 55 57 59 59 62 62 62 66

67 70 71 80 85 85 93 94 96 96 97

数据的十位上的数是 5，6，7，8，9，把它们叫作“茎”，排列在下面茎叶图的第一列；

茎 5 后面的个位数分别是 1，2，2，5，7，9，9，把它们叫作茎 5 的“叶”，排在茎 5 的后面；

按相同的方法把茎 6 的叶 2，2，2，6，7 排在茎 6 的右边；

……

把茎 9 的叶 3，4，6，6，7 排在茎 9 的右边.

如此就得到了如图 12－5 所示的茎叶图.

树茎	树叶
5	1 2 2 5 7 9 9
6	2 2 2 6 7
7	0 1
8	0 5 5
9	3 4 6 6 7

图 12－5

优：可吸入颗粒物在 0～50.

良：可吸入颗粒物在 51～100.

轻度污染：可吸入颗粒物在 101～150.

从茎叶图中看出，尽管这 22 天中可吸入颗粒物都是处于良的水平，但是有较多的时间接近于优，也有较多的时间接近于轻微污染.

在同一个茎叶图中还可以表现两组数据的分布情况，这样做有利于对这两组数据进行比较. 我们称表示两组数据的茎叶图为**双茎**

叶图.

例 2 下面是上海市 2004 年 7 月 11 日至 2004 年 8 月 1 日空气中二氧化硫和二氧化氮的监测数据. 请为这两组数据制作一个双茎叶图，并进行比较.

二氧化硫数据：55，62，54，71，60，51，55，56，51，58，61，62，69，73，72，69，58，42，42，65，77，73.

二氧化氮数据：38，37，30，39，31，19，22，22，18，26，25，31，38，44，42，35，22，19，22，37，50，38.

解 先将两组数据分别从小到大排列，得到：

二氧化硫数据：42，42，51，51，54，55，55，56，58，58，60，61，62，62，65，69，69，71，72，73，73，77.

二氧化氮数据：18，19，19，22，22，22，22，25，26，30，31，31，35，37，37，38，38，38，39，42，44，50.

这两组数据都是十位数，选用十位上的数作“茎”，排在双茎叶图的中间一列，它们是 1，2，…，7.

然后将二氧化硫的各位数作为“叶”，依次排在相应的茎的左边. 例如，从数据 42，42 得到茎 4 的叶 2，2，排在 4 的左边；

从数据 51，51，54，55，55，56，58，58 得到茎 5 的叶 1，1，4，5，5，6，8，8，将它们从大到小排在 5 的左边得到 8，8，6，5，5，4，1，1；

……

把茎 7 的叶 1，2，3，3，7 从大到小排在 7 的左边；

再按照例 1 的方法把二氧化氮数据排在“茎”的右边.

这样就得到了两组数据的双茎叶图（图 12－6）：

二氧化硫 树叶								树茎	二氧化氮 树叶									
								1	8	9	9							
								2	2	2	2	2	5	6				
								3	0	1	1	5	7	7	8	8	8	9
						2	2	4	2	4								
8	8	6	5	5	4	1	1	5	0									
	9	9	5	2	2	1	0	6										
			7	3	3	2	1	7										

图 12－6

从上述茎叶图可以看出，二氧化硫的空气质量指标比二氧化氮的空气质量指标要差很多．二氧化硫的空气质量指标基本都处在良好的水平，而二氧化氮的空气质量指标都处在优的水平．

数据茎叶图的优点是显示了数据的每个信息，从茎叶图中可以直观地看到数据的分布情况．但是数据量很大时，茎叶图的效果就不好了，因为这时的茎叶图会很长或很宽．

多知道一点

数据的茎叶图

茎叶图的茎也可以是两位或三位数．

例 制作以下两组数据的双茎叶图．

数据 1：112，113，115，123，126，127，132，137，138，139，142，143，156，158，159，161，164，165，165．

数据 2：123，124，126，134，135，135，137，142，143，146，147，149，152，158，159，162，163，164．

解 数据都是三位数，我们以前两位数作为茎，它们分别是 11，12，13，14，15，16．按照例 2 的方法可以作出双茎叶图如图 12－7：

数据 1 树叶				树茎	数据 2 树叶				
	5	3	2	11					
	7	6	3	12	3	4	6		
9	8	7	2	13	4	5	5	7	
		3	2	14	2	3	6	7	9
	9	8	6	15	2	8	9		
5	5	4	1	16	2	3	4		

图 12－7

练 习

请分别制作下面的习题中数学、物理、语文考试成绩的茎叶图.

习题 10

学而时习之

甲班的期中考试成绩排列如下：

数学：65，65，67，68，70，71，75，76，77，79，80，82，83，83，83，84，84，85，86，86，87，88，88，88，91，93，93，93，93，93，94，95，97，99，99，99，100，100，100，100.

物理：57，60，62，64，67，67，70，72，73，74，74，74，77，77，78，78，79，80，80，81，81，81，82，82，84，84，84，84，86，87，88，93，95，96，96，98，99，99，100，100.

语文：62，67，67，70，70，70，71，72，72，73，74，75，75，76，76，77，78，78，78，78，79，79，80，80，80，80，81，82，82，82，83，83，84，84，86，88，89，90，93，95.

请分别对数学和物理、物理和语文、数学和语文制作双茎叶图，并通过对茎叶图的观察回答以下问题：

（1）哪科平均成绩最好？

（2）哪科成绩分布最集中？

12.4 数据的相关性

在实际问题中，我们经常遇到有相关关系的变量. 比如讲身高与体重的关系时，虽然身高不能确定体重，但总的来讲，身高者，体也重.

在考虑某一个特定地区居民的身高和体重的关系时，用 x 表示人的身高，用 y 表示体重，总体来讲，y 随着 x 的增大一般也会增大. 这时我们称 x 和 y 有**相关关系**.

在某地区的 12～30 岁居民中随机抽取了 10 个样本，用 x_i 和 y_i 分别表示第 i 个人的身高和体重，得到的数据如下：

身高/cm	143	156	159	172	165	171	177	161	164	160
体重/kg	41	49	61	79	68	69	74	69	68	54

数据 x_i 和 y_i 是成对出现的，所以用 (x_i, y_i) 表示第 i 个人的身高和体重. 这时称数据对

$$(x_i, y_i), i=1,2,\cdots,10$$

为样本或观测数据. 样本是平面直角坐标系中的 10 个点，将这 10 个点画在坐标系上得到的图称为观测数据的**散点图**（scatter diagram）. 见图 12－8.

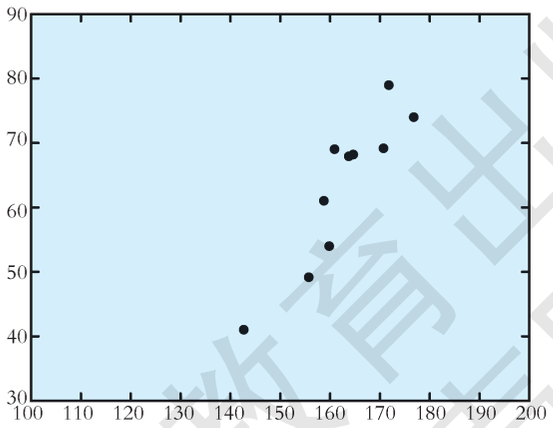


图 12－8　例中数据的散点图

从上面来看，用 (x, y) 泛指总体中某个体的身高和体重时，我们把身高和体重的关系说成是 x 和 y 的关系.

12.4.1 相关性

无论是从抽样调查中得到的成对数据，还是从科学实验、工农业生产中得到的成对数据，在统计学中也都称为观测数据或样本，称数据对的个数为样本量.

样本量是 n 的成对观测数据是用

$$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$$

表示的. 这里，对固定的 i，x_i 和 y_i 或是来自相同的个体，或是同一次试验的观测数据. 对 $i\neq j$，(x_i,y_i) 和 (x_j,y_j) 或是来自不同的个体，或是不同试验的观测数据.

在图 12－8 中，随着身高 x 的增加，体重 y 有明显的增加趋势. 这时称 x 和 y 是正相关的. 当数据

$$(x_1,\ y_1),\ (x_2,\ y_2),\ \cdots,\ (x_n,\ y_n)$$

十分明显地集中在一条上升的直线附近时，我们称 x 和 y 是高度正相关的.

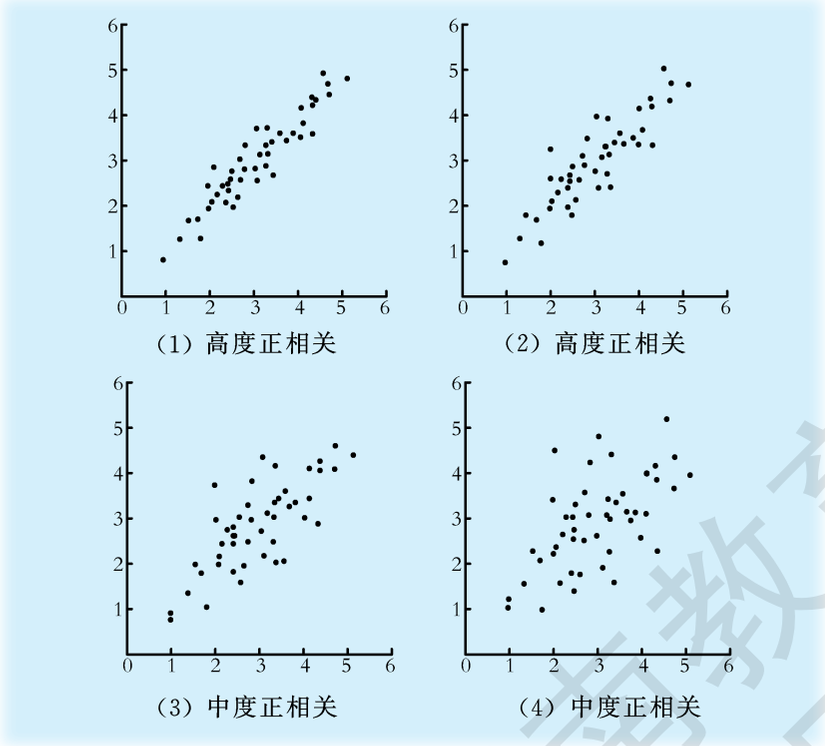


图 12－9

图 12－9 的(1)和(2)展示的数据是高度正相关的. 当上述数据也分布在一条上升的直线附近，但集中的程度不十分明显时，我们称 x 和 y 是中度正相关的. 图 12－9 中(3)和(4)展示的数据是中度正相关的.

当数据 (x_1, y_1)，(x_2, y_2)，…，(x_n, y_n) 十分明显地集中在一条下降的直线附近时，我们称 x 和 y 是高度负相关的. 图 12－10 中(1)和(2)展示的数据是高度负相关的，(3)和(4)展示的数据是中度负相关的.

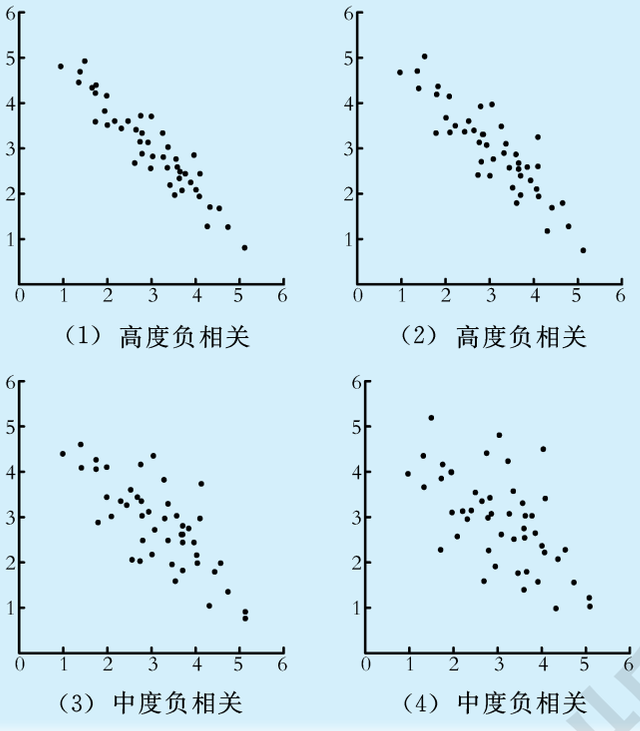


图 12－10

我们把有高度相关性或中度相关性的数据统称为有相关性的数据. 图 12－9 和图 12－10 中展示的数据都是具有相关性的数据，这时也称 x 和 y 是相关的.

