YINGYONG SHULI TONGJ I

应用数理统计

宇世航　张锐梅　王晓霞　编著

YINGYONG SHULI TONGJI

应用数理统计

世航　张锐梅　王晓霞　编著

黑龙江大学出版社
HEILONGJIANG UNIVERSITY PRESS

图书在版编目(CIP)数据

应用数理统计/宇世航,张锐梅,王晓霞编著. —哈尔滨:黑龙江大学出版社,2008.8
ISBN 978-7-81129-109-4

Ⅰ.应… Ⅱ.①宇…②张…③王… Ⅲ.数理统计-高等学校-教材 Ⅳ.O212

中国版本图书馆 CIP 数据核字(2008)第 132766 号

责任编辑 李兴华
封面设计 张 骏

应用数理统计
YINGYONG SHULI TONGJI
宇世航 张锐梅 王晓霞 编著

出版发行 黑龙江大学出版社
地 址 哈尔滨市南岗区学府路 74 号 邮编 150080
电 话 0451-86608666
经 销 新华书店
印 刷 哈尔滨海天印刷设计有限公司
版 次 2008 年 8 月 第 1 版
印 次 2008 年 8 月 第 1 次印刷
开 本 787×1092 毫米 1/16
印 张 11.75
字 数 230 千
书 号 ISBN 978-7-81129-109-4/O·2

定 价 25.00 元

凡购买黑龙江大学出版社图书,如有质量问题请与本社发行部联系调换

前 言

由于近代科学面向的多数是随机问题，因此“概率论”与“数理统计”系列课程作为在本科阶段研究随机现象的学科，他与许多新兴学科如人工智能、信息论、控制论等有着密不可分的联系，掌握这类课程并将其应用于解决实际问题，是新时代人才培养的需要。

1998 年教育部颁布新的本科专业目录中，数学学科包括“数学与应用数学”、“信息与计算科学”两个专业。1999 年在昆明召开的数学专业课程会议上通过的《数学与应用数学专业教学规范》中，已明确将“概率论”与“数理统计”分为两门课程，“概率论”列为七门专业基础课之一，而“数理统计”列为专业课。而《信息与计算科学专业教学规范》中将“数理统计”列为专业基础必修课。

按照“宽口径、厚基础”的办学指导方针，遵循因材施教原理，在新的人才培养计划中我们将原来的“概率论与数理统计”改为“概率论”和“数理统计”两门课程，分两个学期开设。“数理统计”是“概率论”的后续课程。随着专业培养目标和教学计划的不断完善，该课程的内容也在不断地探索、变化和充实。笔者在多年教学经验及参考使用兄弟院校的相关教材的基础上，逐渐形成了适应自己专业培养要求的内容。本书就是根据这些年共同的教学实践而形成的。

本书从对数据描述出发，介绍了数理统计的基本概念，然后叙述了数理统计的核心内容——统计推断（即参数估计和假设检验），最后讨论了回归分析、方差分析和正交试验设计等内容。在本教材的编写和教学中，我们力求做到以下几点：

1. 这是统计学的入门课程，所以立足于从数据出发提出问题和研究问题，注意问题的实际由来，强调统计基本思想的同时介绍实际做法。

2. 对基本概念的叙述上增加了描述性统计的基本内容，让学生能较为全面地认识统计。注意从偶然性中提炼出来的一些规律性的证明和论述，同时图文并茂，尽量使概念和定理成为有源之水、有

本之木。

3. 通过例子和练习帮助读者掌握原理和方法。例子更加贴近人们的经济生活和生产管理。每章配有典型习题。

根据课时需要，书中标有 * 号的章节可适当选修。

本书的编写及出版得到了齐齐哈尔大学有关领导的大力支持，黑龙江大学出版社同志的大力配合，在此表示衷心的感谢。另外我们在编写过程中还参考了国内外的许多著作和文章（见参考文献），很受教诲，在此向有关作者表示深深的感谢。

