陶永德
杨庆霄 编

数理统计

辽宁科学技术出版社

数 理 统 计

陶永德　杨庆霄　编

辽宁科学技术出版社

（辽）新登字4号

数理统计

Shuli Tongji

陶永德　杨庆霄　编

辽宁科学技术出版社出版

（沈阳市和平区北一马路108号）

辽宁省新华书店发行　沈阳新华印刷厂印刷

开本：787×1092 1/32　印张：5　字数：110,000

1991年9月第1版　1991年9月第1次印刷

责任编辑：符　宁　责任校对：刘晓娟

封面设计：庄庆芳

印数：1—3,403

ISBN 7-5381-1103-4/O·57　定价：2.80元

前　言

本书是编者在东南大学（原南京工学院）使用的自编讲义基础上，参照国家教委公布的工科数理统计《基本要求》修改完成的。

本书包含数理统计基本概念；参数估计；假设检验；方差分析；回归分析等五章内容，在每章中介绍了有关方面一些基本知识，书后附有六种数表及习题答案。它适宜作为本专科少学时数理统计课的教学用书，全书教学时数不超过20学时，如果选教本书的前三章，授课只需12—14学时，本书在内容系统的安排上密切衔接高等教育出版社出版的工科类大学本科《概率论》教材，因而它既可以单独用于教学，也可以与上述概率论教材配合作为概率论与数理统计教本使用，同时，本书也可以作为自学参考书。

本书着重讲述数理统计的一些基本概念，概念的背景以及处理问题的思想方法，对一些定理的论证推导凡超出教学要求的都予省略，突出统计推断方法及其应用，在内容符合《基本要求》的前提下，在某些方面做了适当的拓宽，但以方法为主；全书注意理论与实际的联系，通过具体事例，以培养分析问题和解决问题的能力；在习题的选配上既照顾到面又考虑到数量要适中，希望通过这些习题的演算，以达到对基本理论的理解和基本方法的掌握，在文字上力求通俗易懂，便教便学。

本书稿曾得到国家教委高等学校工科数学课程教学指导委员会委员沈恒范教授详细地审阅并提出许多宝贵意见，在此表示深切地感谢。

限于水平，全书肯定有许多不足和错误，敬请读者斧正指教。

编者

1990.11.

目　　录

第一章　数理统计基本概念

数理统计与概率论都是研究大量随机现象规律性的学科，只不过它们所研究的侧重点有所不同，概率论着重对客观随机现象的规律性给出数学模型，譬如给出已知随机变量的分布，并对其性质与相互关系进行研究，而数理统计是以概率论为基础，着重从实际中观察、试验所搜集来的统计资料，进行分析研究其规律性，选择数学模型，对所考察的问题，作出估计与判断或控制与预测。但客观上只允许我们对随机现象进行次数不多的观察和试验，因此，我们所搜集的统计资料只能反映事物的局部特征而不是事物整体的特征，数理统计的任务就在于从统计资料所反映的局部特性来推断事物整体的特性，这种从局部特性推断整体特性的方法，具有普遍的意义，所以它的应用广泛，几乎在人类活动的一切领域中都能程度不等地找到它的应用。诸如产品质量的控制和鉴定；天气与地震的预报；病虫害的防止，良种的选择等等。

§1—1　总体与个体

在数理统计中，把研究对象的全体称为总体或母体，把组成总体的每一个研究对象（元素或单元）称为个体。

例1　考察某窑厂生产的一万块青砖的抗压强度，这一

万块青砖抗压强度的全体就是总体，每块青砖的抗压强度就是个体。

例 2 普查某中学学生的身高，这中学学生身高的全体，就是总体，每个学生的身高就是个体。

例 3 研究长江下游年降雨量的分布情况，若干年来记录下来的年降雨量的全体，就是总体，每年的降雨量就是个体。

例 4 一育苗室各处的温度的全体，就是总体，每处的温度就是个体。

通过以上四例可以看出，在实际中，总体指的是它的某个数量指标可能取值的全体，并非指事物的本身，这是由于在研究中，某个数量指标才是我们主要关心的事，如在例 1 中，我们主要关心的数量指标是这批青砖的抗压强度，因为抗压强度及其分布情况反映了这批青砖的质量，如在例 2 中，我们主要关心的是中学生的身高及其分布情况，因为这个数量指标身高可以从一个角度反映出这批中学生发育的状况，又如在例 3 中，我们关心的不是降雨本身而是年降雨量及其分布情况，因为它有助于预测某年这一地区是否发生旱涝现象。再如例 4 中，我们关心育苗室内的温度，因为温度的变化影响植物幼苗的生长。