练 习

请绘制以下数据的散点图.

年份 x	1975	1976	1977	1978	1979
比萨斜塔倾斜量 y	642	644	656	667	673

习题 11

学而时习之

2000 年的 3 月至 10 月北京和广州的月平均气温（单位：℃）记录如下.

月份	3	4	5	6	7	8	9	10
北京	8	15	20	27	30	26	22	13
广州	19	23	26	28	29	28	27	25

请分别绘制北京和广州月平均气温的散点图，并判断气温平均值与月份的相关性.

12.4.2 回归直线

当$\{x_i\}$和$\{y_i\}$相关时，如果根据数据

$$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$$

描出的散点图中的点大致分布在一条直线的附近，我们将这条直线称为回归直线. 下面就寻找这条直线.

在平面直角坐标系中，两个点(x_1, y_1)，(x_2, y_2)可以决定一条直线.

当 $x_2\neq x_1$ 时，l：$y=\dfrac{y_2-y_1}{x_2-x_1}(x-x_1)+y_1$，

这时，两个点都在直线上，所以这两个点与直线 l 的距离平均最近.

给定三对观测数据

$$(x_1,y_1),(x_2,y_2),(x_3,y_3),$$

当 x_1，x_2，x_3 不全相同，我们也求一条直线 l，使得以上三个点与直线 l 的距离平均最近.

用 $$l：y=bx+a$$

表示要求的直线，在平行于 y 轴的方向，作以上三点到直线 l 的连线，交点 A，B，C 的坐标见图 12－11.

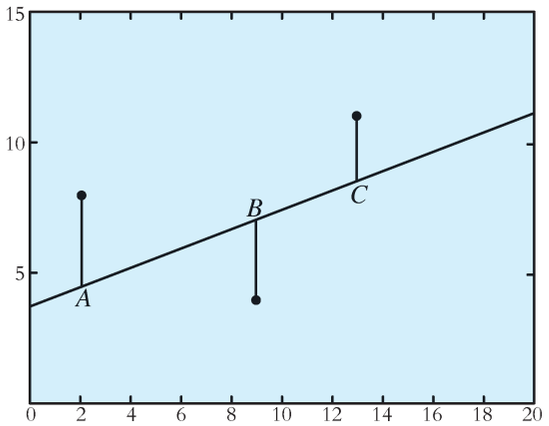


图 12－11

$$A:(x_1,bx_1+a),B:(x_2,bx_2+a),C:(x_3,bx_3+a).$$

三对观测数据和它们与直线 l 交点的距离分别是

$$|y_1-(bx_1+a)|,\ |y_2-(bx_2+a)|,\ |y_3-(bx_3+a)|.$$

我们用这三个距离的平方和

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

衡量这三个观测数据远离直线 l 的程度. 如果 a，b 使得

$$Q(a,b)=(y_1-bx_1-a)^2+(y_2-bx_2-a)^2+(y_3-bx_3-a)^2$$

达到最小，就称直线 l：$\hat{y}=bx+a$ 是回归直线.

不用 $|y_1-(bx_1+a)|+|y_2-(bx_2+a)|+\cdots+|y_n-(bx_n+a)|$ 的原因是数学上不好处理.

一般地，要为样本量是 n 的观测数据

$$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\quad(\text{其中的 } x_i \text{ 不全相同})$$

建立一条直线 l：$\hat{y}=bx+a$，使之与观测数据平均最近时，也采用相同的方法. 沿平行于 y 轴的方向，点 (x_i,y_i) 到它与 l 的交点的距离是

$$|y_i-(bx_i+a)|,\ i=1,2,\cdots,n.$$

我们用这些距离的平方和

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

衡量观测数据远离直线 l 的程度. 如果常数 a，b 使得 $Q(a,b)$ 达到最小，就称直线

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

是 $\{x_i\}$ 与 $\{y_i\}$ 的**回归直线**.

得到了回归直线后，只要 $\{x_i\}$ 与 $\{y_i\}$ 高度正相关或高度负相关，对于新的 x，就可以用回归直线上的点 $\hat{y}=bx+a$ 作为 y 的预测值.

用 $\bar{x}$ 和 $\bar{y}$ 分别表示 $\{x_i\}$ 和 $\{y_i\}$ 的样本均值，用 s_x^2 表示 $\{x_i\}$ 的方差. 引入符号

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

可以证明，只要 x_1，x_2，…，x_n 不全相同，回归直线中的

$$b=\frac{s_{xy}}{s_x^2},\ a=\bar{y}-b\bar{x}.$$

例 1 下面是某冷饮部 8 天中出售的冷饮杯数和当天最高气温的记录数据.

最高气温/℃	26	29	17	23	36	34	5	32
杯数	36	37	29	37	52	49	19	47

请建立坐标系，绘制上述数据的散点图，并建立回归直线方程，且在同一个坐标系中绘制数据的回归直线. 并预测最高气温是 30 ℃ 的一天大约能出售多少杯冷饮.

解 用 x 表示最高气温，用 y 表示杯数，由题意可得数据的散点图如图 12－12 所示.

从数据的散点图看出卖出的冷饮杯数 $\{x_i\}$ 和最高气温 $\{y_i\}$ 是高度正相关的.

先计算出 $\bar{x}=25.25$，$\bar{y}=38.25$，再计算出

$$s_{xy}\approx 95.44,\ s_x\approx 9.59.$$

最后得到

$$b=\frac{s_{xy}}{s_x^2}\approx\frac{95.44}{9.59^2}\approx 1.04;$$

$$a=\bar{y}-b\bar{x}\approx 38.25-1.04\times 25.25\approx 12.$$

于是，回归直线是 $\hat{y}=1.04x+12$.

因而$\{x_i\}$和$\{y_i\}$的回归直线如图 12－12 中直线所示.

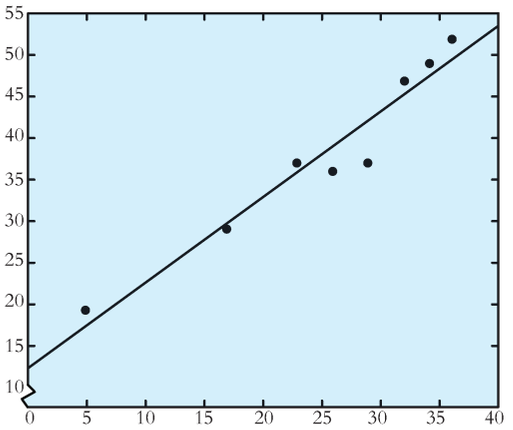


图 12－12

当 $x=30$ ℃时，对 y 的预测值是

$$\hat{y}=1.04\times30+12=43.2.$$

故最高气温是 30 ℃时，大约可以出售 43 杯冷饮.

例 2 在测量一根新弹簧的劲度系数时，测得了如下的结果.

所挂重量/N	1.00	2.00	3.00	5.00	7.00	9.00
弹簧长度/cm	10.18	11.21	11.85	13.01	14.29	15.03

（1）建立坐标系，绘制数据的散点图；

（2）建立回归直线方程，并在同一个坐标系中绘制数据的回归直线.

解 （1）用 x 表示所挂重量，用 y 表示弹簧长度，由题意可得数据的散点图如图 12－13 所示.

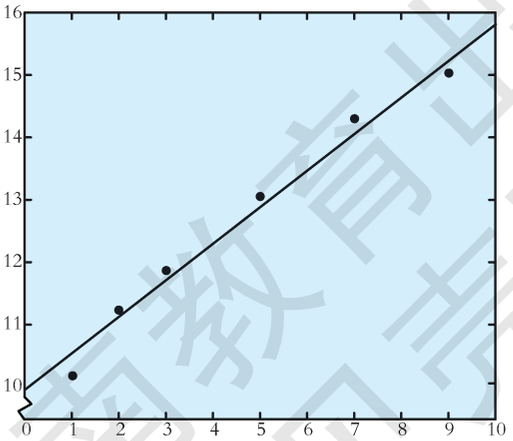


图 12－13

(2) 从数据的散点图 12－13 可以看出$\{x_i\}$和$\{y_i\}$是高度正相关的.

按例 1 中的方法可以计算出

$$\bar{x}=4.50,\ \bar{y}=12.595,\ s_{xy}\approx 4.74,\ s_x\approx 2.81.$$

因此

$$b=\frac{s_{xy}}{s_x^2}=\frac{4.74}{2.81^2}\approx 0.6,$$

$$a=\bar{y}-b\bar{x}\approx 12.595-0.6\times 4.5\approx 9.9.$$

于是，回归直线是

$$l:\ \hat{y}=0.6x+9.9.$$

因而所求回归直线图如图 12－13 中直线所示.

我们知道弹簧的拉伸长度和所挂重量之间的关系服从**胡克**(Hooke)定律：

$$y=b_0x+a_0.$$

其中 x 是弹簧所挂重物的重量 (N)，y 是弹簧拉伸后的长度. 这里的 a_0 和 b_0 是未知的，所以 $b=0.6$ 和 $a=9.9$ 分别是 b_0 和 a_0 的估计.

由于 a_0,b_0 是弹簧本身固有的量，把它们称为**参数**(parameter). 由于 a，b 是通过求平方和 $Q(a,b)$ 的最小值得到的，所以我们称回归直线中的 a，b 分别是参数 a_0，b_0 的**最小二乘估计**. 这里最小是求最小值的意思，二乘是平方和的意思.

练 习

请建立 $\{x_i\}$ 与 $\{y_i\}$ 的回归直线：

年份 x	1984	1985	1986	1987
比萨斜塔倾斜量 y	717	725	742	757

习题 12

学而时习之

1. 当 x_1，x_2，…，x_n 不全相同时，证明（$\bar{x}$，$\bar{y}$）总在回归直线上.

2. 以下数据来自对 5 个职工的随机抽样调查. x 表示月平均收入，y 表示用于购书和买报纸的月平均支出.（单位：元）

职工	1	2	3	4	5
x	1 200	1 400	1 700	1 800	2 100
y	12	11	14	18	21

（1）建立坐标系，绘制数据的散点图；

（2）请建立回归直线方程，并分别对月收入为 1 850 元和 1 500 元的职工预测他们每月平均用于购书和买报纸的支出（结果保留整数）；

（3）在同一个坐标系中绘制数据的回归直线.

3. 以下是 4 个同学每天平均完成作业的时间和每天平均看电视的时间的调查结果，调查在同一个班内用简单随机抽样方法完成.（单位：h）

同学	1	2	3	4
做作业的时间 x	2	2.5	3.2	3
看电视的时间 y	1.5	0.6	0.1	0.3

（1）建立坐标系，绘制数据的散点图；

（2）请建立回归直线方程，并分别对平均每天用 1.9，2.8 和 3.2 h 完成作业的同学预测他们平均每天看电视的时间（结果保留两位小数）；

（3）在同一个坐标系中绘制数据的回归直线.

多知道一点

使用计算机或计算器做统计计算

MatLab 是一个应用广泛的计算机软件，功能强大，使用方便，

易学易懂. MatLab 的各种版本大同小异，低版本的命令可以在高版本上应用. 本书的统计部分使用 MatLab 进行计算.

通过下面的例子可以初步学会使用 MatLab.

1. 计算 156，165，…，161 均值时，用下面的两个语句.［每个语句后面敲换行（Enter). 括弧中的内容是语句的解释，不输入. 后同.］

x=[156;165;…;161];

M=mean(x) （计算 x 的均值).

2. 将数据 156，165，…，161 从小到大排序时，用下面的语句：

x=[156;165;…;161];

sort(x)

3. 计算 $n=50$ 个数据 60，65，65，…，98 的方差时，用下面的语句：

n=50;

y=[60; 65; 65; …; 98];

s2=Std(y)^2*(n−1)/n （计算 y 的方差，n 是样本量).

4. 计算 $n=50$ 个数据 60，65，65，…，98 的标准差时，用下面的语句：

n=50;

y=[60; 65; 65; …; 98];

s1=Std(y)*sqrt((n−1)/n) （计算 y 的标准差).

5. 计算回归直线中的最小二乘估计 a，b 时，用下面的语句.

x=[1; 2; 3; …; 7; 9];

y=[10.18; 11.21; …; 15.03];

ba=polyfit(x, y, 1) （计算 b 和 a，输出的第一个数是 b，第二个数是 a）

b a

0.598 6 9.901 2 （得到的结果).

通过下面的例子学会使用计算器做统计计算.