由于作者水平有限，书中难免不妥之处，请读者不吝指教。

作　者

2008 年 1 月

绪 论

数理统计是概率论的姐妹学科，都是研究随机现象统计规律的数学分支，大体上可以说：概率论是数理统计的基础，而数理统计是概率论的重要应用。概率论是从分布函数出发，先提出数学模型，然后研究它的特点、性质及规律。而数理统计是以试验观测数据为出发点，研究怎样收集带有随机误差的数据，根据试验资料为随机现象选择数学模型，并在此基础上对数据进行分析，以对所研究的对象作出尽可能正确的推断。

统计学自古有之，例如人口统计，社会调查等，但它不是现代意义下的数理统计学，只是数据的记录和整理。

数理统计学是随着概率论的发展而发展起来的。当人们认识到必须把数据看成来自具有一定概率分布的总体，所研究的对象是这个总体而不能局限于数据本身之日，也就是数理统计诞生之日。在19世纪中叶以前已有了若干重要的工作，特别是高斯(C. F. Gauss)和勒让德(A. M. Legendre)关于观测数据的误差分析和最小二乘法。1946年，克拉默(H. Cramer)发表的《统计学的数学方法》，标志着数理统计学的发展进入了成熟阶段。

数理统计是应用性很强的学科。事实上，它几乎渗透到人类活动的一切领域。把数理统计应用到不同的领域就形成了适用于特定领域的统计方法，如农业、生物和医学领域的“生物统计”，教育和心理学领域的“教育统计”，经济和商业领域的“计量经济”，金融领域的“保险统计”等等。而这些统计方法的共同基础是数理统计。

数理统计既然是以形形色色的数据为出发点，那么它所研究的主要内容可归纳为以下几方面：

1. 采集数据。如何设计试验收集数据，包括抽样技术、试验设计等。

2. 整理分析数据。怎样对这些数据进行浓缩加工，提炼信息，这是统计研究的基础。

3. 统计推断。这是数理统计的核心内容，主要包括统计估计和统计假设两种形式。这两种形式的推断渗透到数理统计学的每个分支。

目　录

绪论 …… 1
第1章　数理统计基本概念 …… 1
1.1　总体与个体 …… 1
1.1.1　总体与个体 …… 1
1.1.2　样本 …… 2
1.2　样本数据的整理与描述 …… 4
1.2.1　频数频率分布表 …… 4
1.2.2　样本数据的图形显示 …… 6
1.2.3　经验分布函数 …… 8
1.3　统计量及其分布 …… 9
1.3.1　统计量 …… 9
1.3.2　常用统计量 …… 9
1.3.3　正态总体的抽样分布 …… 12
1.3.4　χ^2、t、F 分布 …… 15
1.3.5　非正态总体的抽样分布 …… 18
1.4　次序统计量及其分布 …… 20
1.4.1　次序统计量(order statistics) …… 20
1.4.2　次序统计量的分布 …… 22
1.4.3　次序统计量的函数 …… 24
习题一 …… 26
第2章　参数估计 …… 29
2.1　矩估计法 …… 29
2.1.1　矩估计法(moment estimate)的原理 …… 29
2.1.2　矩估计法的做法 …… 30
2.2　极大似然估计 …… 32
2.2.1　极大似然估计(maximum－likelihood estimate)的原理 …… 32
2.2.2　极大似然估计的步骤 …… 33
2.3　点估计的优良性准则 …… 36
2.3.1　相合性(congruence) …… 37

