21世纪课程教材 Textbook Series for 21st Century

# 概率论与数理统计（下）

The Theory of Probability and Statistics

3

刘海燕 编　　赵联文 审

国防工业出版社

面 向 21 世 纪 课 程 教 材

Textbook Series for 21st Century

# 概率论与数理统计

## (下 册)

刘海燕 编
赵联文 审

国防工业出版社
·北京·

**图书在版编目(CIP)数据**

概率论与数理统计/李裕奇,刘海燕编.—北京:国防工业出版社,2001.8
ISBN 7-118-02611-5

Ⅰ.概... Ⅱ.①李... ②刘... Ⅲ.①概率论②数理统计 Ⅳ.021

中国版本图书馆 CIP 数据核字(2001)第 043599 号

国防工业出版社出版发行
(北京市海淀区紫竹院南路 23 号)
(邮政编码 100044)
三河腾飞胶印厂印刷
新华书店经售
*
开本 787×960 1/16 印张 28¾ 511 千字
2001 年 8 月第 1 版 2001 年 8 月北京第 1 次印刷
印数:1-8000 册 总定价:42.00 元 上册:24.00 元 下册:18.00 元

# 前　言

本书是依据高等学校工科数学教材编审委员会审定的《概率论与数理统计教学大纲》，及编者多年来从事概率论与数理统计课程的教学经验，专为高等院校学生编写的教材。为方便读者使用，全书分为上、下册，上册(前六章)主要介绍概率论的基础理论知识及简单应用，由李裕奇负责编写；下册(后五章)主要介绍数理统计的基础知识与应用，以及部分多元统计知识介绍，由刘海燕负责编写。

概率论与数理统计是一门专门研究和探索客观世界中随机现象统计规律性的科学，它在自然科学与社会科学的许多领域中得到广泛的应用；它在金融、保险、经济与企业管理、工农业生产、军事、医学、地质学、空间技术、气象与灾害预报，以及许多新兴学科与边缘学科，如信息论、排对论，决策论、生物统计、统计物理、人工智能、控制论、金融数学及其它工程技术的发展中都作出了非常重要的贡献。反过来，这些领域中提出的许多新的课题进一步促进了概率论与数理统计的发展，因而概率论与数理统计是近代数学最活跃的分支之一。鉴于概率论与数理统计的重要性，目前研究生入学数学统考试题中，概率论与数理统计所占比例已达到20%～25%。各高等院校的理、工、医、农、金融、管理与许多文科专业都设置了概率论与数理统计课程，因此，学习与掌握概率论与数理统计的基本理论与应用，不仅是将来从事科学研究与工程实际工作的需要，也是继续学习现代科学技术与个人深造的需要，也是高度发展的现代科学技术对现代化人才提出的要求。

本书系统地阐述了概率论与数理统计的基本概念、基本思想、基本原理与基本方法，注重理论联系实际，突出解题思路，详尽介绍各种概率模型的掌握与应用，介绍多种解题方法与方法之间的联系，并使解题方法条理化，使读者便于学习与记忆，并受到很好的基本训练。在每一章末都注明了对学生的基本学习要求，配备了大量的思考题，基本练习、综合练习与自测题，使学生通过学习容易了解自己学习的成绩。章节安排循序渐进，层次分明，前后呼应，便于教学，方便于读者更好、更快地学习掌握及运用概率论与数理统计知识。

概率论与数理统计讲授各需 34 学时左右，读者只需具备工科高等数学的基本知识，就可顺利阅读全书。

本教材是西南交通大学应用数学系概率统计教研室与西南交通大学统计咨询中心全体同仁通力合作的结晶，是工科数学概率论与数理统计教材改革的成果之一。该教材在编写与出版过程中，得到西南交通大学应用数学系与概率统计教研室、教务处、教材科与国防工业出版社的鼎力支持与帮助，编者谨此一并表示由衷的感谢。

由于编者水平有限、书中难免会有错误与不妥之处，恳请同行及读者批评指正。

编　者

2001年3月于成都

# 内 容 简 介

本书是专为高等院校学生学习概率论与数理统计课程编写的教材，也可作为有关专业的参考书与从事概率论与数理统计相关工作的科研与工程技术人员的参考书。

本书分为上、下册，共十一章，上册包括概率论的基本概念；随机变量及其分布；多维随机变量及其分布；随机变量的数字特征；大数定律与中心极限定说及概率论的简单应用等知识；下册包括数理统计的基本概念；样本分布；参数估计；假设检验；线性统计推断以及最常用的多元统计方法。