很明显，从总体中随意抽取一个个体是数量指标ξ可能取的一个值，ξ是一个随机变量，ξ可能取值的全体就是总体，ξ的分布也就是总体的分布，以后，我们干脆把总体用随机变量ξ来代表。

在实际问题中，什么是总体，什么是个体不是一成不变的，而是要视我们研究任务来确定，如在例 2 中，我们研究的任务是要通过这所中学学生的体重这个数量指标来反映学生

的发育情况，那么体重可能取值的全体就是总体，每个学生的体重就是个体。而总体又分有限总体与无限总体，有限总体是指它含的个体数目为有限个，无限总体是指它包含的个体数目为无限多个，如例1、例2中的总体都为有限总体，而例4中总体是一个无限总体。

§1—2 样本与样本分布

在一个总体ξ中，抽取n个个体$\xi_1,\xi_2,\cdots,\xi_n$，这n个个体就称为总体ξ的样本或子样，n称为样本容量。

如例1，从一万块青砖中抽取100块，它们的抗压强度的全部就是该总体的一个样本。从总体ξ中随机抽取的样本$\xi_1,\xi_2,\cdots,\xi_n$是一个可能结果，可以把它们看成是n个随机变量，但是，一旦被抽取出来以后，它们都是具体数值，常记作$x_1,x_2,\cdots,x_n$，叫做样本值或样本观察值。

为什么要抽取样本？其目的在于通过样本对总体进行分析、估计、推断，为此就要求采用一个适当的抽样方法使所抽取的样本能够反映总体的特性，通常我们采用简单随机抽样法，用这个方法抽取样本$\xi_1,\xi_2,\cdots,\xi_n$，满足以下两点要求：(i)样本中每个分量ξ_i与总体ξ具有相同分布；(ii)$\xi_1,\xi_2,\cdots,\xi_n$是相互独立的。符合这两点的要求的样本，称为简单随机样本简称样本。以后内容中所提到的样本，无特殊说明，都是简单随机样本，不再重述。

对于总体ξ的样本$\xi_1,\xi_2,\cdots,\xi_n$，若ξ的分布函数为$F(x)$，那么样本的联合分布函数为$\prod\limits_{i=1}^{n} F(\xi_i)$；若$\xi$有分布密度函数为

$f(x)$，那么样本的联合分布密度为 $\prod_{i=1}^{n} f(x_i)$，可见，样本分布完全由其总体分布决定，但在一般情况下，总体的分布并不知道，而在数理统计中的任务之一，就是通过样本找到一分布来近似代替总体的分布，下面来介绍在实际中常用的做法。

（1）频率分布

我们从总体ξ中任意抽取出一个样本，就得到样本值x_1，$x_2,\cdots,x_n$，然后对这批数据，按下列步骤进行处理。

1° 找出 $x_1,x_2,\cdots,x_n$ 的最小值与最大值，并分别记作x_1^*、$x_m^*(m\leqslant n)$，即

$$x_1^*=\min\{x_1,x_2,\cdots,x_n\}$$
$$x_m^*=\max\{x_1,x_2,\cdots,x_n\}$$

2° 求出样本值中按由小到大次序排列的不同数值x_1^*，$x_2^*,\cdots,x_m^*(m\leqslant n)$的相应频数$n_1,n_2,\cdots,n_m$，且$\sum_{i=1}^{m} n_i=n$。

于是得样本频数分布：

ξ	x_1^*	x_2^*	$\cdots$	x_m^*
n_i	n_1	n_2	$\cdots$	n_m

与样本频率分布：

ξ	x_1^*	x_2^*	$\cdots$	x_m^*
$\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\cdots$	$\frac{n_m}{n}$

例 1 某厂生产一种200伏25瓦白炽灯泡，现对此厂生

产的这种灯泡重复进行120次抽测所得光通量ξ数据（单位：流明）列表如下，求出这个样本的频数分布与频率分布。

表1

216	203	197	208	206	209	206	208	202	203	206	213	218	207	208	202
194	203	213	211	193	213	208	208	204	206	204	206	208	209	213	203
206	207	196	201	208	207	213	208	210	208	211	211	214	226	211	223
216	224	211	209	218	214	219	211	208	221	211	218	218	190	219	211
208	199	214	207	207	214	206	217	214	201	212	213	211	212	216	206
210	216	204	221	208	209	214	214	190	204	211	201	216	211	209	208
209	202	211	207	202	205	206	216	213	206	207	200	198	200	203	208
216	208	216	206	222	213	209	219								