2.3.2 无偏性(unbiasedness) …… 39
2.3.3 有效性(validity) …… 40
2.3.4 均方误差(mean square error) …… 41
2.4 最小方差无偏估计 …… 42
2.4.1 最小方差无偏估计 …… 42
2.4.2 Cramer-Rao 不等式 …… 43
2.5 区间估计 …… 45
2.5.1 区间估计的概念 …… 45
2.5.2 置信区间的构造方法 …… 46
2.5.3 正态总体参数的置信区间 …… 48
习题二 …… 53
第 3 章 假设检验(Ⅰ) …… 57
3.1 假设检验的基本思想和概念 …… 57
3.1.1 假设检验问题 …… 57
3.1.2 检验法则 …… 58
3.1.3 水平 α 的显著性检验 …… 59
3.2 正态总体参数假设检验 …… 60
3.2.1 单个正态总体参数假设检验 …… 60
3.2.2 两个正态总体参数假设检验 …… 64
3.3 其他分布参数的假设检验 …… 67
3.3.1 指数分布参数的假设检验 …… 67
3.3.2 两点分布参数的假设检验 …… 68
3.3.3 大样本检验 …… 70
3.4* 非独立样本参数的假设检验 …… 71
习题三 …… 75
第 4 章 假设检验(Ⅱ) …… 78
4.1 分布拟合检验 …… 78
4.1.1 概率图纸法 …… 79
4.1.2 χ^2 拟合检验法 …… 82
4.1.3 K -检验法 …… 84
4.2 两总体之间的假设检验 …… 87
4.2.1 独立性检验(test of independence) …… 87
4.2.2 秩和检验(rank sum test) …… 89
习题四 …… 93
第 5 章 回归分析 …… 95

5.1 回归分析基本概念 …… 95
5.2 一元线性回归 …… 97
5.2.1 一元线性回归统计模型 …… 97
5.2.2 回归系数 b_0, b_1 的估计 …… 98
5.2.3 随机误差 ε 的方差估计 …… 102
5.2.4 回归方程的显著性检验 …… 102
5.3 一元非线性回归 …… 106
习题五 …… 113
第6章 方差分析与正交试验设计 …… 116
6.1 方差分析简介 …… 116
6.2 单因子方差分析 …… 117
6.2.1 单因子方差分析的统计模型 …… 117
6.2.2 平方和分解 …… 119
6.2.3 检验方法 …… 120
6.2.4 参数估计 …… 123
6.3 双因子方差分析 …… 125
6.3.1 无交互作用的方差分析 …… 126
6.3.2 有交互作用的方差分析 …… 129
6.4 正交试验设计 …… 134
6.4.1 正交表 …… 134
6.4.2 正交试验方案的设计 …… 135
6.4.3 试验结果的分析 …… 139
习题六 …… 145
附表 …… 149
参考书目 …… 177

第1章　数理统计基本概念

1.1　总体与个体

1.1.1　总体与个体

总体是指与所研究问题有关的对象的全体所构成的集合，而组成总体的每个元素就是个体.

例1.1　某工厂生产大批的电子元件，研究电子元件的寿命情况，那么这一大批元件就是问题的总体，而每一个元件就是一个个体.

例1.2　要研究某大学学生的学习情况，则该校的全体学生构成问题的总体，每一个学生则是该总体中的一个个体.

总体随所研究的范围而定. 如在上例中，若你研究全国大学生的学习成绩，则总体就大多了，它包含全国所有在校的大学生. 总体如何确定，取决于研究目的，也受人力、物力、时间等因素的限制.

对于大多数实际问题，总体中的个体是一些实在的人或物，而问题中所注意的，并不在于这些人或物本身，而在于所关心的某种指标，例如一个学生有身高、体重、姓氏笔划、籍贯出身等特征，当我们研究学生学习成绩时，对这些都不关心，而只注意其考分如何. 在例1.1中，我们只注意元件的寿命如何. 这样，也可以把我们感兴趣的那个指标值作为该个体（例如，大学生 A 得90分，即以90这个数代替 A），而总体就由一些数所组成.

单是这样还不行，这里有两个问题：一是总体中这样一大堆杂乱无章的数没有赋予什么数学或概率的性质，因而无法使用有力的概率论工具去研究它；二是各种总体变得没有区别，例如，大学生的学习成绩也是一堆数，一大批元件的寿命也是一堆数，大家都一样了. 解决这些问题的途径，就涉及总体这个概念的核心——总体的概率分布. 例如，在例1.1中元件寿命分布一般为指数分布，例1.2学生的学习成绩可以假定服从正态分布. 总体分布不同，分析的方法也不同. 赋有一定概率

分布的总体称为**统计总体**.