本书每章节末都配有大量的思考题、基本练习，综合练习与自测题，帮助读者循序渐进地牢固地掌握概率论与数理统计知识。

# 目　　录

## （下册）

# 第七章　样 本 分 布

## §7.1　总体与样本

为了更好地介绍数理统计所研究问题,先引入一些常用的概念和术语。

### 一、总体与样本

**定义 7.1.1**　研究对象的全部元素组成的集合,称为总体,组成总体的每一个元素称为个体。

**例 7.1.1**　考察某地区全体居民的身高情况,则该地区所有人的身高便构成一个总体,而每个人的身高就是一个个体。

**例 7.1.2**　检查一批电视机的质量,则该批电视机的全体就可看成一个总体,其中一个电视机就是一个个体。

首先总体可以是有限的,也可以是无限的,其个数称为总体容量,通常用 $n$ 表示。

在研究总体性质时,我们关心的往往是总体的一个或几个数量指标,如检查电视机质量时往往是检查“电视机的寿命”。考察人时往往是身高,体重等。这些数量指标是通过每一个个体共同表现出来的,而总体中的每个个体不尽相同,因此,刻划总体的数量指标 $X$ 可看作一个随机变量。为了研究方便,我们通常把随机变量 $X$ 取值的全体当作总体,称为总体 $X$。而把每一个体的数值看成是 $X$ 的一个观测值。这样我们可以借助随机变量来研究总体,或者说总体就是一个随机变量。如果要研究的数量指标不止一个,那么可分为几个总体来研究。

对全部个体即总体的检验是困难的或者是不可能做到的,一个可能达到的,最恰当的替代办法是在这一总体中检验一小部分。利用这部分个体提供的信息,对总体特征作出某种程度的推断。例如检验一批电视机的寿命,不可能把所有电视机都用来做试验。但我们可以从这批电视机中随机抽取 $n$ 个,获得总体 $X$ 的 $n$ 个观察值 $x_1,x_2,\cdots,x_n$,然后根据这组观察值对整体电视机的寿命(总体 $X$)作出分析和判断。

从总体中随机抽取的 $n$ 个个体称为容量为 $n$ 的子样。我们要从子样的观察

或试验结果的特性来对总体的特性作出估计与推断，一方面自然要研究应该怎样从总体中抽取子样，使得子样尽可能大的程度上能反映总体的特性；同时必须建立一整套的方法，使能根据所选取的子样的性质，来对总体的特性进行估计与推断。因此，我们在抽取子样来对总体作出估计与推断时，从总体中抽取子样必须是随机的，即每一个个体都有同等概率被抽取（当总体中的个体是有限时，要用有返回抽取方式）。其具体要求为两个方面：独立性，是指 $X_1,X_2,\cdots,X_n$ 相互独立；代表性，是指 $X_1,X_2,\cdots,X_n$ 中每一个都与总体 $X$ 有相同分布。

**定义 7.1.2** （简单随机子样）设 $X_1,X_2,\cdots,X_n$ 为来自总体 $X$ 的容量为 $n$ 的子样，如果 $X_1,X_2,\cdots,X_n$ 相互独立，且每一个都与总体 $X$ 有相同的随机变量，则称 $X_1,X_2,\cdots,X_n$ 为总体 $X$ 的简单随机子样，简称为简单子样或子样。

在以后的章节中，我们所讲的子样 $X_1,X_2,\cdots,X_n$，无特别声明的话，都是指简单子样，即 $X_1,X_2,\cdots,X_n$ 相互独立，且每个都与总体 $X$ 有相同的分布。引入简单随机子样，是基于要从子样所获得的统计特性来对总体的统计特性作出估计与推断。

## 二、统计量

子样 $X_1,X_2,\cdots,X_n$ 也可看作为 $n$ 维随机向量 $(X_1,X_2,\cdots,X_n)$。记 $x_i$ 为 $X_i$ 一次观察值，并称 $(x_1,x_2,\cdots,x_n)$ 为子样的一次观察值。

