SCHAUM'S outlines

全美经典学习指导系列

统计学原理

（上册）

——描述性统计学与概率

〔美〕S. 伯恩斯坦　R. 伯恩斯坦 著

史道济 译

- 获取高分的最佳助手
- 有关数学知识的总复习
- 469道完全解答的习题
- 涵盖本课程的所有基础，是任一教材的补充
- 理想的自学读物

科学出版社

麦格劳-希尔教育出版集团

内 容 简 介

本书论述工程技术、自然科学和生命科学中常用的统计学原理和方法. 全书分为两册,上册内容包括学习本书所需的数学知识、描述性统计学的基本原则和方法以及推断性统计学的理论基础,概率论.

本书每一章都有相同的形式:第一部分以大纲的形式论述所有的新概念和新方法以及有完整解答的例子. 第二部分是习题解答,包括许多理论的应用. 第三部分是补充习题,只有答案,这部分内容是检验读者对本书内容的理解程度.

本书内容丰富,论述严谨. 为了自学的方便,全书有相互参照的系统可以很快找到要学习的内容.

本书适合于高等院校理工科教师与学生和有关工程技术人员阅读.

Schaum's Outlines
Stephen Bernstein and Ruth Bernstein: Elements of Statistics I: Descriptive Statistics and Probability
ISBN:0-07-005023-6

图字:图字 01-2001-2119 号

图书在版编目(CIP)数据
统计学原理(上册)——描述性统计学与概率/〔美〕S. 伯恩斯坦,〔美〕R. 伯恩斯坦著;史道济译. -北京:科学出版社,2002. 1
(全美经典学习指导系列)
ISBN 7-03-009771-8
Ⅰ. 统… Ⅱ. ①伯…②伯…③史… Ⅲ. 统计学 Ⅳ. C8
中国版本图书馆 CIP 数据核字(2001)第 063367 号

科学出版社 出版
北京东黄城根北街16号
邮政编码:100717
http://www.sciencep.com
双青印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2002 年 1 月第 一 版 开本:A4(890×1240)
2002 年 1 月第一次印刷 印张:15 3/4
印数:1—5 000 字数:447 000
定价:24.00 元
(如有印装质量问题,我社负责调换〈环伟〉)

前　言

统计学是论述收集、分析并解释数字信息的科学. 对这门科学有一个基本的了解，不仅对每个做研究工作的科学家是重要的，而且对现代社会中许多人都是重要的，他们必须处理这种信息：如评价有争议的医学研究报告的医生，企图使陪审团相信定量证据是合法的律师、改进质量控制程序的制造商，解释市场趋势的经济学家等等.

统计科学的理论基础是数学中称为数理统计的学科，这里统计学是抽象的公理、定理和严谨的证明紧密结合而形成的完整的结构. 为了使这种理论结构也适用于非数学家，产生了一个称为一般统计学的解释性的学科，其中的描述是最简化的，且常常是非数学的. 每个特殊领域（例如农学、人类学、生物学、经济学、工程学、心理学、社会学）都从这个简化的形式得到了适合于各自数据的材料. 例如有一种称为生物统计学的、就是特别适合于生物学数据资料的一般统计学.

在一般统计学或它的某些特殊分支中，所有的介绍性课程都有相同的核心：统计学基础，多年来，本书作者学习了这些课程的基础，在研究项目中也应用过，并且教过一般统计学和生物统计学课程. 在此经历上写成的本书是一本统计学基础的自学指导，可以自己阅读，也可作为教材的补充，而且由于本书内容完整，实际上还可作为教材使用.

统计科学可以分为两部分：描述性统计与推断性统计. 描述性统计提供了将原始数据整理成有用形式的方法，这些方法包括收集、整理、概括、描述及给出数据信息. 如果在研究中始终利用全部资料（总体），那么要求描述性统计. 但有时可以利用的只是总体的一小部分（样本），此时要求由有限的、不确定的样本信息，对整个总体做出判断及决策的方法，这是推断性统计的领域.

在介绍一般统计学的所有课程中，都以标准的次序给出统计学的这两个部分，本书也依这个次序，但将它们分为二册. 上册（第一—第十章）讲述为学习本书所必须的数学（大学代数的概念），描述性统计的基本原则、方法及推断性统计的主要理论基础：概率论. 下册（第十一—第二十章）论述推断性统计的概念和方法. 全书每章都有相同的形式：第一部分是正文，以大纲的形式论述所有的新概念和新方法以及有完整解答的例子. 下一部分是习题解答，复习同样的材料且让你考虑来自不同视角的材料，最后一部分是补充习题，检查你对内容的掌握，只有答案，而没有解答步骤. 因为这是一本关于一般统计的书，我们努力对代表许多特殊领域的问题作出形式多样的选择，而且还希望用这些问题去说明，如何由实际问题——正在解决的事情的数据信息作出决策.

为了掌握统计学，你必须阅读课文并做习题，我们建议你首先阅读正文及随后的例子，然后在做习题解答及补充习题前，回过头来再看一次正文. 这本书全书有相互参照系统，使你可以很快的复习对理解后面内容所需要的先前的内容.