因此,经过以上几步的分析,我们就得出在数理统计学中"总体"这个基本概念的要旨——**总体就是一个概率分布**. 当总体分布为指数分布时,称为指数分布总体;当总体分布为正态分布时,称为正态分布总体,或简称正态总体,等等. 总体中的每一个个体是随机试验的一个观察值,因此它是某一随机变量 ξ 的值,这样,一个总体对应于一个随机变量 ξ. 我们对总体的研究就是对一个随机变量 ξ 的研究,ξ 的分布函数和数字特征就称为总体的分布函数和数字特征. 今后将不区分总体与相应的随机变量,笼统称为总体 ξ. 两个总体,即使其所含个体的性质根本不同,只要有同一个概率分布,则在数理统计学上就视为同类总体. 例如人的寿命也可以服从指数分布,它与元件寿命的分布一样,处理二者的统计问题的方法也一样,即可视为同一类总体.

例 1.3 考察某厂的产品质量,将其产品分为合格品和不合格品,并以 0 记合格品,以 1 记不合格品,则总体 = {该厂生产的全部合格品和不合格品} = {由 0 或 1 组成的一堆数}.

若以 p 表示这堆数中 1 的比例(不合格品率),则该总体 ξ 服从两点分布:

ξ	0	1
p_i	$1-p$	p

不同的 p 反映总体间的差异. 实际中,分布中的不合格品率是未知的,如何对之进行估计是统计学要研究的问题.

总体还有有限总体和无限总体,这里将以无限总体作为主要研究对象. 总体也常被称做母体.

1.1.2 样本

在实际中,总体的分布一般是未知的,或只知道它具有某种形式,而其中包含着未知参数. 在数理统计中,人们都是通过从总体中抽取一部分个体,根据获得的数据来对总体分布得出推断的. 被抽出的部分个体叫做总体的一个样本.

所谓从总体中取一个个体,就是对总体 ξ 进行一次观察并记录其结果. 我们在相同的条件下对总体 ξ 进行 n 次重复的、独立的观察. 将 n 次观察结果按试验的次序记为 $\xi_1,\xi_2,\cdots,\xi_n$. 由于 $\xi_1,\xi_2,\cdots,\xi_n$ 是对随机变量 ξ 观察的结果,且各次观察是在相同的条件下独立进行的,所以有理由认为 $\xi_1,\xi_2,\cdots,\xi_n$ 是相互独立的,且都是

与 ξ 具有相同分布的随机变量. 这样得到的 $\xi_1,\xi_2,\cdots,\xi_n$ 称为来自总体 ξ 的一个简单随机样本，n 称为这个样本的容量. 以后如无另外说明，所提到的样本都是指简单随机样本.

当 n 次观察一经完成，我们就得到一组实数 $x_1,x_2,\cdots,x_n$，它们依次是随机变量 $\xi_1,\xi_2,\cdots\xi_n$ 的观察值，称为样本值.

对于有限总体，采用放回抽样就能得到简单随机样本，但放回抽样使用起来不方便，当个体的总数 N 比要得到的样本的容量 n 大得多时，在实际中可将不放回抽样近似地当做放回抽样来处理.

至于无限总体，因抽取一个个体不影响它的分布，所以总是用不放回抽样. 例如，在生产过程中，每隔一定时间抽取一个个体，抽取 n 个就得到一个简单随机样本，实验室中的记录，水文、气象等观察资料都是样本. 试制新产品得到的样品的质量指标，也常被认为是样本.

综合上述，我们给出以下定义.