**定义 7.1.3** 设 $X_1,X_2,\cdots,X_n$ 为总体 $X$ 的样本，作子样的函数 $T=T(X_1,X_2,\cdots,X_n)$ 且不含任何未知参数，则 $T(X_1,X_2,\cdots,X_n)$ 称为统计量。

若 $(x_1,x_2,\cdots,x_n)$ 为子样的一次观察值，则 $T$ 的取值为 $t=T(x_1,x_2,\cdots,x_n)$。

我们在用子样 $X_1,X_2,\cdots,X_n$ 获得的信息来对总体 $X$ 作出估计与推断时，按不同的要求而规定子样的各种函数。

**例 7.1.3** 设 $X_1,X_2,\cdots,X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的容量为 $n$ 的样本，其中 $\mu$ 已知，$\sigma$ 未知，则 $\sum\limits_{i=1}^{n}(X_i-\mu)^2$，$X_1+X_2$，都是统计量；而 $\frac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$ 不是统计量，因为它包含未知参数 $\sigma$。

**例 7.1.4** 设 $X_1,X_2,\cdots,X_n$ 为总体 $X$ 的容量为 $n$ 的子样，记作

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

则 $\overline{X}$ 及 $S^2$ 都是统计量，称 $\overline{X}$ 及 $S^2$ 分别为子样 $X_1,X_2,\cdots,X_n$ 的平均值及方差。子样的观察值为 $x_1,x_2,\cdots,x_n$，$\overline{X}$ 及 $S^2$ 的观察值分别记作

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

今后,大写的 $S^2$ 表示统计量,小写的 $s^2$ 表示统计量 $S^2$ 的观察值。

**定义7.1.4(顺序统计量)**　设 $X_1, X_2, \cdots, X_n$ 为总体 $X$ 的子样,今由子样建立 $n$ 个函数:

$$X_k^* = X_k^*(X_1, X_2, \cdots, X_n) \quad k = 1, 2, \cdots, n$$

其中 $X_k^*$ 为这样的统计量,它的观察值为 $x_k^*$,$x_k^*$ 为子样 $X_1, X_2, \cdots, X_n$ 的观察值 $x_1, x_2, \cdots, x_n$ 中按大小由小至大排列:

$$x_1^* \leqslant x_2^* \leqslant \cdots \leqslant x_k^* \leqslant \cdots \leqslant x_n^*$$

后的第 $k$ 个数值,称 $X_1^*, X_2^*, \cdots, X_n^*$ 为顺序统计量。

易见,$X_i^* = \min\{X_1, X_2, \cdots, X_n\}$,$X_n^* = \max\{X_1, X_2, \cdots, X_n\}$。若 $n$ 为奇数称 $X_{\frac{n+1}{2}}^*$ 为子样的中值;若 $n$ 为偶数,则称 $X_{\frac{n}{2}+1}^*$ 为子样的中值。

**定义7.1.5(极差)**　设 $X_1, X_2, \cdots, X_n$ 为总体 $X$ 的子样,则称统计量 $D_n^* = X_n^* - X_1^*$ 为子样的极差。它是子样中最大值与最小值之差,反映了子样观察值的波动幅度。它同方差一样是反映观察值离散程度的数量指标,而且计算最方便。

**例7.1.5**　设 $X_1, X_2, \cdots, X_5$ 为 $X$ 的容量为 5 的子样,今对这个子样作了三次观察,其值如下表所列,试求 $\overline{X}$,$S^2$ 及 $D_n^*$ 的观察值。

| $X$ \ $x$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| 1 | 3 | 1 | 10 | 5 | 6 |
| 2 | 2 | 6 | 7 | 2 | 8 |
| 3 | 8 | 3 | 9 | 10 | 5 |

解:得

| $X_1^*$ | $X_2^*$ | $X_3^*$ | $X_4^*$ | $X_5^*$ | $\overline{x}$ | $s^2$ | $D_n^*$ |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 5 | 6 | 10 | 5 | 11.5 | 9 |
| 2 | 2 | 6 | 7 | 8 | 5 | 8 | 6 |
| 3 | 5 | 8 | 9 | 10 | 7 | 8.5 | 7 |