如果你继续学习统计学，你可能会使用计算机及许多统计程序包中的一个，本书并不论述如何使用这类计算程序，但会告诉你为理解使用程序所要求的概念，以及同样重要的，解释计算机输出结果. 不要求用计算机做本书中的习题，所有的习题只用电子计算器即可解决.

我们感谢对本书出版作出特殊贡献的 McGraw-Hill 公司的以下人员：Barbara Gilson, Elizabeth Zayatz, John Aliano, Fred Perkins, Arthur Biderman, Mary Loebing Giles 以及 Meaghan McGovern，我还要感谢 E. Kirk 允许翻印初等统计学第二版的表 D. 10，感谢所有允许我们使用他们出版物的个人和组织（在材料出处特别提到的），我们也感谢各章不知名的评论员.

目　录

第一章　统计学的数学基础

1.1　什么是统计学

统计学是关于数据资料的收集、整理、分析和推断的一门科学.它可分为**描述统计学**和**推断统计学**两大类.描述统计学给出的是将原始数据资料加工成有用图表的方法，这些方法包括数据的收集、整理、概括和描述等.如果在研究中可以得到整个**总体**，那么描述统计学就足够了.但是，实际中往往只能得到总体的一小部分(称为**样本**)，这就需要通过这些样本的有限的、不确定的信息来确定有关总体的信息，这就是推断统计学的研究领域.

统计学的理论基础是**数理统计学**.数理统计学是数学的一个分支，由一系列的公理、定理以及严格证明来组成.它还涉及到数学的其他一些领域，例如微积分、概率论和高等代数等等.为了使这些理论也适用于一般研究者，人们将其简化，变得非数学化，由此产生了**一般统计学**.不同的专业领域(如建筑学、人类学、生物学、经济学等等)与一般统计学结合，就产生了相应的专业统计学.例如，**生物统计学**即为专门适用于对生物学中的数据资料进行分析的一般统计学.

这本书由两册组成，主要通过许多专业领域的实例和问题来介绍一般统计学.本书采用的是一般统计学课程的典型提纲：首先介绍描述统计学——数据的收集(第二章和第三章)、整理(第四章)、图形化(第五章)和描述(第六章和第七章)；然后介绍概率论(第八章到第十二章)和抽样理论(第十三章)；最后是推断统计学的估计理论和假设检验理论(第十四章到第二十章).

本书要求读者有大学的数学基础，在第一章复习所需的代数基本知识，假定你有一个电子计算器.

1.2　分数运算

若用 m, n 表示两个数，那么 m 与 n 的乘积可以用如下的**等价符号**表示：$m\times n$, $(m)(n)$ 和 $m\cdot n$.类似地，m 除以 n 的商可以等价地表示为 $m\div n$, m/n 和 $\frac{m}{n}$.

例 1.1　设 $m=4$, $n=2$，计算下列式子：$m\times n$, $m\div n$, m/n, $(m)(n)$, $m\cdot n$, $\frac{m}{n}$.

解☞

$$4\times2=(4)(2)=4\cdot2=8$$

$$4\div2=4/2=\frac{4}{2}=2$$

分数运算中，分子、分母同乘以或除以相同的数，分数的值不变.但是分子、分母都加上或减去相同的数，分数的值通常会发生变化.

例 1.2　下列哪些分数与 $\frac{6}{8}$ 是等价的：$\frac{8}{9}$, $\frac{12}{16}$, $\frac{2}{4}$, $\frac{3}{4}$, $\frac{16}{18}$, $\frac{1}{2}$.

解☞

$$\frac{6}{8}=\frac{(2)(6)}{(2)(8)}=\frac{12}{16}=\frac{6/2}{8/2}=\frac{3}{4}$$

而

$$\frac{6}{8}\neq\frac{6-4}{8-4}=\frac{2}{4}=\frac{1}{2}\text{ 和 }\frac{6}{8}\neq\frac{6+10}{8+10}=\frac{16}{18}=\frac{8}{9}$$

分数的加减运算必须转化为同分母才可以进行.分数相乘等于分子分母分别相乘.分数相除等于除数乘以被除数的倒数.

例 1.3　计算下列各式：(a) $\frac{4}{5}+\frac{5}{6}$，(b) $\frac{5}{7}\times\frac{1}{4}\times\frac{2}{3}$，(c) $\frac{f}{g}\div\frac{m}{n}$.

解☞　(a) $\frac{4}{5}+\frac{5}{6}=\frac{4\times6}{5\times6}+\frac{5\times5}{6\times5}=\frac{24}{30}+\frac{25}{30}=\frac{49}{30}=1\,\frac{19}{30}$

(b) $\frac{5}{7}\times\frac{1}{4}\times\frac{2}{3}=\frac{5\times1\times2}{7\times4\times3}=\frac{10}{84}=\frac{5}{42}$

(c) $\frac{f}{g}\div\frac{m}{n}=\frac{f}{g}\times\frac{n}{m}=\frac{fn}{gm}$

1.3 带符号数的运算

同号数相加,取原来的符号,并把数值相加.异号数相加,用数值大的减去数值小的,并取数值大的符号.带符号数相减等于被减数改变符号后相加.

例 1.4 对下列带符号数进行加减运算:(a) $5+7+9+2$,(b) $5+(-7)+9+(-2)$,(c) $(-5)-(+7)+9-(-2)$.