定义　设 ξ 是具有分布函数 F 的随机变量，若 $\xi_1,\xi_2,\cdots,\xi_n$ 是具有同一分布函数 F 的、相互独立的随机变量，则称 $\xi_1,\xi_2,\cdots,\xi_n$ 为从分布函数 F（或总体 F、或总体 ξ）得到的**容量为 n 的简单随机样本**，简称**样本**（**sample**），它们的观察值 $x_1,x_2,\cdots,x_n$ 称为**样本值**，又称为 ξ 的 n **个独立的观察值**.

也可以将样本看成是一个随机向量，写成 $(\xi_1,\xi_2,\cdots,\xi_n)$，此时样本值相应地写成 $(x_1,x_2,\cdots,x_n)$. 若 $(x_1,x_2,\cdots,x_n)$ 与 $(y_1,y_2,\cdots,y_n)$ 都是相应于样本 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的样本值，一般来说它们是不相同的.

例 1.4　啤酒厂生产的瓶装啤酒规定净含量为 640 g，由于随机性，事实上不可能使所有的啤酒净含量均为 640 g. 现从某厂生产的啤酒中随机抽取 10 瓶测定其净含量，结果如下：

641　635　640　637　642　638　645　643　639　640

这是一个容量为 10 的样本的观测值，对应的总体为该厂生产的瓶装啤酒的净含量.

设总体 ξ 具有分布函数 $F(x)$，$\xi_1,\xi_2,\cdots,\xi_n$ 为取自这一总体的容量为 n 的样本，则样本的联合分布函数为

$$F(x_1,x_2,\cdots,x_n) = P(\xi_1 < x_1,\xi_2 < x_2,\cdots,\xi_n < x_n) = \prod_{i=1}^{n} F(x_i).$$

联合分布密度

$$p(x_1,x_2,\cdots,x_n) = \prod_{i=1}^{n} p(x_i).$$

联合分布列

$$P(\xi_1 = x_1,\xi_2 = x_2,\cdots,\xi_n = x_n) = \prod_{i=1}^{n} P(\xi_i = x_i).$$

续例 1.3 设这批产品共 N 件,从中作不放回抽取 n 件,则

$$P(\xi_2 = 1 \mid \xi_1 = 1) = \frac{Np-1}{N-1},$$

$$P(\xi_2 = 1 \mid \xi_1 = 0) = \frac{Np}{N-1}.$$

显然如此得到的样本不是简单随机样本. 但是当 N 很大时,上述两种情形的概率都近似等于 p,所以当 N 很大,而 n 不大时,可以把该样本近似看成简单随机样本. 其联合分布列为

$$P(\xi_1 = x_1,\xi_2 = x_2,\cdots,\xi_n = x_n) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i},$$

其中 $x_i = 1,0;i = 1,2,\cdots,n.$

在数理统计中,总体或者说总体分布是我们研究的目标,而样本是从总体中随机抽取的一部分个体. 通过对这些个体进行具体研究,我们得到的统计结论,都反映或体现着总体的信息. 因此我们总是着眼于总体,而着手于样本,用样本去推断总体.

1.2 样本数据的整理与描述

1.2.1 频数频率分布表

样本数据的整理是统计研究的基础,整理数据的最常用方法之一是给出其频数分布表或频率分布表.

例 1.5 为研究某厂工人生产某种产品的能力,我们随机调查了 20 位工人某天生产的该产品的数量,数据如下

160	196	164	148	170
175	178	166	181	162
161	168	166	162	172
156	170	157	162	154

对这20个数据(样本值)进行整理,具体步骤如下:

(1)对样本值进行分组. 首先确定组数 k,作为一般性原则,组数通常在5~20个,对容量较小的样本,通常将样本值分为5组或6组,容量在100左右的样本可分为7~10组,容量在200左右的样本可分为9~13组,容量在300左右的样本可分为12~20组. 目的是使用足够的组来表示数据的变异. 本例是20个数据,我们将之分为5组,即 $k=5$.