## 三、样本平均数和样本方差的简算公式

当样本值很复杂时,样本平均数和样本方差的计算常采用以下简算公式。

设 $x_1, x_2, \cdots, x_n$ 为样本的 $n$ 个观察值。

(1) 选择适当常数 $a$,令 $y_i = x_i - a\ (i = 1, 2, \cdots, n)$ 获得一组较简单数据 $y_1, y_2, \cdots, y_n$。

由 $$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{n}\sum_{i=1}^{n}(y_i+a)=\frac{1}{n}\sum_{i=1}^{n}y_i+a$$

则 $$\bar{x}=\bar{y}+a$$

$$S_x^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n-1}\sum_{i=1}^{n}[(y_i+a)-(\bar{y}+a)]^2=$$
$$\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

即 $$S_x^2=S_y^2$$

**例 7.1.6** 取某型号火箭 8 枚进行射程试验,测得数据如下(单位:km):

54, 52, 49, 57, 43, 47, 50, 51

试计算样本平均数和样本方差。

**解:**取 $a=50$,由 $y_i=x_i-50$ 把样本值变为

$y_i$:4, 2, −1, 7, −7, −3, 0, 1,

则 $$\bar{y}=\frac{1}{8}\times(4+2-1+7-7-3+0+1)=0.375$$

$$S_y^2=\frac{1}{8-1}\left(\sum_{i=1}^{8}y_i^2-8\times0.375^2\right)=18.27$$

所以 $$\bar{x}=\bar{y}+50=0.375+50=50.375$$
$$S_x^2=S_y^2=18.27$$

(2) 选择适当常数 $a$ 及非零常数 $c$,作变换,使

$y_i=\dfrac{x_i-a}{c}$,$i=1,2,\cdots,n$,获得一组简单数据 $y_1,y_2,\cdots,y_n$。容易计算。

$$\bar{x}=c\bar{y}+a,\quad S_x^2=c^2S_y^2$$

简化计算的思路是通过常数 $a$,$c$ 的选择,使得样本值经过变换尽量简单。用简单数据算出 $\bar{y}$ 和 $S_y^2$ 后,再计算 $\bar{x}$ 和 $S_x^2$。

**例 7.1.7** 计算下列样本值的平均数和方差:

0.806, 0.882, 0.794, 0.800, 0.785, 0.793, 0.817, 0.820, 0.771

**解:**取 $a=0.800$,$c=0.001$,通过 $y_i=\dfrac{x_i-0.800}{0.001}$ 把样本值变换为

$y_i$:6, 82, −6, 0, −15, −7, 17, 20, −29

容易计算:$\bar{y}=\dfrac{1}{9}\times(6+82-6+0-15-7+17+20-29)\approx7.56$

$$S_y^2=\frac{1}{9-1}\left(\sum_{i=1}^{9}y_i^2-9\times7.56^2\right)\approx1010.7$$

则

$$\overline{x}_2 = 0.001 \times 7.56 + 0.800 = 0.80756, S_x^2 = 0.001^2 \times 1010.7 = 0.0010107$$

## §7.2　样本分布函数

### 一、样本分布函数

**定义7.2.1**　设从总体 $X$ 中抽取容量为 $n$ 的子样 $X_1, X_2, \cdots, X_n$，当顺序统计量 $X_1^*, X_2^*, \cdots, X_n^*$ 的值固定时，对任何实数 $x$，我们定义函数 $F_n^*(x)$：

$$F_n^*(x) = \begin{cases} 0 & x < X_1^* \\ \dfrac{k}{n} & X_k^* \leqslant x < X_{k+1}^* \quad k = 1, 2, \cdots, n-1 \\ 1 & x \geqslant X_n^* \end{cases}$$

称 $F_n^*(x)$ 为总体 $X$ 样本分布函数或经验分布函数(见图7.2.1)。

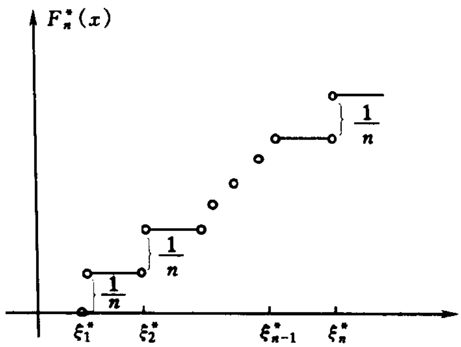


图7.2.1

易见，$F_n^*(x)$ 是单调、非降、右连续且在 $x = X_k^*$ 点有间断点，在每个间断点上跳跃量都是 $\dfrac{1}{n}$。显然，$0 \leqslant F_n^*(x) \leqslant 1$，并具有分布函数的其它性质。