1. 计算 3，5，8，12，19，21.5，32.3 的样本均值、样本标准

差和样本方差.

MODE 2 （进入计算模式）

SHIFT SCL = （清空统计存储器）

3 DT 5 DT 8 DT 12 DT 19 DT 21.5 DT 32.3 DT

SHIFT $\bar{x}$ = 14.4 （计算的样本均值）

SHIFT $x\sigma n$ = 9.688 137 076 （计算的样本标准差）

x^2 = 93.86 （计算的样本方差）

2. 计算回归直线中的参数 a，b，对 $x=19$ 预测 y.

x	10	15	20	25	30	35
y	103	105	110	115	120	124

MODE 3 1 （进入回归模式）

SHIFT SCL = （清空统计存储器）

10，103 DT 15，105 DT 20，110 DT 25，115 DT 30，120 DT 35，124 DT

SHIFT A = 92.904 761 9 （计算的 a）

SHIFT B = 0.885 714 285 （计算的 b）

19 SHIFT $\hat{y}$ 109.733 333 3 （对 $x=19$ 预测的 y）

数学实验

用计算机画回归直线和做统计计算

打开用“Z＋Z 超级画板”制作的课件“回归直线．zjz”，屏幕画面如图 12－14：

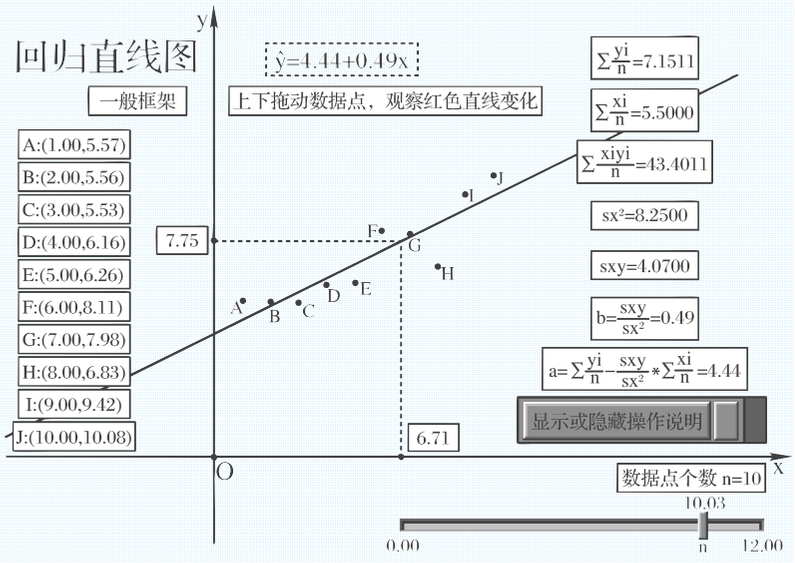


图 12－14

图上显示出有 10 个数据点的散点图和对应的回归直线．这些数据点可以上下拖动，回归直线随之变化．左边显示出每个点的坐标，即数据对；右边显示出为画出回归直线而必须做的统计计算结果．

鼠标单击右下方“显示或隐藏操作说明”按钮，仔细阅读出现的文本内容；根据操作说明，输入本章 12.4.2 节例 2 中的数据，并把点的个数调整为 6，就得到书中图 12－13 所示的回归直线．

单击上方的“下一页”图标，屏幕上出现课件的第二页，

如图 12－15，即本章 12.4.2 节例 2 的回归直线图. 对照一下，和你作的是否相同？如果不同，检查自己的操作是否有误.

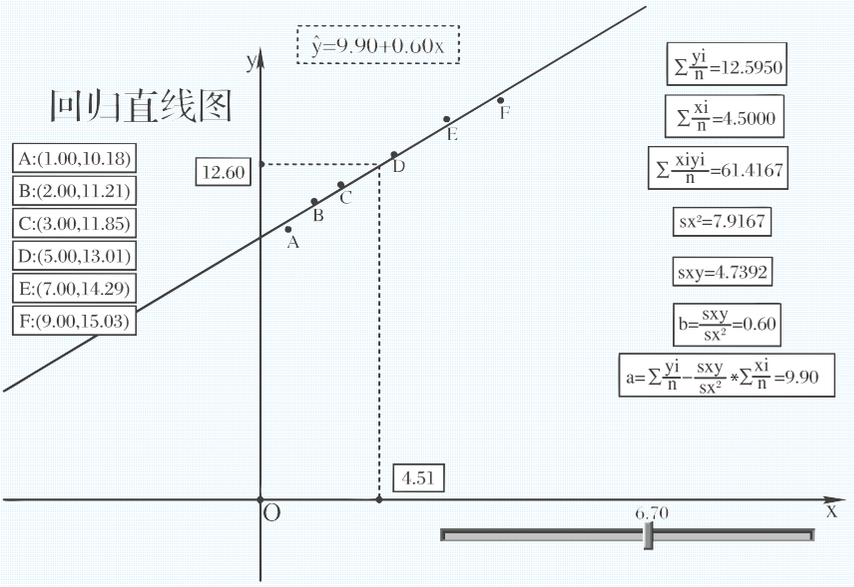


图 12－15

拖动横坐标轴上的红点，注意淡蓝色小框中数字的变化，想一想，这些数字的意义是什么？能利用这些数字预测弹簧所挂重量为 4.5 kg 时的长度吗？

类似地，可作出本章 12.4.2 节例 1 的图，并在图上预测最高气温为30 ℃时出售冷饮的数量.

课件中设置了自动计算统计量的功能. 为了熟悉统计量的计算，也可以应用前一章学到的算法知识，自己写程序计算统计量.

下面举例说明如何在“Z＋Z 超级画板”的程序工作区计算统计量.

最直接的办法是把数据和公式直接输入. 如 12.1.3 节的例 3，要计算 10 个数据

9.5，9.9，9.9，9.9，9.8，9.7，9.5，9.3，9.6，9.6

的均值并用 mx 表示，就在程序工作区键入：

```
mx=(9.5+9.9+9.9+9.9+9.8+9.7+9.5+9.3+
    9.6+9.6)/10;
```

按 Ctrl＋Enter 键执行（下同）得到

≫(967)/(100) #

如果想要用小数表示，可执行

Float(1);

返回：

≫计算结果显示浮点数 #

再键入

mx;

执行得

≫(967)/(100)=9.67 #

要计算这组数据的方差 s^2，可键入

s^2=(9.5^2+9.9^2+9.9^2+9.9^2+9.8^2+9.7^2+9.5^2+9.3^2+9.6^2+9.6^2)/10−mx^2;

执行得到

≫(381)/(10000)=0.0381 #

计算标准差则再键入：

sx=s2^0.5;

执行即得到

≫((381)^(1/2))/(100)=0.195192 #

用这种方法计算的好处，在于能在计算过程中多次复习这些统计量的意义和公式. 用了计算机，计算统计量的主要工作就是输入数据. 为了避免多次输入“+”号和平方运算符号“^2”的重复性工作，可以做一个“模板”，也就是先打一行符号，如

(^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2+^2)

复制一下，用时随时粘贴，再在符号“^”前面依次键入数据就是了.

小结与复习

一、总体和个体

1. 样本：是从总体中抽取的一部分个体，也称为观测数据.

2. 总体均值：总体的平均，常用 μ 表示.

3. 样本均值：样本的平均，常用 $\bar{x}$ 表示. 样本平均是总体均值的估计.

4. 均值的性质：

（1）如果每个数据增加 b，均值也增加 b；

（2）如果每个数据增加到原来的 a 倍，均值也增加到原来的 a 倍.

5. 总体方差：当 y_1，y_2，…，y_N 是总体的全部个体，μ 是总体均值时，总体方差

$$\sigma^2=\frac{(y_1-\mu)^2+(y_2-\mu)^2+\cdots+(y_N-\mu)^2}{N}.$$

6. 样本方差：用 $\bar{x}$ 表示样本 x_1，x_2，…，x_n 的均值时，样本方差

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

样本方差是总体方差的估计.

7. 方差的性质：

（1）方差描述数据向均值的集中程度：方差越小，个体向均值集中得越好. 方差也描述数据的整齐程度或波动幅度：方差越小，数据就越整齐.

（2）$s^2=\frac{1}{n}(x_1^2+x_2^2+\cdots+x_n^2)-\bar{x}^2$.

8. 标准差：标准差是方差的算术平方根. 数据带有单位时，标准差的单位和数据的单位是一致的.

二、抽样调查方法

1. 抽样调查的必要性：在很多实际问题中，采用抽样的方法来确定总体性质不仅是必要的，也是必须的.

2. 随机抽样：指总体中的每个个体都有相同的机会被抽中.

3. 简单随机抽样：无放回随机抽样.

4. 随机抽样的重要性：不进行随机抽样，就可能得到有偏的样本. 利用有偏的样本很容易得到错误的结论.

5. 随机数：是从 1 至 n 个号码中用简单抽样方法抽到的 m（$m \leqslant n$）个号码. 随机数可以在计算机上产生.

6. 分层抽样：将总体分成若干层（子总体），然后在每层中独立地进行简单随机抽样.

7. 分层抽样的特点：

（1）分层抽样在获得总体均值估计的同时，也得到各层的均值估计；

（2）将差别不大的个体分在同一层，使得分层抽样得到的样本更具有代表性，从而提高估计的准确度；

（3）抽样调查的实施更加方便，调查数据的收集、处理也更加方便.

8. 系统抽样：当总体中的个体按一定的方式排列，在规定的范围内随机抽取一个个体，然后按照制定好的规则确定其他个体的抽样方法.

9. 系统抽样的特点：实施简单，如果了解总体中个体排列的规律，设计合适的系统抽样规则可以增加估计的精度.

三、用样本分布估计总体分布

1. 频率分布表、频率分布直方图、频率折线图和数据的茎叶图都是用来展示样本的分布性质的，这些分布性质是总体分布性质的估计或近似.

2. 绘制频率分布表时，数据分段个数 K 可以参考经验公式

$$K=1+4\lg n$$

确定. 但是经验公式只起参考作用. 实际应用时，应当根据样本量的大小和数据的特点以及分析的要求灵活确定.

说明：由于频率分布表的制作没有统一的数据分段方法，所以对相同的数据，可以作出不同的频率分布表，因而也可以作出不同的直方图和频率折线图. 但是好的频率分布表、直方图和频率折线图应当是简单明了的.

四、数据的相关性

样本量是 n 的成对观测数据是用

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

表示的. 这里，对固定的 i，x_i 和 y_i 或是来自相同的个体，或是同一次试验的观测数据. 对 $i\neq j$，(x_i, y_i) 和 (x_j, y_j) 或是来自不同的个体，或是不同次试验的观测数据.

1. 当 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 十分明显地集中在一条上升的直线附近时，称 x 和 y 是高度正相关的；当它们分布在一条上升的直线附近时称 x 和 y 是中度正相关的. 高度正相关和中度正相关统称为正相关.

2. 当 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 十分明显地集中在一条下降的直线附近时，称 x，y 是高度负相关的；当它们分布在一条下降的直线附近时，称 x 和 y 是中度负相关的. 高度负相关和中度负相关统称为负相关.

3. 当 x 和 y 正相关或负相关时，称 x 和 y 是相关的.

4. 回归直线：当 $\{x_i\}$ 和 $\{y_i\}$ 有相关性时，可以用一条直线描述数据 $\{x_i\}$ 和 $\{y_i\}$ 的关系，这条直线就是回归直线. 用

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

表示这条回归直线时，其中的 a，b 可以用下面的公式进行计算：

$$b=\frac{s_{xy}}{s_x^2},\ a=\bar{y}-b\bar{x}.$$

5. 用回归直线进行预测：得到了回归直线后，只要 $\{x_i\}$ 与

$\{y_i\}$ 高度相关，对于新的 x，就可以用回归直线上的点 $\hat{y}=bx+a$ 作为 y 的预测值.

复习题十二

学而时习之

1. 在调查某城市的居民对自来水调价的意见时，能否只在该市的娱乐场所进行随机抽样？原因是什么？能否只在该市的公共汽车站进行随机抽样？原因是什么？

2. 以下是对某个公交车站候车的乘客候车时间的调查数据（单位：min）.

21	1	22	19	9	23	15	19	5	2	7	13
4	20	24	18	6	16	19	5	1	21	16	2
14	16	7	2	5	3	4	19	5	14	13	20
18	9	7	22	10	9	9	23	11	1	18	21

请计算样本均值和样本标准差.