解 ☞ (a) $5+7+9+2=23$

(b) $5+(-7)+9+(-2)=14-9=5$

(c) $(-5)-(+7)+9-(-2)=-5-7+9+2=-1$

两数相乘或相除,同号得正,异号得负.

例 1.5 对下列带符号数进行乘法和除法运算:(a) $(-4)(-2)$,(b) $(-4)(2)$,(c) $(-4)\div(-2)$,(d) $(-4)/(2)$.

解 ☞ (a)$(-4)(-2)-8$

(b)$(-4)(2)=-8$

(c)$(-4)\div(-2)=2$

(d)$(-4)/(2)=-2$

若一个算式中含有加、减、乘、除运算时,先做乘除法,再做加减法.如果有括号,就先做括号中的运算.

例 1.6 计算:(a)$2\times10-9$,(b)$[(-2)+(-3)]\div[(-2)+(-3)]$.

解 ☞ (a)$2\times10-9=20-9=11$

(b)$[(-2)+(3)]\div[(-2)+(-3)]=[1]\div[-5]=-0.2$

1.4 舍入运算

用舍入原则进行取整运算中,若小数部分小于 0.5 则舍去小数部分,整数部分保持不变;若小数部分大于 0.5 则去掉小数部分,并且整数部分加 1;若小数部分恰好等于 0.5,则通常采用如下原则:当个位数为奇数时,整数部分加 1,当个位数为偶数时,整数部分保持不变.

例 1.7 对下列数做取整运算:(a) 2.2,(b) 1.89,(c) 2.5,(d) 1.50.

解 ☞ (a) 2.2 舍入到 2

(b) 1.89 舍入到 2

(c) 2.5 舍入到 2

(d) 1.50 舍入到 2

将数舍入到一个小数位置的运算与上述取整运算的基本原则相同,只是用在小数位置以后的小数.

例 1.8 舍入运算:(a) 1.933 舍入二位小数,(b) 0.01791 舍入到二位小数,(c) 1.23915 舍入到三位小数,(d) 0.0015 舍入到三位小数.

解 ☞ (a) 1.933 舍入到 1.93

(b) 0.01791 舍入到 0.02

(c) 1.23915 舍入到 1.239

(d) 0.0015 舍入到 0.002

1.5 绝对值

数 n 的绝对值就是这个数的数值部分,而不考虑符号.通常用 $|n|$ 表示.

例 1.9　计算下列数的绝对值:(a) -5, (b) $10/2$, (c) $\frac{43-52}{9}$

解 ☞　(a) $|-5|=5$

(b) $|10/2|=5$

(c) $\left|\frac{43-52}{9}\right|=1$

1.6　阶乘

符号 $n!$ (读做 n 阶乘)表示从 n 到 1 所有正整数的乘积

$$n!=n\times(n-1)\times(n-2)\times(n-3)\times\cdots\times1$$

例 1.10　计算下列阶乘:(a) $2!$, (b) $4!$, (c) $9!$.

解 ☞　(a) $2!=2\times1=2$

(b) $4!=4\times3\times(2!)=24$

(c) $9!=9\times8\times7\times6\times5\times(4!)=362,880$

1.7　开方和根

在表达式 $a=\sqrt[n]{b}$ 中,符号 $\sqrt{\ }$ 称为开方符号,$\sqrt[n]{b}$ 称为根式,a 称为 b 的 n 次方根,b 称为被开方数,n 称为根指数.

例 1.11　求下列根:$\sqrt[2]{4}$.

解 ☞　$\sqrt[2]{4}=\sqrt{4}=4$ 的二次根(或平方根). 4 的平方根是 $+2$,或者 -2. 但是按照惯例我们记 $\sqrt{4}=+2$, $-\sqrt{4}=-2$.

一个数的主 n 次方根是它的一个实数根,或者是正、负根中的正根.

例 1.12　求解主 n 次方根:(a) $\sqrt{16}$, (b) $-\sqrt{16}$.

解 ☞　(a) $\sqrt{16}=4$.

(b) $-\sqrt{16}=-4$.

1.8　平方根运算

两个平方根相乘等于被开方数相乘再开方. 两个平方根相除等于被开方数相除再开方.

例 1.13　计算:(a) $\sqrt{5}\sqrt{5}$, (b) $\frac{\sqrt{125}}{\sqrt{5}}$.

解 ☞　(a) $\sqrt{5}\sqrt{5}=\sqrt{5\times5}=\sqrt{25}=5$

(b) $\frac{\sqrt{125}}{\sqrt{5}}=\sqrt{\frac{125}{5}}=\sqrt{25}=5$

平方根与数相乘等于被开方数乘以此数的平方再开方. 平方根除以一数等于被开方数除以此数的平方再开方.

例 1.14　计算:(a) $2\sqrt{20.25}$, (b) $\frac{\sqrt{25}}{5}$.