由定义知，对于 $x$ 的每一数值而言，经验分布函数 $F_n^*(x)$ 为子样 $X_1, X_2, \cdots, X_n$ 的函数，它是一统计量，即为一随机变数，其可能取值为 $0, \dfrac{1}{n}, \cdots, \dfrac{n-1}{n}$, 1。事件"$F_n^*(x) = \dfrac{k}{n}$"发生的概率，由于 $X_1, X_2, \cdots, X_n$ 相互独立且有相同的分布函数 $F(x)$，因而它等价于 $n$ 次独立重复试验的贝努利概型中事件"$X \leqslant x$"发生 $k$ 次而其余 $n-k$ 次不发生的概率，即有

$$P\left\{F_n^*(x) = \frac{k}{n}\right\} = C_n^k \{F(x)\}^k \{1 - F(x)\}^{n-k}$$

其中 $F(x)=P\{X\leqslant x\}$,它是总体 $X$ 的分布函数。

对于任一实数 $x$,$F_n^*(x)$ 是事件“$X\leqslant x$”发生的频率,即在 $n$ 个样本值中小于或等于 $x$ 的个数与样本容量 $n$ 的比值,这与分布函数的概念是一致的。从频率与概率的关系知道,$F_n^*(x)$ 可以作为未知总体分布函数 $F(x)$ 的一个近似,且由格列汶科定理知,对任何实数 $x$,记

$$D_n=\sup_{-\infty<x<+\infty}|F_n^*(x)-F_n(x)|$$

则有

$$P\{\lim_{n\to\infty}D_n=0\}=1$$

这也是为什么可以依据样本的性质来推断总体特征的依据。

**例7.2.1** 从一批标准重量为500g的罐头中,随机抽取8听,测得误差如下(单位;g):

$$8,-4,6,-7,-2,1,0,1$$

求经验分布函数,并作出图形。

**解:**将样本值按大小顺序排列为

$$-7<-4<-2<0<1=1<6<8$$

则其样本分布函数为

$$F_8^*(x)=\begin{cases}0 & x<-7\text{时}\\ \frac{1}{8} & -7\leqslant x<-4\\ \frac{2}{8} & -4\leqslant x<-2\\ \frac{3}{8} & -2\leqslant x<0\\ \frac{4}{8} & 0\leqslant x<1\\ \frac{6}{8} & 1\leqslant x<6\\ \frac{7}{8} & 6\leqslant x<8\\ 1 & x\geqslant 8\end{cases}$$

其图形为图 7.2.2 所示。

### 二、概率密度函数的近似解 —— 直方图

设$(x_1,x_2,\cdots,x_n)$是从总体 $X$ 中抽得的容量为 $n$ 的样本$(X_1,X_2,\cdots,X_n)$的样本值;则利用直方图求概率密度函数的近似解的步骤如下:

(1) 从 $x_1,x_2,\cdots,x_n$ 中找出最小值 $x_1^*$ 和最大值 $x_n^*$;

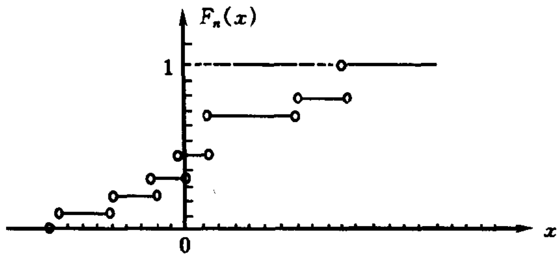


图 7.2.2

(2) 选择数 $a$ 略小于最小值 $x_1^*$，$b$ 略大于最大值 $x_n^*$，使区间 $(a,b]$ 易于等分，将其分为 $m$ 个子区间，得出分点为

$$a = t_0 < t_1 < t_2 < \cdots < t_m = b$$

(3) 记录样本值在第 $i$ 个区间 $(t_{i-1},t_i]$ $(i = 1,2,\cdots,m)$ 中的个数 $v_i$（称为频数），总体 $X$ 落在区间 $(t_{i-1},t_i]$ 中的频率 $f_i$ 的近似值为 $\frac{v_i}{n}$ $\left(f_i \approx \frac{v_i}{n}\right)$。