3. 甲、乙两篮球运动员上赛季每场比赛的得分如下，试通过画茎叶图比较这两名运动员的得分水平.

甲　12，15，24，25，31，31，36，37，39，44，49，50.

乙　8，13，14，16，23，26，28，33，38，39，51.

4. 要在全年级 450 名同学中随机选取 45 人参加暑假的夏令营时，需完成以下工作：

(1) 设计一个随机抽样方案；

(2) 设计一个系统抽样方案；

(3) 设计一个分层抽样方案，使得选取出男生 23 名，女生 22 名；

(4) 如果全年级有 9 个班，设计一个分层抽样方案，使得各班随机选取 5 人.

温故而知新

5. 从某个渔场通过简单随机抽样方法检查了 30 条鱼的质量（单位：g）如下：

254 298 320 325 347 354 362 367 369 372
379 380 382 386 391 397 397 411 412 419
422 435 438 443 447 452 480 506 524 539

(1) 制作频率分布表；

(2) 制作频率直方图；

(3) 制作频率折线图；

(4) 制作茎叶图；

(5) 计算样本均值；

(6) 计算样本方差.

6. 按以下要求分别设计随机抽样调查方案，调查本年级的近视率.

(1) 要求随机调查 50 个同学；

(2) 要求调查男女同学各 25 人；

(3) 要求调查每班 10 个同学；

(4) 要求调查每班的男女生各 5 人.

7. 2000 年的 3—10 月北京和广州的月平均气温（单位：℃）记录如下：

月份	3	4	5	6	7	8	9	10
北京 x	8	15	20	27	30	26	22	13
广州 y	19	23	26	28	29	28	27	25

请为北京和广州月平均气温建立回归直线.

上下而求索

8. 下面是世界上 10 个地区的人均收入（百元）和受教育比例的数据：

人均收入 x	61	165	125	645	398	208	289	311	246	86
受教育比例 y	6	43	50	87	80	71	30	77	96	77

(1) 建立坐标系，绘制数据的散点图；

(2) 请建立回归直线方程，并在同一个坐标系中绘制数据的回归直线；

(3) 对 $x=60$，70，200 分别预测 y 的值；

(4) 对于 $y=85$ 预测 x 的值.

9. 下面是某种合成纤维的拉伸强度 y 和拉伸倍数 x 之间的测量数据.

编号	1	2	3	4	5	6	7	8	9	10	11	12
拉伸倍数 x	1.9	2.0	2.1	2.5	2.7	2.7	3.5	3.5	4.0	4.0	4.5	4.6
强度 y (Mpa)	1.4	1.3	1.8	2.5	2.8	2.5	3.0	2.7	4.0	3.5	4.2	3.5
编号	13	14	15	16	17	18	19	20	21	22	23	24
拉伸倍数 x	5.0	5.2	6.0	6.3	6.5	7.1	8.0	8.0	8.9	9.0	9.5	10
强度 y (Mpa)	5.5	5.0	5.5	6.4	6.0	5.3	6.5	7.0	8.5	8.0	8.1	8.1

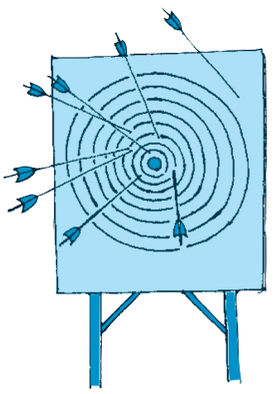
(1) 建立坐标系，绘制数据的散点图；

(2) 请建立回归直线方程，并在同一个坐标系中绘制数据的回归直线（精确到小数点后两位）；

(3) 对 $x=5.0$，5.2 分别预测 y 的值（精确到小数点后两位）；

(4) 对于 $y=4.5$ 预测 x 的值（精确到小数点后两位）.

第13章 概 率



沙场百胜古来稀，
九密一疏已足奇.
祸福偶然存概率，
风云变幻识玄机.

在考虑一个未来事件是否会发生的时候，人们常关心该事件发生的可能性的大小. 概率就是用来衡量一个未来事件发生的可能性大小的度量. 概率通过对简单随机事件的研究，将人们逐步引入复杂随机现象规律的研究. 概率是研究复杂随机现象规律的有效方法和工具.

投掷一枚均匀的硬币，出现正面朝上的机会是 50%. 在多次投掷这枚硬币后，正面朝上和反面朝上出现的次数大致相同. 南非数学家凯瑞（Kerrich）是在非常困难的情况下发现这一点的. 他的掷币试验是在德国人的集中营里进行的. 二战爆发时他访问哥本哈根，德国人入侵丹麦时他被关进集中营，在那里度过了漫长的岁月. 为了消磨时间，他一次次地掷硬币，并记录了下面的结果：

掷币次数 N	正面朝上次数 n	频率 $f=n/N$	掷币次数 N	正面朝上次数 n	频率 $f=n/N$
10	4	0.400	600	312	0.520
20	10	0.500	700	368	0.526
30	17	0.567	800	413	0.516
40	21	0.525	900	458	0.509
50	25	0.500	1 000	502	0.502
60	29	0.483	2 000	1 013	0.507
70	32	0.457	3 000	1 510	0.503
80	35	0.438	4 000	2 029	0.507
90	40	0.444	5 000	2 533	0.507
100	44	0.440	6 000	3 009	0.502
200	98	0.490	7 000	3 516	0.502
300	146	0.487	8 000	4 034	0.504
400	199	0.498	9 000	4 538	0.504
500	255	0.510	10 000	5 067	0.507

从上述掷币结果看出，随着掷币次数的增加，正面朝上的频率稳定在 0.5 附近. 我们将这个性质称为**频率的稳定性**.

由学习过的概率知识知道，投掷一枚硬币，出现正面朝上的概率是 0.5. 在多次重复试验中，正面朝上的频率稳定在正面朝上的概率附近.

13.1 试验与事件

13.1.1 事 件

投掷一枚硬币，用 H 表示硬币正面朝上，用 T 表示硬币反面朝上，则试验有两个可能的结果：H 和 T. 我们把 H 和 T 叫作试验的元素，把集合 $\{H, T\}$ 叫作试验的全集.

投掷一枚骰子，用 1 表示掷出点数 1，用 2 表示掷出点数 2，…，用 6 表示掷出点数 6，则试验的可能结果是 1，2，3，4，5，6. 我们称这 6 个数是试验的元素. 称

$$\{j \mid j=1,2,\cdots,6\}$$

为试验的全集.

对于一个试验，我们将该试验的可能结果称为元素，称所有元素构成的集合为试验的全集.

以后总用 Ω 表示试验的全集，用 ω 表示 Ω 的元素.

Ω 音 omega.

这里 ω（音 omega）是 Ω 的小写.

例 1 同时投掷一枚 1 角的硬币和一枚 1 元的硬币，写出试验的元素和全集.

解 试验一共有 4 个元素，它们是

HH：1 角硬币正面朝上，1 元硬币正面朝上；

HT：1 角硬币正面朝上，1 元硬币反面朝上；

TH：1 角硬币反面朝上，1 元硬币正面朝上；

TT：1 角硬币反面朝上，1 元硬币反面朝上.

全集是 $\Omega=\{HH, HT, TH, TT\}$.

在上面的例子中，HT 和 TH 是不同的元素.

例 2 投掷两枚硬币，写出试验的元素和全集.

解 将一个硬币视为硬币 1，另一个硬币视为硬币 2. 试验有 4 个元素，它们是

HH：硬币 1 正面朝上，硬币 2 正面朝上；

HT：硬币 1 正面朝上，硬币 2 反面朝上；

TH：硬币 1 反面朝上，硬币 2 正面朝上；

TT：硬币 1 反面朝上，硬币 2 反面朝上.

全集是 $\Omega=\{HH,HT,TH,TT\}$.

和例 1 中的情况相同，这里 HT 和 TH 也是不同的元素.

投掷一枚骰子的全集是

$$\Omega=\{j\mid j=1,2,\cdots,6\}.$$

用 $A=\{3\}$ 表示掷出 3 点，A 是 Ω 的子集. 我们称 A 是**随机事件** (random event)，简称为事件. 掷出 3 点，就称事件 A 发生，否则称事件 A 不发生.

用 $B=\{2,4,6\}$ 表示掷出偶数点，B 是 Ω 的子集，我们也称 B 是随机事件，简称为事件. 当掷出偶数点，称事件 B 发生，否则称事件 B 不发生. 事件 B 发生和掷出偶数点是等价的.

当 Ω 是试验的全集，我们称 Ω 的子集 A 是 Ω 的事件，简称为事件 (event). 当试验结果（即试验的元素）ω 属于 A 时，就称事件 A 发生，否则称 A 不发生.

对于全集 Ω，A 是事件和 $A\subseteq\Omega$ 等价. 元素 $\omega\in A$ 和事件 A 发生等价.

空集$\varnothing$也是 Ω 的子集，所以空集$\varnothing$是事件. 空集$\varnothing$中没有元素，永远不会发生，所以我们称$\varnothing$是**不可能事件**.

Ω 也是 Ω 的子集，并且包括了所有的元素，所以必然发生. 我们称全集 Ω 是**必然事件**.

事件是随机事件的简称.

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

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

练 习

1. 叙述什么是试验的元素和全集.
2. 简述事件和全集的关系.
3. 简述事件和元素的关系.

习题 1

学而时习之

1. 口袋中有标号 1～3 的球各 1 个. 为以下的试验写出全集.

(1) 从中任取 1 个;

(2) 从中一次随机地取出 2 个.

2. 投掷一枚骰子和一枚硬币，写出全集.

3. 同时投掷一枚骰子和一枚硬币，写出以下事件.

(1) 硬币是正面，骰子的点数是奇数;

(2) 硬币是正面，骰子的点数是偶数;

(3) 硬币是正面;

(4) 骰子的点数是 5.

13.1.2 事件的运算

由于事件就是集合，所以对事件可以进行并、交和补的运算.

例 1 投掷两枚骰子，一枚是红色，一枚是蓝色. 写出全集和以下事件.

(1) $A=$“红骰子的点数是 2”;

(2) $B=$“蓝骰子的点数是 3”;

(3) $A\cap B$;

(4) $A\cup B$.

解 用(i,j)表示红色骰子的点数是 i，蓝色的点数是 j. 试验的全集是

$$\Omega=\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),$$
$$(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),$$

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

根据事件的定义，得到

(1) $A=\{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)\}$;

(2) $B=\{(1, 3), (2, 3), (3, 3), (4, 3), (5, 3), (6, 3)\}$;

(3) $A\cap B=\{(2, 3)\}$ = “红骰子是 2 点，蓝骰子是 3 点”;

(4) $A\cup B=\{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (1, 3), (3, 3), (4, 3), (5, 3), (6, 3)\}$

= “红骰子是 2 点或蓝骰子是 3 点”.

用 $\Omega\setminus A$ 表示 A 的补集.

在例 1 中，$\Omega\setminus A$ 表示红骰子的点数不是 2.

对于试验的全集 Ω 和事件 A，由于 A 和 $\Omega\setminus A$ 有且只能有一个发生，所以我们称 $\Omega\setminus A$ 是 A 的**对立事件**.

必修第一册 P8 的“多知道一点”中已经介绍可以用 $A\setminus B$ 表示 A 与 B 的差集，其含义为 $A\setminus B=\{x\mid x\in A\text{ 且 }x\notin B\}$.

A 的对立事件在不产生歧义的条件下也可以简记为 $\overline{A}$.

A 与 $\Omega\setminus A$ 互为对立事件.

例 2 铅笔盒中有圆珠笔 3 支，钢笔 2 支. 从中无放回地任取 3 支，用集合 A，B，C 表示下面 (1)，(2)，(3) 中的事件.

(1) 3 支都是圆珠笔;

(2) 恰有 2 支圆珠笔;

(3) 恰有 1 支圆珠笔;

(4) 用 A，B，C 表示 Ω;

(5) 解释事件 $A\cup B$，$A\cap B$，$A\setminus B$，$\Omega\setminus A$ 的含义.

解 将 3 支圆珠笔标号 1，2，3，将 2 支钢笔编号 1，2. 用 y_i 和 g_j 分别表示取出的有第 i 支圆珠笔和第 j 支钢笔. 用 $y_1y_2y_3$ 表示取出的是 1 号，2 号和 3 号圆珠笔，$y_1y_2g_1$ 表示取出的是 1 号，2 号圆珠笔和 1 号钢笔……按照事件的定义，得到

(1) $A=\{y_1y_2y_3\}$.