解 ☞　(a) $2\sqrt{20.25}=\sqrt{2^2}\sqrt{20.25}=\sqrt{4(20.25)}=\sqrt{81}=9$

(b) $\frac{\sqrt{25}}{5}=\frac{\sqrt{25}}{\sqrt{25}}=\sqrt{\frac{25}{25}}=1$

1.9　幂运算

b^n 是 b 的 n **次幂**,它是 n 个 b 的乘积. 例如,$b^2=b\times b$ 是 b 的二次幂(b **的平方**),$b^3=b\times b\times b$ 是 b 的三次幂(b **的立方**). 在表达式 b^n 中,n 称为**指数**,b 称为**底数**.

例 1.15　计算下列幂运算：(a) b^0，(b) 12^0，(c) 12^2，(d) 5^5.

解 ☞　(a) 任何不等于零的实数的零次幂都等于 1：$b^0=1$

(b) $12^0=1$

(c) $12^2=12\times12=144$

(d) $5^5=5\times5\times5\times5\times5=3,125$

例 1.16　将下列表达式用分数、根式表示：(a) b^{-n}，(b) $b^{1/n}$，(c) $b^{-1/n}$.

解 ☞　(a) 对于任何不等于零的实数，如果它的指数为负数，那么有下列关系成立：$b^{-n}=\frac{1}{b^n}$

(b) $b^{1/n}=\sqrt[n]{b}$

(c) $b^{-1/n}=\frac{1}{\sqrt[n]{b}}$

任何数均可表示为 10 的整数次幂与数 a 的乘积，其中 a 的绝对值大于等于 1 且小于 10，习惯上称之为科学记数法.

例 1.17　用科学记数法表示下列各数：(a) 237，(b) 0.000237，(c) 116,270,000.

解 ☞　(a) $237=2.37\times10^2$

(b) $0.000237=2.37\times10^{-4}$

(c) $116,270,000=1.1627\times10^8$

同底数的幂相乘，指数相加；同底数的幂相除，指数相减.

例 1.18　计算：(a) 10×10^3，(b) $\frac{10^7}{10^3}$.

解 ☞　(a) $10\times10^3=10^{1+3}=10^4=10,000$

(b) $\frac{10^7}{10^3}=10^{7-3}=10^4=10,000$

1.10　对数运算

若 $n=c^b$，那么数 b 就称为以 c 为**底**的 n 的**对数**(其中 $c\neq1$ 且 $c>0$)，记作 $b=\log_c n$.

例 1.19　求(a) 若 $4=2^b$，计算 $\log_2 4$，(b) 若 $\log_{10} n=2$，计算 n.

解 ☞　(a) 若 $4=2^b$，那么 $b=2$，于是 $\log_2 4=2$

(b) 若 $\log_{10} n=2$，那么 $n=10^2=100$.

通常，若 $\log_c n=b$，则 b 的**反对数**是 n.

例 1.20　根据下列式子写出反对数：(a) $\log_2 16=4$，(b) $\log_{10}10=1$.

解 ☞　(a) 4 的反对数是 16

(b) 1 的反对数是 10

两个正数的积的对数，等于这两个数的对数的和. 两个正数的商的对数，等于被除数的对数减去除数的对数的差. 一个正数的幂的对数，等于幂的底数的对数乘以幂指数.

例 1.21　计算下列各数的对数：(a) bc，(b) b/c，(c) a^b.

解 ☞　(a) $\log(bc)=\log b+\log c$

(b) $\log(b/c)=\log b-\log c$

(c) $\log(a^b)=b(\log a)$

1.11　代数表达式

代数表达式是指用四种基本运算符号[(+)，(−)，(×)，(÷)]把**算术数**(有指定的数值)和**一般数**(表示数值的字母)连接而成的式子. **代数表达式的项**是指用+或−号分开的数、字母，或者用+或−号分开的它们的乘积或商. 只有一项的代数表达式称为**单项式**，有两项的代数表达式称为**二项式**，通常有两项或两项以上的代数表达式称为**多项式**.

例 1.22　指出下列代数表达式的项：$14a^2+10b-3c^4$.

解 ☞　该代数表达式的项为：$(14a^2)$，$(10b)$，$(3c^4)$.

1.12　方程和公式

方程是表示两个代数式相等关系的式子. 例如：$a-b=c$；$\frac{15}{5}=3$；$y+3=4$；$\log x+y=2$. 每一个方程都由**等号**(=)和等号两边的表达式组成. **公式**是用代数符号表示定理或规则的方程. 例如：$c=\pi d$[圆的周长 c=它的直径 $d\times$常数 $\pi(\pi=3.14159\cdots)$]. 方程两边若进行相同运算(被 0 除除外)则得到等价方程.

例 1.23　写出下列方程进行指定运算的等价方程：(a) $a+b=c$，方程两边同时加 b，(b) $a+b=c$，方程两边同时平方.

解 ☞　(a) $a+2b=c+b$

(b) $(a+b)^2=c^2$，或 $a^2+2ab+b^2=c^2$

恒等方程是指等号两边的值恒等的方程. 例如$(y-3)(y-1)=y^2-4y+3$ 和 $7+2=9$，都是恒等方程. 这类方程也称为**恒等式**，通常用符号≡代替=表示. 包含字母的恒等式，对字母取任何值均成立.

例 1.24　通过取 $y=5$，说明$(y-3)(y-1)=y^2-4y+3$ 是恒等式.