(4) 在坐标平面上，自左至右在各个小区间 $(t_{i-1},t_i]$ $(i = 1,2,\cdots,m)$ 上画竖直的长方形。对于第 $i$ 个长方形，以 $x$ 轴上小区间 $(t_{i-1},t_i]$ 的一段为底，以 $y_i$ 为高，其中 $y_i = \frac{f_i}{t_i - t_{i-1}}$ $(i = 1,2,\cdots,m)$。画出的图形如图 7.2.3 所示，大致经过每个竖直长方形上边的光滑曲线，它就是概率密度函数 $f(x)$ 的近似解。

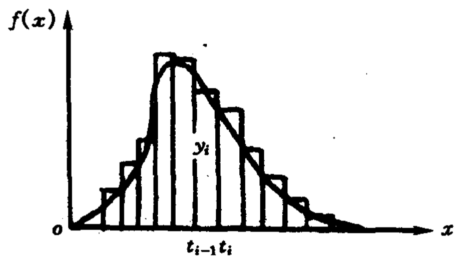


图 7.2.3

习惯上把竖直长方形构成的图称作直方图，实用中经常将它作为密度函数 $f(x)$ 近似解，而不构作光滑曲线。

当样本总容量 $n$ 较大时，$m$ 取 10 ~ 20；当 $n < 550$ 时，$m$ 取 5 ~ 6。但是，如果 $m$ 取得过大，就会出现某些小区间上频数为零，应尽量避免这种情况的出现。

为什么直方图可以近似代替概率密度函数 $f(x)$ 的曲线呢？从上面步骤(4)可知

$$y_i(t_i - t_{i-1}) = \frac{f_i}{t_i - t_{i-1}}(t_i - t_{i-1}) = f_i \approx P\{t_{i-1} < x \leqslant t_i\}$$

又因为
$$P\{t_{i-1} < x \leqslant t_i\} = \int_{t_{i-1}}^{t_i} f(x)\mathrm{d}x \approx f(t_i)(t_i - t_{i-1})$$

比较上面两式,自然有

$$y_i(t_i - t_{i-1}) \approx f(t_i)(t_i - t_{i-1})$$

亦即
$$y_i \approx f(t_i)$$

所以,所画直方图是概率密度函数 $f(x)$ 的近似解。

画直方图前,为了方便,一般先把数据填入设计的表格中,如下表:

| 组数 | 区间$(t_{i-1},t_i]$ | 频数 $v_i$ | 频率 $f_i$ | $y_i = \frac{f_i}{t_i - t_{i-1}}$ | 累积频率 |
|---|---|---|---|---|---|
| | | | | | |

其中 $f_i = \frac{v_i}{n}$。表中最后一项累积频率是为了画经验分布函数 $F_n^*(x)$ 的图形而设计的。为了使数据不要重复落在两个区间中,划分小区间时按半开区间 $(t_{i-1},t_i]$ 设计。

**例 7.2.2** 从某学校学生升学考试试卷中随机抽取 150 份,记录成绩如下:

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 83 | 90 | 70 | 64 | 56 | 92 | 68 | 78 | 80 | 70 | 88 | 64 | 92 | 79 | 77 |
| 71 | 76 | 98 | 52 | 65 | 91 | 71 | 63 | 85 | 85 | 75 | 50 | 75 | 81 | 88 |
| 70 | 63 | 61 | 72 | 78 | 81 | 72 | 86 | 65 | 87 | 85 | 65 | 80 | 82 | 84 |
| 72 | 67 | 80 | 73 | 84 | 50 | 85 | 84 | 48 | 85 | 58 | 60 | 85 | 87 | 99 |
| 60 | 75 | 94 | 76 | 64 | 71 | 95 | 58 | 65 | 60 | 69 | 92 | 62 | 73 | 70 |
| 43 | 76 | 70 | 82 | 65 | 93 | 64 | 70 | 58 | 80 | 80 | 81 | 66 | 62 | 88 |
| 75 | 81 | 82 | 76 | 75 | 56 | 68 | 68 | 85 | 88 | 90 | 66 | 84 | 65 | 92 |
| 79 | 60 | 91 | 62 | 93 | 64 | 57 | 68 | 74 | 86 | 74 | 76 | 95 | 79 | 101 |
| 55 | 81 | 98 | 96 | 70 | 73 | 86 | 82 | 77 | 65 | 94 | 75 | 66 | 75 | 65 |
| 76 | 90 | 104 | 58 | 54 | 95 | 76 | 86 | 83 | 72 | 80 | 76 | 78 | 75 | 89 |