(2)确定每组组距. 每组区间长度可以相同也可以不同,实用中常选用长度相同的区间以便进行比较,此时各组区间的长度称为**组距**,其近似公式为:

组距 = (样本最大观测值 − 样本最小观测值)/ 组数

$$d = \frac{196 - 148}{5} = 9.6,$$

为方便起见,取组距为10.

(3)确定每组组限. 各组区间端点为 $a_0, a_1 = a_0 + d, a_2 = a_0 + 2d, \cdots, a_k = a_0 + kd$,形成如下的分组区间

$$(a_0, a_1], (a_1, a_2], \cdots, (a_{k-1}, a_k],$$

其中 a_0 略小于最小观测值,a_k 略大于最大观测值,本例中可取 $a_0 = 147, a_5 = 197$,于是本例的分组区间为:

$$(147,157], (157,167], (167,177], (177,187], (187,197].$$

通常可用每组的组中值来代替该组的变量取值,

组中值 = (组下限 + 组上限)/2.

(4)统计样本数据落入每个区间的个数——频数,并列出其频数、频率分布表. 本例的频数、频率分布表见表1-1,从表中可以读出许多信息,如:40%的工人产量在157到167之间;产量少于167个的有12人,占60%;产量高于177的有3人,占15%.

表1-1 例1.5的频数频率分布表

组序	分组区间	组中值	频数	频率	累计频率%
1	(147,157]	152	4	0.20	20
2	(157,167]	162	8	0.40	60
3	(167,177]	172	5	0.25	85
4	(177,187]	182	2	0.10	95
5	(187,197]	192	1	0.05	100
合计			20	1	

1.2.2 样本数据的图形显示

对于样本数据的整理,除了用列表记录频数、频率分布情况外,还可以用图形表示,这在许多场合更直观.

1. 直方图

频数分布最常用的图形是直方图,它在组距相等场合常用宽度相等的长条矩形表示,矩形的高低表示频数的大小. 在图形上,横坐标表示所关心变量的取值区间,纵坐标表示频数,这样就得到频数直方图,若把纵坐标改成频率就得到频率直方图.

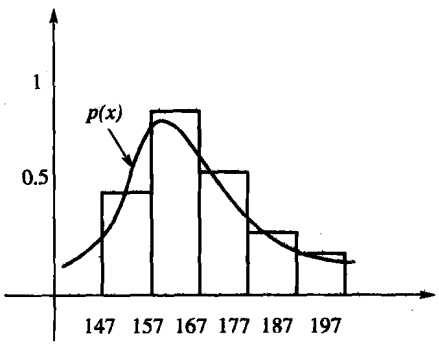

图 1-1 例 1.5 的单位频率直方图

为使诸长条矩形面积和为 1,可将纵轴取为频率/组距,如此得到的直方图称为单位频率直方图. 凡此三种直方图的差别仅在于纵坐标刻度的选择,直方图本身并无变化.

如果 n 愈大,分组时各组组距愈小(当然组数增多),直方图越接近总体的分布密度 $p(x)$ 的图形.

2. 茎叶图

除直方图外,另一种常用的方法是**茎叶图**,即将一个数值分为两部分,前面一部分(百位和十位)称为**茎**,后面部分(个位)称为**叶**,如

数值		分开		茎	和	叶
112	→	11|2	→	11	和	2

然后画一条竖线,在竖线的左侧写上茎,右侧写上叶,就形成了茎叶图.

例 1.6 某公司对应聘人员进行能力测试,测试成绩总分为 150 分,下面是 50 位应聘人员测试成绩(已经排过序):

64	67	70	72	74	76	76	79	80	81
82	82	83	85	86	88	91	91	92	93
93	93	95	95	95	97	97	99	100	100
102	104	106	106	107	108	108	112	112	114
116	118	119	119	122	123	125	126	128	133

应聘人员测试成绩的茎叶图见图1－2.

茎叶图的外观很像横放的直方图，但茎叶图中叶增加了具体的数值，使我们对数据具体取值一目了然，从而保留了数据中全部的信息.