解 ☞

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

$$8=8$$

恒等式对于字母取任意值均成立，而**条件方程**只对某些数值是成立的.

例 1.25　找出使条件方程 $2y+4=10$ 成立的 y 值.

解 ☞　只有 $y=3$ 时，$2y+4=10$ 成立.

解方程是指找出使方程两边的值相等的一般数(字母)的值，这样的值称为**方程的解**. 对于恒等式$(y-3)(y-1)=y^2-4y+3$，y 的任意取值都是方程的解. 对于条件方程 $2y+4=10$，它只有一个**解**：$y=3$.

例 1.26　求解方程：$\frac{x}{10}+18=x$.

解 ☞

$$\frac{x}{10}+18=x$$

$$\frac{x}{10}-x=-18$$

$$x\left(\frac{1}{10}-1\right)=-18$$

$$x(-0.9)=-18$$

$$x=20$$

1.13　变量

在某一过程中可以取不同数值的一般数(见 1.11 节)称为**变量**. 因此，对于公式 $c=\pi d$(见 1.12 节)，如果问题要求计算直径分别为 1 in，2 in 和 3 in(1in(英寸)=2.540cm)的三个圆的周长，那么在这个问题中 c 和 d 都是变化的，因此它们都是变量.

在过程中保持同一数值的量称为**常量**. 常量分为**绝对常量**和**任意常量**两种. 绝对常量是指任何情况下都取相同值的量，例如算术数(例如 $5,\frac{1}{2},100$)或取固定值的一般数[例如 e(自然对数的底，见习题 1.23)，π(见 1.12 节)]. 任意常量是指在某个过程中保持不变的一般数，但过程不同其取值可能会改变.

本书到这为止都是采用小写字母来表示方程或公式中的一般数. 如果没有特别指出，**对于**

表示变量的一般数,将用字母表中后几个字母(Z,Y,X 等等)的大写表示,而对于表示常量的一般数,用字母表中前几个字母(a,b,c 等等)的小写来表示.

1.14　单变量方程和二次公式

单变量线性方程,也称为**一元一次方程**,只含有常量和一个变量,并且变量的次数为1.例如 $2X-3=2$ 和 $aY-b=cY$. **单变量二次方程**也称为**一元二次方程**,含有常量和一个变量,并且变量的最高次数为2.例如 $aX^2+bX=c$, $aX^2=c$ 和 $3X^2-5X=0$.如果方程中既有一次项又有二次项(例如 $aX^2+bX=c$),则称方程为**完全二次方程**;如果方程中只含有二次项(例如 $aX^2=c$),则称方程为**不完全二次方程**.对于完全二次方程 $aX^2+bX+c=0$,它的解是

$$X=\frac{-b\pm\sqrt{b^2-4ac}}{2a} \tag{1.1}$$

上式称为**二次求根公式**.对于任一二次项系数不为零的一元二次方程都可以用该公式求解(该公式的推导见习题1.31).

例 1.27　用二次求根公式求解方程:$2X^2=-3X+9$.

解 ☞　先将方程化为一般的二次形式:$2X^2+3X-9=0$,于是 $c=-9,b=3,a=2$

$$X=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

$$=\frac{-3\pm\sqrt{3^2-[4\times(-9)\times2]}}{2\times2}$$

$$=\frac{-3\pm\sqrt{9-(-72)}}{4}=\frac{-3\pm\sqrt{81}}{4}=\frac{-3\pm9}{4}$$

因此,

$$X=\frac{6}{4}=\frac{3}{2},\ X=\frac{-12}{4}=-3$$

1.15　统计中的变量

在统计中,变量是指所研究对象(人,物体,地点等)的特征,这些特征是可以测量的,并且对于一组对象中的不同对象可以取不同的值.例如,如果我们研究一群儿童,那么由于儿童的体重是可以测量的,并且对于不同的儿童它可以取不同的值,所以儿童的体重是一个变量.或者,如果我们研究一组马铃薯苗,那么我们可以考虑测量如下变量:马铃薯苗的高度、宽度、叶子的数量和马铃薯的个数.变量的测量问题将在第二章中讨论.

变量的每一个测量值称为**观测值**.在这本书中,用大写字母表示变量,而用同一字母的小写表示它的观测值(例如 X 表示变量,x 表示它的观测值).

观测值用数值表示的变量称为**定量变量**.定量变量的观测值可以根据特征的数量(更重,更高,更富有等等)进行排序.例如儿童的体重是一个定量变量,一个儿童体重为45.2磅(1磅=0.4536千克),另一个更重的儿童体重为50.9磅.类似的,马铃薯的个数也是一个定量变量.

如果变量的观测值只是一些不同的类别,不能按照其量值进行排序,则称该变量为**定性变量**.例如树的种类就是一个定性变量,其观测值有:松树、枫树、白杨树以及山胡桃等.

1.16　可观测变量,假设变量和测量变量

统计中的变量可以分为**直接测量变量**和**间接测量变量**.为理解这样的分类,考虑下面的例子.

遗传学家为了研究奥林匹克男子长跑冠军与短跑冠军之间的遗传差别,可以测量如下变量:身高、腿长、小腿和大腿的周长等.