依据这些资料分别作数学成绩的频数直方图,频率分布直方图和累积频率直方图。

**解**:在此例中,选 $a = 40, b = 105$,组数为 13,组距为 5,列表如下:

| 组数 | 区间$(t_{i-1}, t_i]$ | 频数 $v_i$ | 频率 $f_i$ | $y_i = \dfrac{f_i}{t_i - t_{i-1}}$ | 累积频率 |
|---|---|---|---|---|---|
| 1 | 40～45 | 1 | $\frac{1}{150}$ | $\frac{1}{750}$ | $\frac{1}{150}$ |
| 2 | 45～50 | 1 | $\frac{1}{150}$ | $\frac{1}{750}$ | $\frac{2}{150}$ |
| 3 | 50～55 | 4 | $\frac{4}{150}$ | $\frac{4}{750}$ | $\frac{6}{150}$ |
| 4 | 55～60 | 8 | $\frac{8}{150}$ | $\frac{8}{750}$ | $\frac{14}{150}$ |
| 5 | 60～65 | 15 | $\frac{15}{150}$ | $\frac{15}{750}$ | $\frac{29}{150}$ |
| 6 | 65～70 | 17 | $\frac{17}{150}$ | $\frac{17}{750}$ | $\frac{46}{150}$ |
| 7 | 70～75 | 19 | $\frac{19}{150}$ | $\frac{19}{750}$ | $\frac{65}{150}$ |
| 8 | 75～80 | 24 | $\frac{24}{150}$ | $\frac{24}{750}$ | $\frac{89}{150}$ |
| 9 | 80～85 | 21 | $\frac{21}{150}$ | $\frac{21}{750}$ | $\frac{110}{150}$ |
| 10 | 85～90 | 18 | $\frac{18}{150}$ | $\frac{18}{750}$ | $\frac{128}{150}$ |
| 11 | 90～95 | 12 | $\frac{12}{150}$ | $\frac{12}{750}$ | $\frac{140}{150}$ |
| 12 | 95～100 | 8 | $\frac{8}{150}$ | $\frac{8}{750}$ | $\frac{148}{150}$ |
| 13 | 100～105 | 2 | $\frac{2}{150}$ | $\frac{2}{750}$ | $\frac{150}{150}$ |

可依据列表做出图形。图 7.2.4 为频数直方图，图 7.2.5 为频率直方图，累积频率直方图留给学生完成。

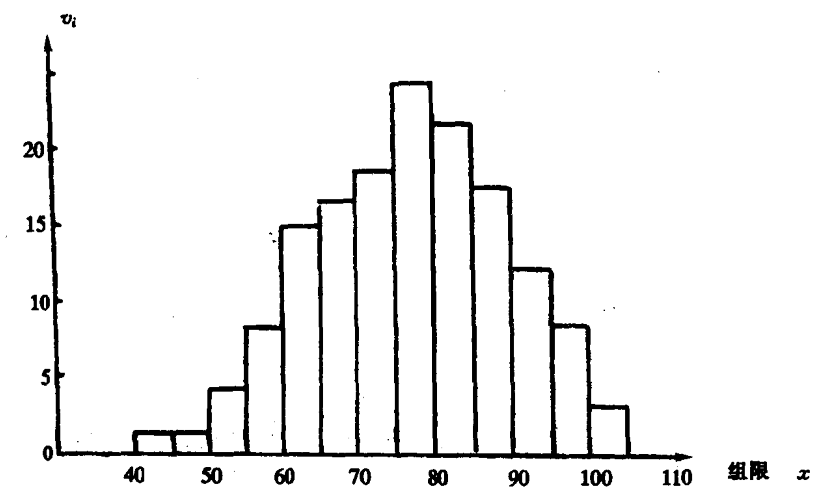


图 7.2.4