(2) $B=\{y_1y_2g_1, y_1y_2g_2, y_1y_3g_1, y_1y_3g_2, y_2y_3g_1, y_2y_3g_2\}$.

(3) $C=\{y_1g_1g_2, y_2g_1g_2, y_3g_1g_2\}$.

(4) 因为必有事件 A，B，C 之一发生，所以全集 $\Omega=A\cup B\cup C$.

(5) $A \cup B$ = “至少有 2 支圆珠笔”；

$A \cap B = \varnothing$ = “不可能事件”；

$A \setminus B = A$ = “3 支都是圆珠笔”；

$\Omega \setminus A$ = “至少有 1 支钢笔”.

若两事件对立，则它们互斥吗？

在例 2 中，事件 A，B 的交集是空集，所以 A 发生，B 就不能发生；B 发生，A 就不能发生.

当 $A \cap B = \varnothing$，我们称 A，B **互斥**.

练 习

投掷 3 枚硬币，观察结果. 写出全集，分别用集合 A，B，C 表示以下 (1)，(2)，(3) 中的事件.

(1) 恰好 2 个正面；

(2) 至少 1 个正面；

(3) 都是反面；

(4) 计算 $B \setminus A$，$B \setminus C$，$\Omega \setminus A$，并解释它们的含义.

习题 2

学而时习之

1. 投掷 4 枚硬币，观察结果. 不写出全集，直接用集合 A，B，C 表示以下 (1)，(2)，(3) 中的事件.

(1) 至少 1 个反面朝上；

(2) 至少 2 个反面朝上；

(3) 恰好 2 个反面朝上；

(4) 计算 A，$A \cap C$，$\Omega \setminus A$，并解释含义.

2. 盒子中有标号 1～3 的白球各 1 个，标号 1～2 的黑球各 1 个. 从中倒出 3 个，

观察结果. 写出全集，用集合 A,B,C 表示下面(1),(2),(3)中的事件.

(1) 3 个都是白球；

(2) 至少 2 个白球；

(3) 至少 1 个白球；

(4) 计算 $A\cup B$，$A\cap B$，$A\setminus B$，$C\setminus B$，并解释它们的含义.

13.2 概率及其计算

13.2.1 古典概率模型

1. 概率的定义.

投掷一枚均匀的硬币，全集 $\Omega=\{H,T\}$ 中有两个元素. 事件 $A=\{H\}$,$B=\{T\}$ 各有一个元素. 根据已有的概率知识,得

$$P(A)=\frac{1}{2}=\frac{A\text{ 中元素数}}{\Omega\text{ 中元素数}},\qquad P(B)=\frac{1}{2}=\frac{B\text{ 中元素数}}{\Omega\text{ 中元素数}}.$$

由初中所学的列举法可以计算出.

投掷一枚均匀的骰子,则全集是

$$\Omega=\{j\mid j=1,2,\cdots,6\}.$$

若 $A=\{3\}$ 表示掷出的点数是 3,$B=\{2,4,6\}$ 表示掷出偶数点. 于是全集 Ω 中元素的个数是 6,A 中元素的个数是 1,B 中元素的个数是 3. 根据已有的概率知识,有

$$P(A)=\frac{A\text{ 中元素数}}{\Omega\text{ 中元素数}}=\frac{1}{6},\qquad P(B)=\frac{B\text{ 中元素数}}{\Omega\text{ 中元素数}}=\frac{1}{2}.$$

若 $A=\{1\}$ 或 $\{4\}$ 或 $\{6\}$，则 $P(A)$ 仍会相同吗?

由以上的例子引出概率的如下定义.

定义 设试验的全集 Ω 有 n 个元素，且每个元素发生的可能性相同. 当 Ω 的事件 A 包含了 m 个元素时，称

$$P(A)=\frac{m}{n}$$

为事件 A 发生的概率，简称为 A 的**概率** (probability).

我们把上述定义描述的概率模型称为古典概率模型，简称为**古典**

概型.

因此古典概型具有以下特点：

（1）试验中所有可能出现的元素只有有限个；

（2）每个元素出现的可能性相等.

对于古典概型，事件 A 的概率计算公式为：

$$P(A)=\frac{A\text{ 中包含的元素个数}}{\text{全集中的元素个数}}.$$

例 1 同时投掷 3 枚硬币. 计算以下事件的概率.

（1）至少 1 个反面朝上；

（2）至少 2 个反面朝上；

（3）恰好 2 个反面朝上.

解 试验的全集为：

$$\Omega=\{HHH,HHT,HTH,THH,HTT,THT,TTH,TTT\}.$$

（1）用 A 表示至少有一个反面朝上，则 A 中的元素至少含 1 个 T.

$\therefore A=\{HHT,HTH,THH,HTT,THT,TTH,TTT\}$.

$\therefore A$ 含有 7 个元素，于是 $P(A)=\frac{7}{8}$.

（2）用 B 表示至少 2 个反面，则

$$B=\{HTT,TTH,THT,TTT\}.$$

$\therefore B$ 含有 4 个元素，于是 $P(B)=\frac{4}{8}=\frac{1}{2}$.

（3）用 C 表示恰好 2 个反面朝上，则

$$C=\{HTT,TTH,THT\}.$$

$\therefore C$ 含有 3 个元素，于是 $P(C)=\frac{3}{8}$.

事件作为集合经过并、交、差和补的运算后得到的结果还是事件，于是可以计算经过集合运算后的事件的概率.

例 2 在例 1 中，计算事件 $A\cup B$，$A\cap C$，$B\setminus C$，$\Omega\setminus C$ 发生的概率.

解 全集 Ω 有 8 个元素. $A\cup B=A$，含有 7 个元素，

$A\cap C=C$，含有 3 个元素，$B\setminus C=\{TTT\}$，含有 1 个元素，

$\Omega\backslash C=\{HHH,HHT,HTH,THH,TTT\}$，含有 5 个元素. 所以有

$P(A\cup B)=\frac{7}{8}$，$P(A\cap C)=\frac{3}{8}$，$P(B\backslash C)=\frac{1}{8}$，$P(\Omega\backslash C)=\frac{5}{8}$.

2. 概率的性质.

概率有如下的简单性质：

(1) $0\leqslant P(A)\leqslant 1$（概率总是 $[0,1]$ 中的数）；

(2) $P(\Omega)=1$（必然事件的概率是 1）；

(3) $P(\varnothing)=0$（不可能事件的概率是零）.

草花就是我们常说的梅花.

在一副扑克的 54 张牌中随机抽取 1 张. 用 A 表示得到的是草花，用 B 表示得到的是黑桃，则 A，B 互斥. 全集 Ω 有 54 个元素，A 和 B 分别有 13 个元素. 不用写出具体的全集 Ω 和事件 A，B，也可以直接计算出

$$P(A)=\frac{13}{54},\quad P(B)=\frac{13}{54}.$$

由于 $A\cup B$ 也是事件，含有 26 个元素，所以

$$P(A\cup B)=\frac{26}{54}=\frac{13}{54}+\frac{13}{54}=P(A)+P(B).$$

概率的加法公式： 如果 Ω 的事件 A，B 互斥，则

$$P(A\cup B)=P(A)+P(B).$$

我们把概率的加法公式称为概率的**可加性**. 可加的前题是两个事件互斥.

证 设 Ω 有 n 个元素，A 有 m（$m\leqslant n$）个元素，B 有 k（$k\leqslant n$，且 $m+k\leqslant n$）个元素. 则 $P(A)=m/n$，$P(B)=k/n$. 由于 A，B 互斥，所以

$A\cup B$ 中元素个数 $=A$ 中元素个数 $+B$ 中元素个数 $=m+k$.

于是得到

$$P(A\cup B)=\frac{m+k}{n}=\frac{m}{n}+\frac{k}{n}=P(A)+P(B).$$

对立事件的概率公式： 如果 A 是全集 Ω 的事件，则

$$P(\Omega\backslash A)=1-P(A).$$

证 $\Omega\backslash A$ 和 A 互斥，并且 $\Omega=(\Omega\backslash A)\cup A$. 由概率的加法公式得到

$$P(\Omega)=P(\Omega\backslash A)+P(A).$$

再利用 $P(\Omega)=1$ 得到 $1=P(\Omega\backslash A)+P(A)$，于是 $P(\Omega\backslash A)=1-P(A)$.

例 3 在一副扑克的 54 张牌中随机抽取 1 张.

(1) 计算抽到是草花或黑桃的概率；

(2) 计算抽到的不是草花的概率；

(3) 计算抽到的不是草花也不是黑桃的概率.

解 (1) 用 A 表示抽到的是草花，用 B 表示抽到的是黑桃，则 $A\cup B$ 表示抽到的是草花或黑桃，并且 A，B 互斥.

因此由题意可得，$P(A)=\frac{13}{54}$，$P(B)=\frac{13}{54}$，所以

$$P(A\cup B)=P(A)+P(B)=\frac{13}{54}+\frac{13}{54}=\frac{13}{27}.$$

(2)$\Omega\backslash A$ 表示抽到的不是草花，是 A 的对立事件，所以

$$P(\Omega\backslash A)=1-P(A)=1-\frac{13}{54}=\frac{41}{54}.$$

(3) $C=A\cup B$ 表示抽到的是草花或黑桃，$\Omega\backslash C$ 表示抽到的不是草花也不是黑桃.

$$P(\Omega\backslash C)=1-P(C)=1-P(A\cup B)=1-\frac{13}{27}=\frac{14}{27}.$$

例 4 袋中有红球和白球各 1 个，每次抽 1 个，有放回地随机抽取 3 次. 计算：

(1) $A=$“至少有 1 个红球”的概率；

(2) $B=$“至少有 1 个白球”的概率；

(3) 有红球或有白球的概率.

解 (1) 用 BHB 表示第 1、第 2 和第 3 次分别取到白、红、白球等. 全集是

$$\Omega=\{BBB,BBH,BHB,HBB,BHH,HHB,HBH,HHH\},$$

$$A=\{BBH,BHB,HBB,BHH,HHB,HBH,HHH\},$$

因此，$P(A)=\frac{7}{8}=0.875$.

(2) 由于红球和白球处于对称的地位，所以 $P(B)=P(A)=0.875$.

(3) $A\cup B$ 表示有红球或有白球，这是必然事件，所以 $P(A\cup B)=1$.

例 5 彩票的中奖率是$\frac{1}{2}$，每次抽 1 张，有放回地随机抽取 3 次. 计算：

(1) $A=$“至少抽中 1 次”的概率；

(2) 1 次也没抽中的概率.

解 由于彩票的中奖率是$\frac{1}{2}$，因此可将中奖的彩票视为红球，不中奖的彩票视为白球，于是 A 等价于至少抽中 1 个红球.

(1) 由例 4 的结论知道 $P(A)=0.875$.

(2)由于 $\Omega\backslash A$ 是 1 次也没抽中的概率，因此

$$P(\Omega\backslash A)=1-0.875=0.125.$$

例 6 有 10 万张彩票，中奖率是 1/2. 每次抽 1 张，无放回地随机抽取 3 次，计算至少抽中 1 次的概率.

解 由于彩票数量很大，抽取一两张基本不会影响彩票的中奖比例，所以无放回抽奖的中奖概率和有放回抽奖的中奖概率基本是一样的. 因此，由例 5(1)可知，至少抽中 1 次的概率仍然是 0.875.

有人认为既然每次抽中的概率是 1/2，抽 2 次必然抽中，这是不对的. 上面的例 4、例 5 告诉我们，抽奖 3 次时，至少抽中 1 次的概率只有 0.875.

完全相同的道理，当彩票的中奖率是 1/100 时，你购买 100 张彩票中奖的概率是严格小于 1 的（实际上只有 0.634).

练 习

投掷两枚骰子，一枚是红色，一枚是蓝色. 计算以下事件的概率.

(1) $A=$“两枚骰子的点数相同”；

(2) $B=$“红色骰子的点数小于蓝色骰子的点数”；

(3) $C=$“两枚骰子的点数之和是 6”；

(4) $A\cup B$，$A\cup C$，$C\backslash B$.

习题 3

学而时习之

1. 假设每个人的生日在一年的 365 天中是等可能的. 在全校随机挑选一名同学，计算以下事件的概率.

   (1) 该同学的生日在 5 月份；

   (2) 该同学的生日在 5 月或 7 月份；

   (3) 该同学的生日是 1 日；

   (4) 该同学的生日是 1 日或 2 日.