在这个例子中,这些解剖学变量是直接测量变量,而遗传学变量(特定基因或基因组之间

的差别)是间接测量变量. 直接测量变量又称为**可观测变量**,间接测量变量又称为**假设变量**(或称为**介入性变量**).

测量变量是可观测变量,是被测量事物的直接可测量特性,可以用具体测量尺度上的值来表示. 测量变量可以是定性变量也可以是定量变量. 以英尺测量的高度、以克测量的重量、一棵植物上叶子的数量以及花的种类等都是测量变量.

1.17　函数和关系

在讨论函数和关系之前,我们先定义两个概念:**集合**与**实数系**.

集合是指一些事物(物体、符号、数等)的全体,集合中的每一项称为集合的**元素**(或**成员**).

实数系是指由**有理数**(所有整数的比$\frac{a}{b}$,其中$b\neq 0$)和**无理数**(不能写成两个整数的比的数,如$\sqrt{2}$,π)构成的集合.

考虑两个变量X,Y,如果对于X在某一范围内的每一个确定的值,Y都有唯一确定的值y与它对应,那么就称Y是X的**函数**. 函数通常有**定义域**(即X的取值范围)、**值域**(即Y的取值范围)和联系x值与y值的**对应规则**.

例 1.28　指出函数$Y=X^2$的定义域、值域和对应规则.

解 ☞　对于函数$Y=X^2$,X可以取任一实数,因此函数的定义域是实数系. 对每一个特定的x值,对应的y值是零或正数,因此函数的值域是非负实数集合. 函数的对应规则是$Y=X^2$. 图 1-1 通过一些具体的值举例说明了这些概念.

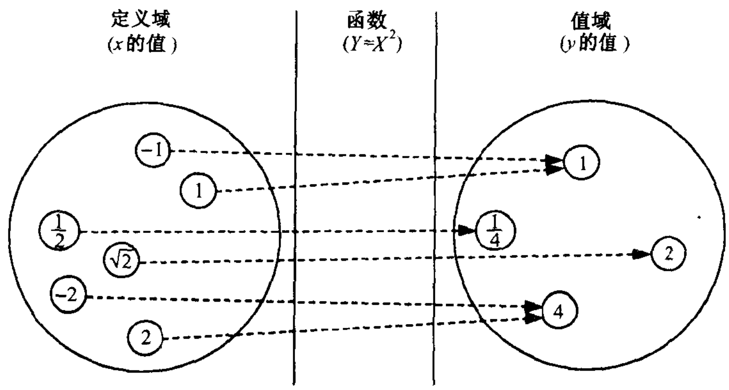

图 1-1

对于X在取值范围内的一个确定值,函数有唯一确定的对应值Y. 因此,可以说Y值"依赖"于X值,所以我们称X为**自变量**,Y为**因变量**.

关系与函数的区别在于对应规则的不同. 函数的对应规则是一对一,即一个X值对应一个且只能对应一个Y值;而关系的对应规则是一对多,即一个X值可以对应多个Y值. 所以也称函数为**单值函数**,而称关系为**多值函数**.

例 1.29　指出关系$Y=X\pm 3$的定义域、值域和对应规则.

解 ☞　对于关系$Y=X\pm 3$,定义域和值域均为实数系,其对应规则表明每一个x都有两个y值相对应. 图 1-2 举例说明了这一关系.

1.18　函数记号

函数可以写成变量间的关系式,如$Y=X^2$,或者写成以具体的变量值之间的联系. 因此,$Y=X^2$可以用函数记号表示如下

$$y=f(x)=x^2$$

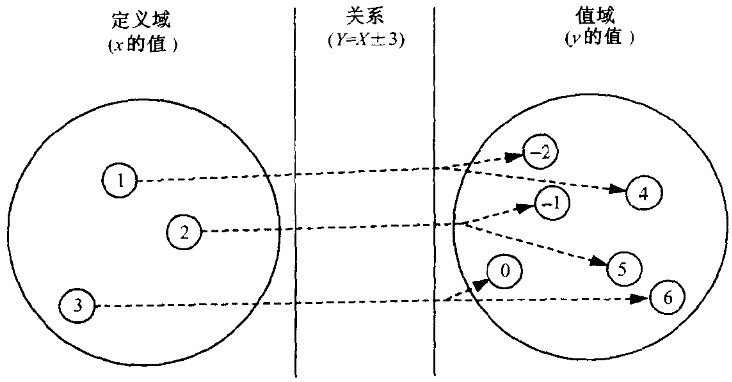

图 1-2

其中 $y=f(x)$读作 y 等于 x 的函数，而不能读成 y 等于 f 乘以 x. $f(x)$是一个一般的函数记号，我们也可以用 $F(z)$、$g(x)$、$h(y)$等其他记号来表示函数.

例 1.30　对函数 $y=f(x)=-3+2x+x^2$，计算(a) $f(0)$，(b) $f(1)$.

解 ☞　(a) $f(0)=-3+2(0)+0^2=-3+0+0=-3$

(b) $f(1)=-3+2(1)+1^2=-3+2+1=0$

1.19　统计中的函数

所有研究的根本目标都是探讨**因果关系**，找出**导致**事件发生的原因. 例如，植物学家想研究影响植物生长(果)的土壤的特性(因)，经济学家想确定影响汽车销售(果)的广告因素(因).