在要比较两组样本时，可画出他们**背靠背的茎叶图**，这是一个简单直观而有效的对比方法.

6	47
7	024669
8	01223568
9	112333555779
10	002466788
11	2246899
12	23568
13	3

图1－2　测试成绩茎叶图

例1.7　下面的数据是某厂两个车间某天各40名员工生产的产品数量(表1－2)

表1－2　某厂两车间各40名员工的产量

甲车间						乙车间					
50	52	56	61	61	62	56	66	67	67	68	68
64	65	65	65	67	67	72	72	74	75	75	75
67	68	71	72	74	74	75	76	76	76	76	78
76	76	77	77	78	82	78	79	80	81	81	83
83	85	87	88	90	91	83	83	84	84	84	86
86	92	86	93	93	97	86	87	87	88	92	92
100	100	103	105			93	95	98	107		

为对其进行比较，我们将这些数据放到一个背靠背的茎叶图上(图1－3).

甲车间			乙车间
	620	5	6
	87775554211	6	67788
	877664421	7	22455556666889
	8766532	8	01133344466778
	733210	9	22358
	5300	10	7

图1－3　两车间产量背靠背的茎叶图

在茎叶图1－3中，茎在中间，左边表示甲车间的数据，右边表示乙车间的数据. 从茎叶图可以看出，甲车间员工的产量偏于上方，而乙车间员工的产量大多位于中间，乙车间的平均产量要高于甲车间，乙车间各员工的产量比较集中，而甲车间员工的产量则比较分散.

1.2.3 经验分布函数

设 $\xi_1,\xi_2,\cdots,\xi_n$ 是取自总体分布函数为 $F(x)$ 的样本，$x_1,x_2,\cdots,x_n$ 是它的一组观测值，若将样本观测值由小到大顺序排列为 $x_{(1)},x_{(2)},\cdots,x_{(n)}$，则 $x_{(1)},x_{(2)},\cdots,x_{(n)}$ 称有序样本，用有序样本定义如下函数：

$$F_n(x)=\frac{N_n(x)}{n},(-\infty<x<+\infty),$$

其中 $N_n(x)$ 为 $x_1,x_2,\cdots,x_n$ 中小于 x 的个数. $F_n(x)$ 有如下形式：

$$F_n(x)=\begin{cases}0, & x\leqslant x_{(1)},\\ k/n, & x_{(k)}<x\leqslant x_{(k+1)},(k=1,2,\cdots,n-1)\\ 1, & x>x_{(n)},\end{cases}$$

则 $F_n(x)$ 是单调不减左连续函数，且满足 $F_n(-\infty)=0,F_n(+\infty)=1$，由此可见 $F_n(x)$ 是一个分布函数，并称 $F_n(x)$ 为**经验分布函数**.

其中，
$$I(x_i<x)=\begin{cases}1, & x_i<x,\\ 0, & x_i\geqslant x\end{cases}$$

称**示性函数**.

例 1.8 从织布车间取 7 匹布，检查每匹上疵点数，得样本观测值 0,3,2,1,1,0,1. 则样本经验分布函数为

$$F_n(x)=\begin{cases}0, & x\leqslant 0,\\ 2/7, & 0<x\leqslant 1,\\ 5/7, & 1<x\leqslant 2,\\ 6/7, & 2<x\leqslant 3,\\ 1, & x>3.\end{cases}$$

其图形如图 1-4 所示.

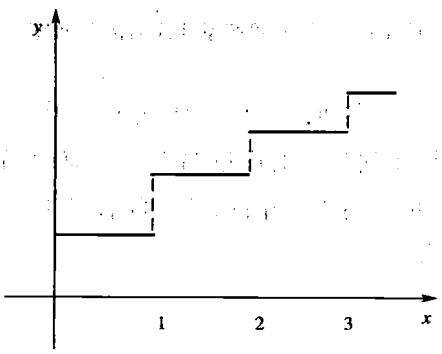

图 1-4 例 1.8 样本经验分布函数