2. 一批产品有 100 个，其中含有 10 个次品，从中随机抽取 1 个. 计算：

   (1) 这件产品是次品的概率；

   (2) 这件产品是正品的概率；

   (3) 这件产品是次品或是正品的概率.

3. 某电视台要招聘两名播音员，现在有三名符合条件的女士和两名符合条件的男士前来应聘. 如果每个应聘人员被录用的概率相同，计算以下概率.

   (1) 一名男士和一名女士被录用的概率；

   (2) 两名男士被录用的概率；

   (3) 两名女士被录用的概率.

4. 投掷两枚骰子，不用写出全集，计算以下概率.

   (1) 两枚骰子的点数相同；

   (2) 两枚骰子的点数之和是 6；

   (3) 两枚骰子的点数之和不是 6；

   (4) 至少一枚骰子的点数是 3.

5. 一个口袋内装有大小、质地均相同的 5 只球，其中 3 只为白球，2 只为黑球，从中一次摸出 2 只球. 求：

   (1) 一共可能出现多少种不同的结果；

   (2) 摸出的 2 只球均是白球的概率；

   (3) 摸出的 2 只球是 1 只白球与 1 只黑球的概率.

6. 如果全集 Ω 的事件 A，B，C 两两互斥，用 $A\cup B\cup C$ 表示事件 A，B，C 中至

少有一个发生. 证明：

$$P(A\cup B\cup C)=P(A)+P(B)+P(C).$$

7. 彩票的中奖率是 1/3，每次抽 1 张，有放回地随机抽取 3 张. 计算至少抽中 1 张的概率.

13.2.2 几何概率

例 1 在区间$[0,3)$中随机投掷一个质点，分别求质点落在区间$[0,1)$和$[1,2)$中的概率.

解 用$A=[0,1)$表示质点落在$[0,1)$中，用$B=[1,2)$表示质点落在$[1,2)$中，用$C=[2,3)$表示质点落在$[2,3)$中. 把A,B,C看作试验的元素时，试验的全集$\Omega=A\cup B\cup C=[0,3)$. A,B,C发生的可能性相同，所以

$$P(A)=\frac{1}{3}=\frac{A\text{ 的长度}}{\Omega\text{ 的长度}}.$$

$$P(B)=\frac{1}{3}=\frac{B\text{ 的长度}}{\Omega\text{ 的长度}}.$$

在概率的语言中，我们将“随机”解释成“等可能”.

几何概率定义 1 设试验的全集Ω是长度为正数的区间，A是Ω的子区间. 如果试验的结果随机地落在Ω中，则称

$$P(A)=\frac{A\text{ 的长度}}{\Omega\text{ 的长度}}$$

为事件A发生的概率，简称为A的概率.

在几何概率定义 1 中，并不指定所述的Ω和A是开区间、闭区间，还是半开半闭的区间. 这是因为区间(a,b)，$[a,b)$，$(a,b]$，$[a,b]$有相同的长度.

例 2 公共汽车在 0～5 min 内随机地到达车站.

(1) 求汽车第 3 min 到达车站的概率；

(2) 求汽车在 1～3 min 到达车站的概率.

解 试验的全集是 $\Omega=[0,5]$，集合 $A=[3,3]$ 表示汽车在第 3 min 时到达，$B=[1,3]$ 表示汽车在 1～3 min 到达. 根据几何概率定义：

$$P(A)=\frac{0}{5}=0,\quad P(B)=\frac{2}{5}.$$

在例 2 中也可以用 $\Omega=(0,5)$ 表示全集，用 $B=(1,3)$ 表示汽车在 1～3 min 到达. 计算的 $P(B)$是一样的.

下面把几何概率的定义推广到平面上. 我们把平面上的矩形、圆、椭圆等统一称为区域.

几何概率定义 2 设试验的全集 Ω 是面积为正数的区域，A 是 Ω 的子区域，如果试验的结果随机地落在 Ω 中，则称

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}$$

为事件 A 发生的概率，简称为 A 的概率.

几何概率也有如下的基本性质.

(1) $0\leqslant P(A)\leqslant 1$ （概率总是[0,1]中的数）；

(2) $P(\Omega)=1$ （必然事件的概率是 1）；

(3) $P(\varnothing)=0$ （不可能事件的概率是零）；

(4) 如果 A,B 互斥，则 $P(A\cup B)=P(A)+P(B)$；

(5) $P(A)+P(\Omega\backslash A)=1$ （对立事件概率之和等于 1）.

例 3 设雨点等可能地落在半径是 1 m 的圆 Ω 中，A 是半径为 0.5 m 的圆（如图 13－1）.

(1) 计算雨点落在小圆内的概率；

(2) 计算雨点落在小圆外的概率.

解 (1) 大圆的面积是 π m^2，雨点等可能地落入大圆. 小圆的面积是 $\pi\cdot 0.5^2$ m^2，雨点落入小圆的概率是

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}=\frac{\pi\cdot 0.5^2}{\pi}=0.25.$$

(2)根据几何概率的性质 5，雨点落入小圆外的概率是

$$P(\Omega\backslash A)=1-0.25=0.75.$$

例 4 陨石等可能地掉落在方圆为 200 km^2 的区域内，该区域内有面积为 80 km^2 的湖泊. 求陨石溅落在湖泊中的概率.

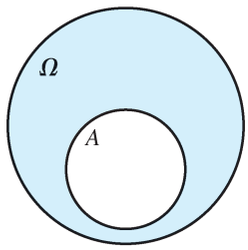


图 13－1

解 全集 Ω 的面积是 200 km^2，湖泊 A 的面积是 80 km^2. 元素等可能地落在 Ω 中，由几何概率的定义得到

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}=\frac{80}{200}=0.4.$$

陨石溅落在湖泊中的概率是 0.4.

练 习

每天的整点（如 9 时，10 时，11 时等）北京站都有列车发往天津. 一名乘客在 9 时至 10 时之间随机到达北京站. 计算：

(1) 他候车多于 20 min 的概率；

(2) 他候车恰好 15 min 的概率；

(3) 他候车少于 30 min 的概率；

(4) 他候车时间在 25～45 min 的概率.

习题 4

学而时习之

1. 一只麻雀随机地落在面积是 400 m^2 的广场上觅食. 广场内有一个长 20 m、宽 5 m的草坪，还有一个半径是 5 m 的圆形花坛. 计算：

(1) 麻雀落在草坪中的概率；

(2) 麻雀落在花坛内的概率；

(3) 麻雀落在草坪或花坛内的概率；

(4) 麻雀落在草坪或花坛外的概率.

2. 在长方形 $\Omega=A\cup B\cup C\cup D$ 中随机投掷一个质点（图 13－2）. 计算：

(1) $P(A)$；

(2) $P(A\cup B)$；

(3) $P(A\cup C\cup D)$；

(4) 证明 $P((A\cup B)\cap(A\cup C))=P(A\cup B)\cdot P(A\cup C)$.

A	B
C	D

图 13－2

3. 在区间 [0，1] 中随机地投掷一点，计算该点落在 [0，0.3] 中的概率.

4. 在区间 [0，1] 中随机地取两点，用 A 表示它们的平方和小于 1.

(1) 写出试验的全集 Ω；

(2) 用集合表示出事件 A；

(3) 计算 $P(A)$.

5. 在区间[1,2]中随机地投掷两个点,用 B 表示它们的差的绝对值小于 1/3.

(1) 写出试验的全集 Ω；

(2) 用集合表示出事件 B；

(3) 计算 $P(B)$.

6. 两人在某天的 1 时至 2 时间各自独立随机到达某地会面，先到者等候 20 min 后离去.

(1) 写出试验的全集 Ω；

(2) 用集合表示出事件 $B=$“两人相遇”；

(3) 计算这两人能相遇的概率.

13.3 频率与概率

设 Ω 是某个试验的全集，A 是 Ω 的事件. 在相同的条件下将该试验独立地重复 N 次，我们称

$$f_N=\frac{N\text{ 次试验中 }A\text{ 发生的次数}}{N}$$

是 N 次独立重复试验中，事件 A 发生的**频率**.

这个结论首先由伯努利给出数学的证明.

理论和事实都证明：在相同的条件下，将一试验独立重复 N 次，

用 f_N 表示事件 A 在这 N 次试验中发生的频率. 当 N 增加时，f_N 将在一个固定的数值 p 附近波动，这个数值 p 就是事件 A 的概率 $P(A)$. 于是，f_N 是 $P(A)$的估计.

历史上许多著名的统计学家对概率和频率的关系进行过验证. 他们的试验结果总结在表 13.1 中.

表 13.1

试验者	掷币次数 N	正面朝上次数	频率 f_N
德·摩根	2 048	1 061	0.518 1
蒲丰	4 040	2 048	0.506 9
凯瑞	7 000	3 516	0.502 2
凯瑞	9 000	4 538	0.504 2
费勒	10 000	4 979	0.497 9
皮尔逊	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5
罗曼诺夫斯基	80 640	40 173	0.498 2

现在的随机试验工作可以在计算机上方便地进行.

例 1 表 13.2 是用计算机进行的投掷一枚均匀的骰子的试验总结. 其中 N 是试验的次数，表中的百分数是频率. 例如表中第 3 行第 2 列的 15.00%，表示试验次数 $N=10^2$ 时，点数 2 出现的频率是 15.00%.

表 13.2

点数	$N=10^2$	$N=10^3$	$N=5000$	$N=10^4$	$N=10^5$	$N=10^6$
1	17.00%	16.50%	16.28%	16.61%	16.72%	16.69%
2	15.00%	15.50%	17.12%	16.62%	16.44%	16.62%
3	18.00%	17.10%	16.78%	16.94%	16.84%	16.69%
4	18.00%	16.00%	16.68%	16.97%	16.76%	16.64%
5	13.00%	16.60%	15.50%	15.94%	16.69%	16.64%
6	19.00%	18.30%	17.64%	16.92%	16.55%	16.72%

从表 13.2 可以看出，当试验的次数逐步增加时，每个点数出现的频率都向概率 $1/6\approx16.67\%$靠近.

例 1 中的计算机试验称为计算机**模拟试验**. 计算机模拟试验还可以解决很多其他的计算问题.

例 2 （利用几何概率估算圆周率 π）在平面上作一个边长是10 cm

的正方形，在正方形内作一个半径等于 5 cm 的圆，见图13-3.

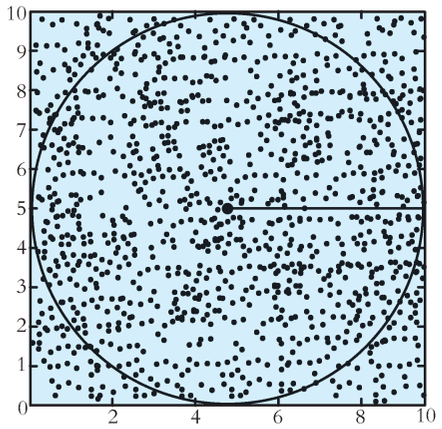


图 13-3　在 Ω 内随机投掷的1 000个质点

（1）在该正方形内随机投掷 1 个质点，计算质点落入圆 A 的概率；

（2）利用计算机模拟的方法估计 π 的值.

解　（1）试验的全集是正方形

$$\Omega=\{(x,y)\,|\,0\leqslant x\leqslant 10,0\leqslant y\leqslant 10\}.$$

事件

$$A=\{(x,y)\,|\,x^2+y^2\leqslant 25\}$$

是 Ω 的子集，根据圆面积的计算公式和几何概率定义 2 得到

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}=\frac{25\pi}{100}=\frac{\pi}{4}.$$

（2）如果 π 是未知的，可以用如下的方法进行模拟计算.

独立重复地在 Ω 中投掷 N 个质点，对于较大的 N，质点落入圆 A 的频率

$$f_N=\frac{\text{落入 }A\text{ 的质点数}}{N}.$$

由频率和概率的关系知道 f_N 是 $P(A)$ 的近似，所以对较大的 N，

$$f_N\approx\frac{\pi}{4}.$$

f_N 是可以计算的，于是

$$\hat{\pi}=4f_N$$

是 π 的估计.

利用计算机在 Ω 中随机投掷 $N=10^2$，10^3，10^4，10^5 个质点，见图 13－3，把依次得到的 $\hat{\pi}$ 列入表 13.3. 从表 13.3 可以看出，对于较大的 N，$\hat{\pi}$ 对 π 的近似是不错的.

表 13.3

N	10^2	10^3	10^4	10^5
$\hat{\pi}$	3.080	3.148	3.160	3.149

为了看清计算机模拟结果的随机性，再次进行模拟计算时，得到的结果如表 13.4 所示.