为了研究因果关系，研究人员使用统计技术确定自变量和因变量之间的函数关系. 在研究问题中，自变量是与"果"相联系的一个测量变量，因变量则是与"因"相联系的一个测量变量. 因变量的值在某种程度上是依赖于自变量的值. 因此，在上面的植物例子中，植物学家想说明植物高度(因变量)是土壤中氮含量(自变量)的函数；而在汽车销售例子中，经济学家想调查一个汽车公司过去每十年中销售的汽车数量(因变量)是否是该公司过去每十年中广告费用总数(自变量)的函数.

例 1.31　指出下列试验中的自变量和因变量

为了确定水的温度对鲑鱼生长的影响作用，我们饲养两组鲑鱼(每组 10 条)，一组水温控制在 20℃，另一组水温则控制在 24℃，其他条件都相同. 在孵化后 200 天，称一下两组中每一条鲑鱼的重量.

解 ☞　在试验中，通过自变量的变化来研究其对因变量的影响. 这个试验中是通过水温的变化来研究其对鲑鱼的体重变化的影响，所以水温是自变量而鲑鱼的体重是因变量.

统计中的函数只需要定义域、值域和两个变量间的对应关系，而并不需要写成等式表达式. 例如，一个函数的定义域是一组男生的名字，值域是他们头发的颜色，对应规则为对每一个名字有且仅有唯一的头发颜色相对应.

1.20　实数轴和直角笛卡儿坐标系

实数轴(也称为**数轴**或**实轴**)，是用来表示实数系中所有实数的一条直线. 实数系(见 1.17 节)中的每一个数都可用实数轴上的一个点来表示.

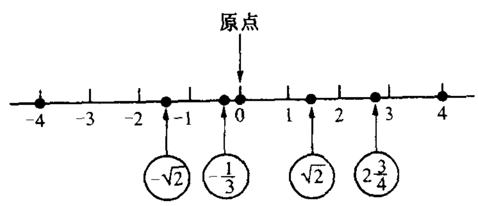

图 1-3

例 1.32　在实数轴上画出下列各数：-4，$-\sqrt{2}$，$-\dfrac{1}{3}$，0，$\sqrt{2}$，$2\dfrac{3}{4}$，4.

解 ☞　如图 1-3 所示.

直角笛卡儿坐标系(也称为**直角坐标系**)是由平面内有公共**原点**而且互相垂直的两条数轴构成的,如图 1-4 所示. 横轴,也称为 X **轴**,取向右方向为正方向. 纵轴,也称为 Y **轴**,取向上方向为正方向. X 轴和 Y 轴把平面(**XY 平面**)分成四个部分:**第一象限**、**第二象限**、**第三象限**和**第四象限**. 在平面内建立了直角坐标系后,对于平面内的任意一点,都有一对有序实数(x,y)和它对应. 第一个数,称为 x **坐标**或**横坐标**,是点到 Y 轴的水平距离. 第二个数,称为 y **坐标**或**纵坐标**,是点到 X 轴的垂直距离.

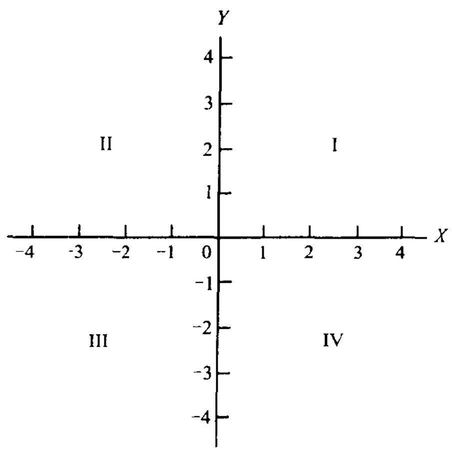

图 1-4

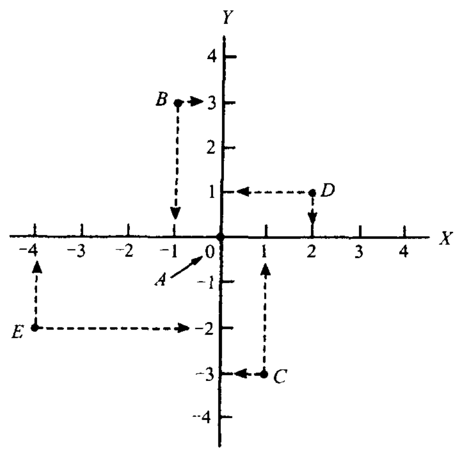

图 1-5

例 1.33　在直角坐标系中画出下列各点:$A(0,0)$;$B(-1,3)$;$C(1,-3)$;$D(2,1)$;$E(-4,-2)$.

解 ☞　这些点在直角坐标系中的位置如图 1-5 所示,其中虚线表示点到 X 轴和 Y 轴的距离.

1.21　函数的图象

函数的**图象**是函数中变量间关系的图形表示. 把自变量 x 的值和函数 y 的对应值分别作为点的横坐标和纵坐标,可以在直角坐标系内绘出函数的图象. 函数可以像方程(见 1.14 节)那样分类,因此,函数 $y=f(x)=c+bx$ 是一个线性或一次函数,它的图象是一条直线.

例 1.34　在直角坐标系内画出函数 $y=f(x)=4+2x$ 的图象.

解 ☞　该函数的图象如图 1-6 所示.

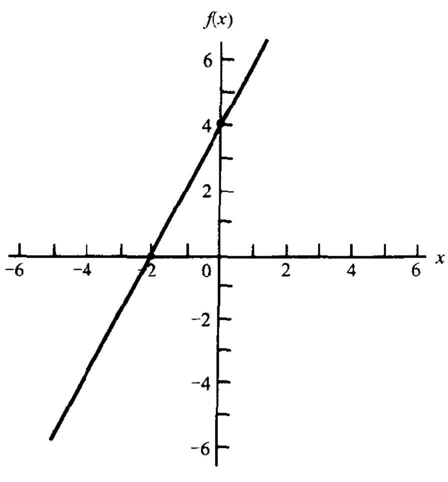

图 1-6

我们把形如 $y=f(x)=c+bx+ax^2$(其中 a,b,c 是常数,且 $a\neq 0$)的函数叫做**二次函数**. 一个二次函数的图象是一条**抛物线**,这条抛物线关于一条与 Y 轴平行的直线对称,这条直线称为**对称轴**. 抛物线如果开口向下,则有**最大值**;如果开口向上,则有**最小值**. 二次函数的图象的构造方法见习题 1.40.

1.22 序列、级数和求和符号

定义域为相继正整数的函数称为**序列**. 如果定义域包括所有正整数,此序列称为**无限序列**. 例如:$f(i)=i+1, i=1,2,3,\cdots,\infty$. 这个序列的第一个数是 $f(1)=1+1=2$;第二个数是 $f(2)=2+1=3$;第三个数是 $f(3)=3+1=4$;等等. 这个序列为:$2,3,4,\cdots,\infty$. 序列中的每一个数称为序列的项. 如果定义域只包括部分正整数,此序列称为**有限序列**. 例如:$f(i)=x_i, i=1,2,3$. 其中 x_i 中的 i 称为**下标**. 这个序列中只有三项:x_1, x_2, x_3.

例 1.35 写出序列的各项:$f(i)=i^2-3, i=2,3,4$.

解 ☞ 这个序列有三项:$f(2)=2^2-3=1$;$f(3)=3^2-3=6$;$f(4)=4^2-3=13$.

级数是序列的各项之和. 对于无限序列 $f(i)=i+1, i=1,2,3,\cdots,\infty$,其级数是

$$2+3+4+\cdots+\infty$$

对于有限序列 $f(i)=x_i, i=1,2,3$,其级数是

$$x_1+x_2+x_3$$

符号 $\sum\limits_{i=1}^{n} x_i$ 称为**求和符号**,它是级数 $x_1+x_2+x_3+\cdots+x_n$ 的记号. 符号 $\sum$ 是希腊字母 σ 的大写形式,表示对它的右边序列求和. $\sum$ 下面的字母称作**求和下标**,或**求和变量**,它的数值表示序列定义域的下限. $\sum$ 上面的数值表示序列定义域的上限. 对于 $\sum\limits_{i=1}^{n}$,其上下限表示序列从 $i=1$ 加到 $i=n$(从 x_1 到 x_n). 一般,求和下标用小写字母 i,j,k 表示.

例 1.36 求和:$\sum\limits_{i=1}^{4} i^2$.

解 ☞ $\sum\limits_{i=1}^{4} i^2=(1)^2+(2)^2+(3)^2+(4)^2=1+4+9-16=30$

如果 x_i 代表变量 X 的测量值,那么在统计学中 n 个测量值的全集是从 x_1 加到 x_n,这一级数记为 $\sum\limits_{i=1}^{n} x_i$. 通常我们省去上下限,直接记为 $\sum x_i$ 或 $\sum x$.

例 1.37 五个三年级的男孩的身高形成如下序列(1ft=0.3048 米):$x_1=2.1$ ft, $x_2=2.0$ ft, $x_3=1.9$ ft, $x_4=2.0$ ft, $x_5=1.8$ ft. 求和:$\sum x_i$.

解 ☞

$$\begin{aligned}\sum x_i=\sum_{i=1}^{n} x_i=\sum_{i=1}^{5} x_i&=x_1+x_2+x_3+x_4+x_5\\&=2.1\text{ ft}+2.0\text{ ft}+1.9\text{ ft}+2.0\text{ft}+1.8\text{ft}\\&=9.8\text{ ft}\end{aligned}$$

1.23 不等式

符号($<,>,\leq,\geq$)称为**不等号**. 用这些符号表示两个代数表达式**不相等关系**(小于,大于,小于等于,大于等于)的式子叫做**不等式**. 不等号的箭头总是指向较小的表达式.

例 1.38 解释下列数学表达式:(a) $3<4$, (b) $5>2$, (c) $b>a$, (d) $b\geq a$, (e) $b\leq a$.

解 ☞ (a) $3<4$ 是指 3 小于 4

(b) $5>2$ 是指 5 大于 2

(c) $b>a$ 是指 b 大于 a

(d) $b\geq a$ 是指 b 大于或等于 a