表 13.4

N	10^2	10^3	10^4	10^5
$\hat{\pi}$	3.163	3.221	3.121	3.144

例 3 A 是平面上的不规则区域，作一个长 12 m、宽 8 m 的矩形 Ω，使得 $A\subseteq\Omega$（图 13－4）. 利用计算机在 Ω 中随机投掷了 2 万个质点后，发现有 1.12 万个质点落入区域 A 中，估算 A 的面积.

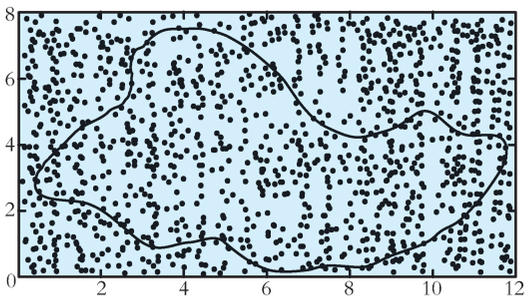


图 13－4 估算 A 的面积

解 质点落入 A 的频率是

$$f_N=\frac{1.12}{2}=0.56.$$

根据频率和概率的关系知道

$$f_N\approx\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}.$$

所以，

$$A\text{ 的面积}\approx f_N\times\Omega\text{ 的面积}=0.56\times12\times8=53.76\ (\mathrm{m}^2).$$

练 习

A 是平面上的不规则区域，作一个半径为 12 cm 的圆 Ω，使得 $A\subseteq\Omega$. 在 Ω 中随机投掷了 3 000 个质点后，发现有 1 440 个质点落入区域 A 中，估算 A 的面积（结果精确到小数点后两位）.

习题 5

学而时习之

1. 某老师在某大学连续 3 年主讲高等数学这门课，3 年来学生学习这门课的成绩汇总如下：

成　绩	人　数
90 分以上	50
80～89 分	180
70～79 分	260
60～69 分	90
60 分以下	60

学生甲下学期将学习该老师的高等数学课，用已有的信息估计他得以下分数的概率：

（1）90 分以上；　（2）60～69 分；　（3）60 分以上.

2. 某人捡到不规则形状的五面体石块，他将每个面分别标上 1，2，3，4，5 后，投掷了 100 次，并且记录了每个面落在桌面上的次数（如下表）. 如果再投掷一次，请估计标记为 4 的这一面落在桌面上的概率是多少.

石块的面	1	2	3	4	5
频　数	32	18	15	13	22

数学文化

概率简史

概率的概念形成于16世纪，与用投掷骰子的方法进行赌博有密切的关系.

重复投掷一枚硬币1万次，你会得到什么结果呢？如果硬币是均匀的，你会判断正面出现的频率大约是1/2吗？初看起来这是一个简单的问题，数学上首先证明这个结论的人是**伯努利**（Bernoulli），尽管他说：哪怕最笨的人，不通过别人的教诲也能理解频率大约是1/2. 但是要在数学上证明它却不容易.

伯努利1654年出生于瑞士的巴赛尔. 在他的家族成员中，程度不同地对数学的许多方面作出过贡献，其中至少有5人在概率论方面作出过贡献. 他的父亲希望他成为神职人员，但是伯努利自己更喜欢数学，他和同时代的牛顿等人保持密切的通信联系. 现在国际上的伯努利统计期刊和伯努利统计学会就是以他的名字命名的.

学习数学的人对**费马**（Fermat，1601－1665）是不陌生的. 因为“费马大定理”在前些年得到证明，费马的名声早已传播到数学的领域之外. 但是费马和概率论的关系并不为很多人所了解.

费马和**笛卡儿**（Descartes，1596－1650）同享发明解析几何的荣誉，但是费马最重要的研究工作是在数论方面. 费马不写论文发表，只是通过书信的形式和朋友们交流数学研究的思想和成果. 他和**帕斯卡**（Pascal，1623－1662）的通信是建立概率论的数学基础的起点.

帕斯卡出身于贵族家庭，16岁时就发表了圆锥曲线方面的数学论文. 为了帮助他父亲管理账目，他还发明了一个早期的计算机. 帕斯卡对于概率论的贡献体现在他和费马的通信中.

促使帕斯卡和费马通信的人是德梅尔，他向帕斯卡请教几个有关赌博的问题. 1654 年 7 月 29 日帕斯卡首先给费马写信，转达了德梅尔的以下问题：投掷两个骰子 24 次，至少掷出一对 6 的概率小于 1/2. 这个概率实际上近似等于 0.491 4.

概率论的数学理论基础是由著名的苏联数学家**柯尔莫哥罗夫**(Kolmogorov,1903—1987)在 1933 年建立的.

在我国，**许宝騄**教授是概率论和统计学研究的先驱，有很大的学术成就，在国际上享有盛誉，对概率论和统计学作出了杰出的贡献. 1979 年，世界著名的统计期刊《**数理统计年鉴**》(*The Annals of Statistics*) 邀请了一些著名学者撰文介绍他的生平，高度评价了他在概率论和统计学两方面的研究工作.

多知道一点

使用计算机模拟随机试验

利用计算机和 MatLab 进行计算机模拟时，可以仿照下面程序进行.

1. 用计算机进行投掷 10^3 次硬币试验，计算正面出现的频率时，直接输入下面的语句. 括弧的内容是语句的解释，不输入.

N＝10^3；

n＝unidrnd（2，1，N）－1（产生 10^3 个取值 0 或 1 的随机数，相当于投掷 10^3 次硬币）；

S＝sum（n）（计算正面朝上的次数）；

M＝mean（n）（计算正面出现的频率）.

2. 用计算机进行独立重复投掷一枚均匀的骰子的试验时，用下面的语句.

n＝unidrnd（6，1，10^3）（产生 10^3 个 1～6 中的随机数）；

tabulate（n）（计算出现的次数和频率）.

3. 利用计算机和几何概率估算圆周率 π 时，直接输入以下语句：

x＝rand（2，10^3）*10；（在［0，10］×［0，10］中投掷 10^3 个质点）

n＝zeros（1，10^3）；

for j＝1：10^3

if （x(1,j)－5)^2＋(y(1,j)－5)^2＜25

n（j）＝1；

end

end

M＝mean（n）*4（计算 $\hat{\pi}$）.

数学实验

用计算机模拟随机试验

打开用“Z＋Z 超级画板”制作的课件“圆中投豆．zjz”，屏幕画面如图 13－5：

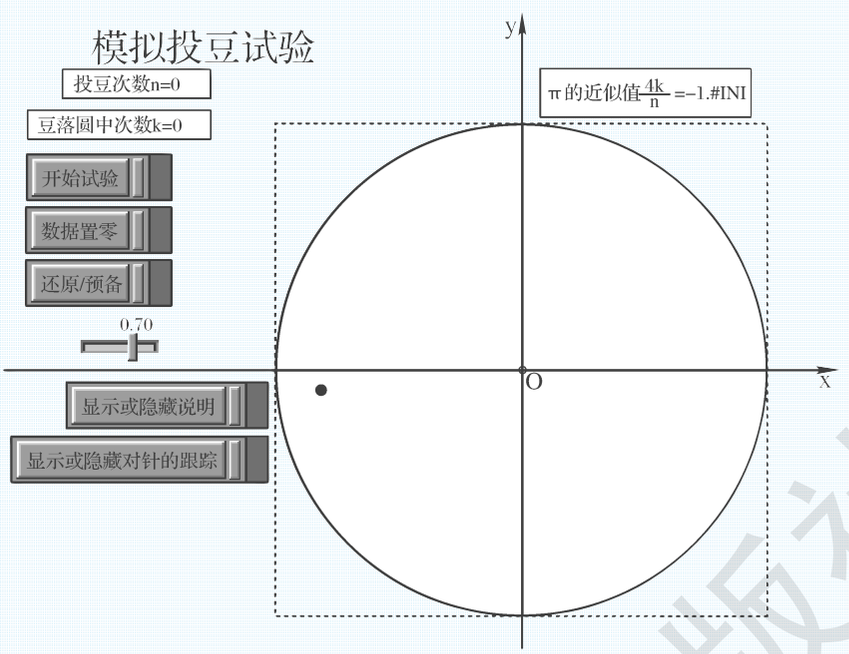


图 13－5

鼠标单击右下方“显示或隐藏说明”按钮，仔细阅读出现的文本内容；根据操作说明，单击灰色按钮上的副钮使还原；待变量尺指向 0，再单击蓝色按钮两次使数据置零；单击灰色按钮上的主钮做预备；待变量尺指向 1，单击绿色按钮开始试验．

图 13－6 是试验若干次后的画面．

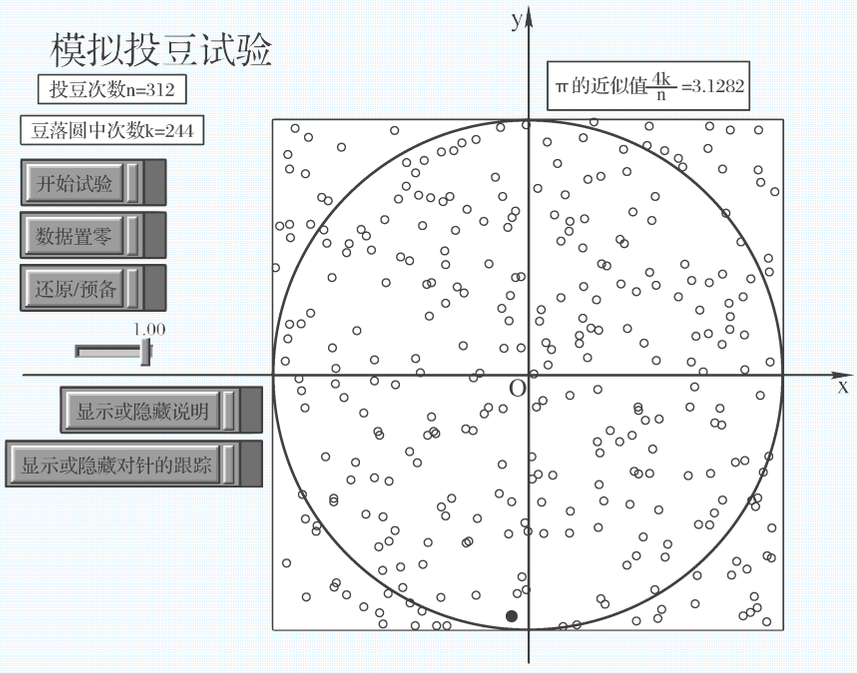


图 13－6

利用几何概率求圆周率的另一个著名的试验是投针试验. 在纸上画一些等距离的平行直线，使相邻直线距离为 d. 将一枚长度为 $a(a<d)$ 的针随机地投到纸上，可以算出针和直线相交的概率为

$$p=\frac{2a}{\pi d}$$

想一想，若在 n 次投针中和直线相交 k 次，当 n 较大时相交频率与 π 有何关系？

打开用“Z＋Z 超级画板”制作的课件“投针试验 . zjz”，屏幕画面如图 13－7：

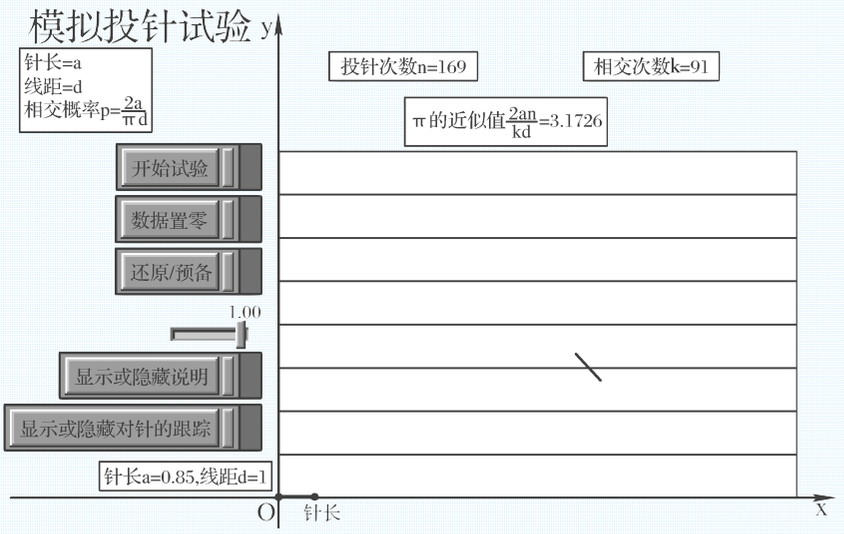


图 13－7

单击“显示或隐藏说明”按钮，仔细阅读出现的说明，依法操作，模拟投针若干次后的效果类似图 13－8：

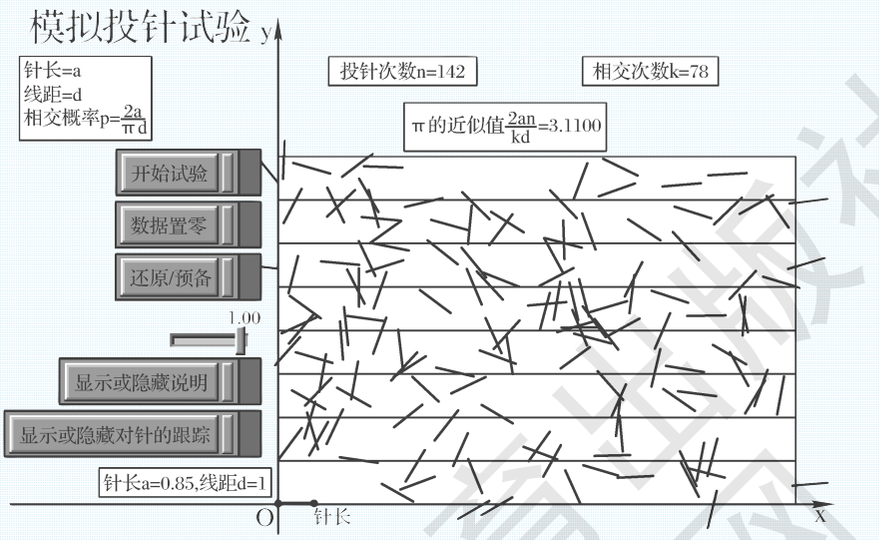


图 13－8

下面举例说明，如何在“Z＋Z 超级画板”的程序工作区直接计算模拟随机试验.

根据学过的算法知识，不难理解下面的几个程序. 注意其中使

用的函数 rand(a，b)，可以每次随机地产生 a，b 之间的一个具有指定的有效数字的实数.

(1) 模拟投掷硬币输出正面向上的频率的函数 yb(m)，其中 m 是投掷次数.

键入程序：

```
yb(m) {k=m; n=0;
  while (k>0)
    {k=k-1;
     n=n+sign (rand (0, 1), 0.5);}
   n/m;}
```

执行（用 Ctrl+Enter 键，下同）后返回：

≫yb(m) #

如希望将频率表示成小数，可键入

Float(1);

返回：

≫计算结果显示浮点数 #

要模拟试验 1 000 次，可键入

yb (1000);

执行后返回：

≫(503)/(1000)=0.503 #

(2) 模拟掷骰子 m 次，输出 d 点向上的频率的函数 sz(m)：

```
sz(m,d){k=m; n=0;
          while(k>0)
            {k=k-1;
       if (floor(rand(1,7))==d) {n=n+1;}}
         n/m;}
```

键入程序执行后返回：

≫sz(m,d)#

要模拟 300 次投掷，求出现 2 点的频率，键入程序：

sz(300,2);

执行后返回：

≫(47)/(300)＝0.156667 ＃

(3) 模拟投豆试验用几何概率求圆周率的近似值的程序.

```
geo(m){
k＝m; n＝0;
while (k＞0)
{k＝k－1;
if((rand(0,10)－5)^2＋(rand(0,5)－5)^2＜25) {n＝n＋1;}}
4＊n/m;}
```

键入程序执行后返回：

≫geo(m) ＃

要模拟投豆 1 000 次的试验，只要输入

```
geo(1000);
```

执行后返回：

≫(388)/(125)＝3.104 ＃

当然，你在计算机上运行的结果和这里可能不同.

有了这些程序，你可以在不长的时间做大量的有关随机数的试验了.

小结与复习

1. 元素：是试验的可能结果，也称为样本点或基本事件.

2. 全集：是试验的元素的集合，常用 Ω 表示. 也称为样本空间.

3. 事件：是全集的子集.

4. 古典概型：设全集 Ω 中有 n 个元素，事件 A 包含了 m 个元素. 如果 Ω 的每个元素发生的可能性相同，就称

$$P(A)=\frac{m}{n}$$

是事件 A 的概率，称这个模型是古典概型.

5. 几何概率 1：设试验的全集 Ω 是长度为正数的区间，A 是 Ω 的子区间. 如果试验的结果随机地落在 Ω 中，则称

$$P(A)=\frac{A\text{ 的长度}}{\Omega\text{ 的长度}}$$

为事件 A 的概率.

6. 几何概率 2：设试验的全集 Ω 是面积为正数的区域，A 是 Ω 的子区域. 如果元素随机地落在 Ω 中，则称

$$P(A)=\frac{A\text{ 的面积}}{\Omega\text{ 的面积}}$$

为事件 A 的概率.

7. 概率的性质：

(1) $0\leqslant P(A)\leqslant 1$；

(2) $P(\Omega)=1$；

(3) $P(\varnothing)=0$；

(4) 如果 A，B 互斥，则 $P(A\cup B)=P(A)+P(B)$；

(5) $P(A)+P(\Omega\backslash A)=1$.

8. 概率和频率：在相同的条件下，将一试验独立重复 N 次，用 f_N 表

示事件 A 在这 N 次试验中发生的频率. 当 N 增加时，f_N 将在一个固定的数值 p 附近波动，这个 p 就是事件 A 的概率 $P(A)$. f_N 是 $P(A)$ 的估计.

复习题十三

学而时习之

1. 口袋中有标号 1～5 的球各 1 个. 为以下的试验写出全集.

   (1) 从中任取 1 个；

   (2) 从中一次任取出 2 个.

2. 投掷一枚骰子和两枚硬币，写出全集.

3. 同时投掷一枚骰子和一枚硬币，计算概率.

   (1) 硬币是正面，骰子的点数是 3；

   (2) 硬币是正面，骰子的点数是 2 或 4.

4. 豌豆的高矮性状的遗传由其一对基因决定，其中决定高的基因记为 D，决定矮的基因记为 d，则杂交所得第一子代的一对基因为 Dd. 若第二子代的 D，d 基因的遗传是等可能的，求第二子代为高茎的概率（只要有基因 D 则其就是高茎，只有两个基因全是 d 时，才显现矮茎）.

5. 将一枚骰子先后抛掷 2 次，观察向上的点数，问：

   (1) 共有多少种不同的可能结果？

   (2) 点数之和是 3 的倍数的可能结果有多少种？

   (3) 点数之和是 3 的倍数的概率是多少？

6. 投掷两枚骰子，计算以下事件的概率.

   (1) A＝“两枚骰子的点数之和是 2”；

   (2) B＝“两枚骰子的点数之和是 4”；

   (3) C＝“两枚骰子的点数之和是 6”；

（4）以上三个事件中，哪个概率最大，为什么？

7. 一批产品有 30 个，其中含有 3 个次品，从中随机抽取 1 个．计算：

（1）这个产品是次品的概率；

（2）这个产品是正品的概率．

8. 黄色人种种群中各种血型的人所占的比如下表所示：

血　型	A	B	AB	O
该血型的人所占比/%	28	29	8	35

已知同种血型的人可以输血，O 型血可以输给任一种血型的人，任何人的血都可以输给 AB 型血的人，其他不同血型的人不能互相输血．小明是 B 型血，若小明因病需要输血，问：

（1）任找一个人，其血可以输给小明的概率是多少？

（2）任找一个人，其血不能输给小明的概率是多少？

9. 某射击手在同一条件下进行射击，结果如下：

射击次数（n）	10	20	50	100	200	500
击中靶心次数（m）	8	19	44	92	178	455
击中靶心频率$\left(\frac{m}{n}\right)$						

（1）计算表中击中靶心的各个概率；

（2）这个射手射击一次，击中靶心的概率约是多少？

温故而知新

10. 如果全集 Ω 的事件 A，B，C，D 两两互斥，用 $A\cup B\cup C\cup D$ 表示事件 A，B，C，D 中至少有一个发生．证明：

$$P(A\cup B\cup C\cup D)=P(A)+P(B)+P(C)+P(D).$$

11. 设 Ω 是长 2 cm、宽 3 cm 的长方形，A 是以长方形对角线交点为圆心，以 2 cm 长为直径的圆．如果质点等可能地落在 Ω 中，

（1）计算质点落在 A 内的概率；

（2）计算质点落在 A 外的概率．

12. 在区间［1，3］中随机地投掷两个质点，计算这两个质点都落在［1，2］中的概率．

13. 两人在某天的 5 时至 7 时间相互独立随机到达某地会面，先到者等候 30 min 后离去.

（1）写出试验的全集 Ω；

（2）用集合表示出事件 $B=$“两人相遇”；

（3）计算这两人能相遇的概率.

上下而求索

智者千虑，必有一失！

14. 一金融公司的主要工作是进行投资，尽管每次投资前都有缜密的投资分析，但是投资失败的概率仍保留在 5%. 利用频率和概率的关系，说明该公司一次次相互独立的投资一定有失败的时候.

15. 某射击运动员脱靶的概率是 0.01%，如果他独立重复射击下去，必有一次脱靶发生.（利用频率和概率的关系说明.）

愚者千虑，必有一得！

16. 张三和好友李四下棋时，赢李四的概率只有 10%. 张三不服输，不断约李四下棋. 试说明张三总有赢棋的时候.

这个赌徒破产的概率是 95%. 如果这个赌徒的欲望增加，他破产的概率会跟着增加.

是赌徒就要破产！

17. 一个赌徒手中有 1 000 元本金，赌博时每次赌注是 1 元，输赢的概率都是 1/2. 在赌博期间，一旦输光则宣告破产. 现在该赌徒决心赢到手中有 2 万元后再停止赌博. 你认为他的目的可以达到吗？（对掷硬币的试验再次理解后给出答案.）

附 录

数学词汇中英文对照表

（按词汇所在页码的先后排序）

中文名	英 文 名	页 码
算法	algorithm	2
顺序结构	sequence structure	7
条件结构	conditional structure	10
循环结构	cycle structure	14
输入语句	input statement	21
输出语句	output statement	21
赋值语句	assignment statement	22
条件语句	conditional statement	24
循环语句	cycle statement	30
总体	population	60
个体	individual	60
均值	mean	61
样本	sample	62
观测数据	observed data	62
样本量	sample size	62
抽样	sampling	62
估计	estimator	63
方差	variance	65
标准差	standard deviation	67
随机数	random number	72
层权	weight	79
系统抽样方法	systematic sampling method	80
频率	frequency	82
频率分布表	frequency distribution table	83

直方图	histogram	86
茎叶图	stemplot	90
散点图	scatter diagram	94
参数	parameter	101
随机事件	random event	117
事件	event	117
样本点	sample outcome	117
样本空间	sample space	117
概率	probability	121
数理统计年鉴	*The Annals of Statistics*	136

后　记

本套教科书是按照《基础教育课程改革纲要(试行)》的精神和要求，以《普通高中数学课程标准（实验)》为依据进行编写，经国家基础教育课程教材专家工作委员会2005年审查通过.

本书在实验过程中，得到了广大数学家、数学课程专家、教研人员以及一线教师的大力支持和帮助，我们对他们的辛勤付出表示衷心的感谢.同时，我们还要特别感谢重庆市教育科学研究院张晓斌、重庆市巴蜀中学校费春斌、宋晓宇、饶志明、马洪超、来章润，育才中学校邓启华等老师，他们为本书的修订提出了宝贵的意见，在此，我们一并表示衷心的谢意.

教材建设是一项长期的任务，我们真诚地希望广大教师、学生在使用本册教科书的过程中提出宝贵意见，并将这些意见和建议及时反馈给我们. 让我们携起手来，共同完成基础教育教材建设这一光荣的使命!

湖南教育出版社

普通高中课程标准实验教科书
数　　学
第五册（必修）
责任编辑：邹楚林
湖南教育出版社出版发行（长沙市韶山北路 443 号）
电子邮箱：hnjycbs@ sina. com
客　　服：电话 0731 – 85486979
重庆市新华书店经销
湖南天闻新华印务邵阳有限公司印刷
890 × 1240　16 开　印张：10　字数：250000
2005 年 8 月第 1 版　2019 年 7 月第 2 版第 18 次印刷
ISBN 978 – 7 – 5355 – 4601 – 2
定价：8. 55 元
批准文号：渝发改价格[2019]946 号　举报电话：12358