解：据数据表1.1，求得样本频数分布：

ξ	190	193	194	196	197	198	199	200	201	202	203	204
n_i	1	1	1	1	1	1	2	2	3	5	5	4
ξ	205	206	207	208	209	210	211	212	213	214	215	216
n_i	1	12	7	14	7	2	12	2	8	7	0	7
ξ	217	218	219	220	221	222	223	224	225	226		总计
n_i	1	4	3	0	2	1	1	1	0	1		120

与样本频率分布：

ξ	190	193	194	196	197	198	199	200	201	202
$n_i/120$	1/120	1/120	1/120	1/120	1/120	1/120	2/120	2/120	3/120	5/120
ξ	203	204	205	206	207	208	209	210	211	212
$n_i/120$	5/120	4/120	1/120	12/120	7/120	14/120	7/120	2/120	12/120	2/120
ξ	213	214	215	216	217	218	219	220	221	222
$n_i/120$	8/120	7/120	0/120	7/120	1/120	4/120	3/120	0/120	2/120	1/120
ξ	223	224	225	226	合计					
$n_i/120$	1/120	1/120	0/120	1/120	100/100					

（2）直方图

样本频率分布反映出数据的某种规律性，如在上例中样本值从190到226，稍加注意就可以看出光通量ξ取两头的数频率小，取中间一些数频率大，但它所反映的规律性还不够明确，为了获得进一步信息，需要对数据再进行处理，以便使数据所遵循的规律性能够比较明显地呈现出来，为此在上述两步的基础上，对数据进行分组作出频率分布直方图，一般我们采取如下步骤：

1° 决定组矩与组数。

选取起点与终点，一般起点a选得比x_1^*略小些，终点b选得比x_m^*略大些，选定a、b后，确定组距：

$$d=\frac{b-a}{h}$$

将$[a,b]$进行等分，也就是在$[a,b]$内插入$h-1$个分点$a<x_1'<x_2'<\cdots x'_{h-1}<b$，将$[a,b]$分成$h$个组（即小区间），使每个组必须有$x_i$落入，通常在试验数据较多（即样本容量较大）时，可分成10—20组，数据在100以内时，分成5—12组，但是对起点、终点、组矩、组数可视具体情况选取，并无硬性要求。

2° 数出频数，列出分组频率分布。

用选举唱票的办法，数出样本值$x_1,x_2,\cdots,x_n$落在每个组的数目，计算每个组的频数与频率和其它需要的数据。

3° 绘出频率分布直方图。

以样本值作横轴，以频率÷组距作为纵轴，在横轴上标出各分组的点，以各组的组距为底，画出高度等于频率÷组矩的小矩形，整个图形，叫做频率分布直方图（简称直方图）。

例2　根据例1所给出光通量ξ试验数据表，绘出频率分布直方图。

解　在例1的基础上

1°　决定组距与组数，我们为便于计算起见选取 $a=189.5$，$b=228.5$，$b-a=39$，组数$h=13$，组距 $d=3$。

2°　数出频数，列出分组频率分布：由例1与1°，得下列频率分布表：

表 1.2

分　　组	频　数	频　率	频率/组距	累积频率
189.5—192.5	1	0.008	0.003	0.008
192.5—195.5	2	0.017	0.006	0.025
195.5—198.5	3	0.025	0.008	0.050
198.5—201.5	7	0.057	0.019	0.107
201.5—204.5	14	0.117	0.039	0.225
204.5—207.5	20	0.167	0.055	0.392
207.5—210.5	23	0.192	0.064	0.583
210.5—213.5	22	0.183	0.061	0.766
213.5—216.5	14	0.117	0.039	0.883
216.5—219.5	8	0.067	0.022	0.950
219.5—222.5	3	0.025	0.008	0.975
222.5—225.5	2	0.017	0.006	0.992
225.5—228.5	1	0.008	0.003	1.000
合　　计	120	1.000		

3°　绘频率分布直方图：

在平面上，以光通量x作横轴，以$y=$频率/组距作为纵轴由上面频率分布表，作出如图1—1直方图。

这个图的明显优点在于每个小长方形的面积，近似地代表ξ取值落入“底边”的概率，如在图中有阴影的小矩形面

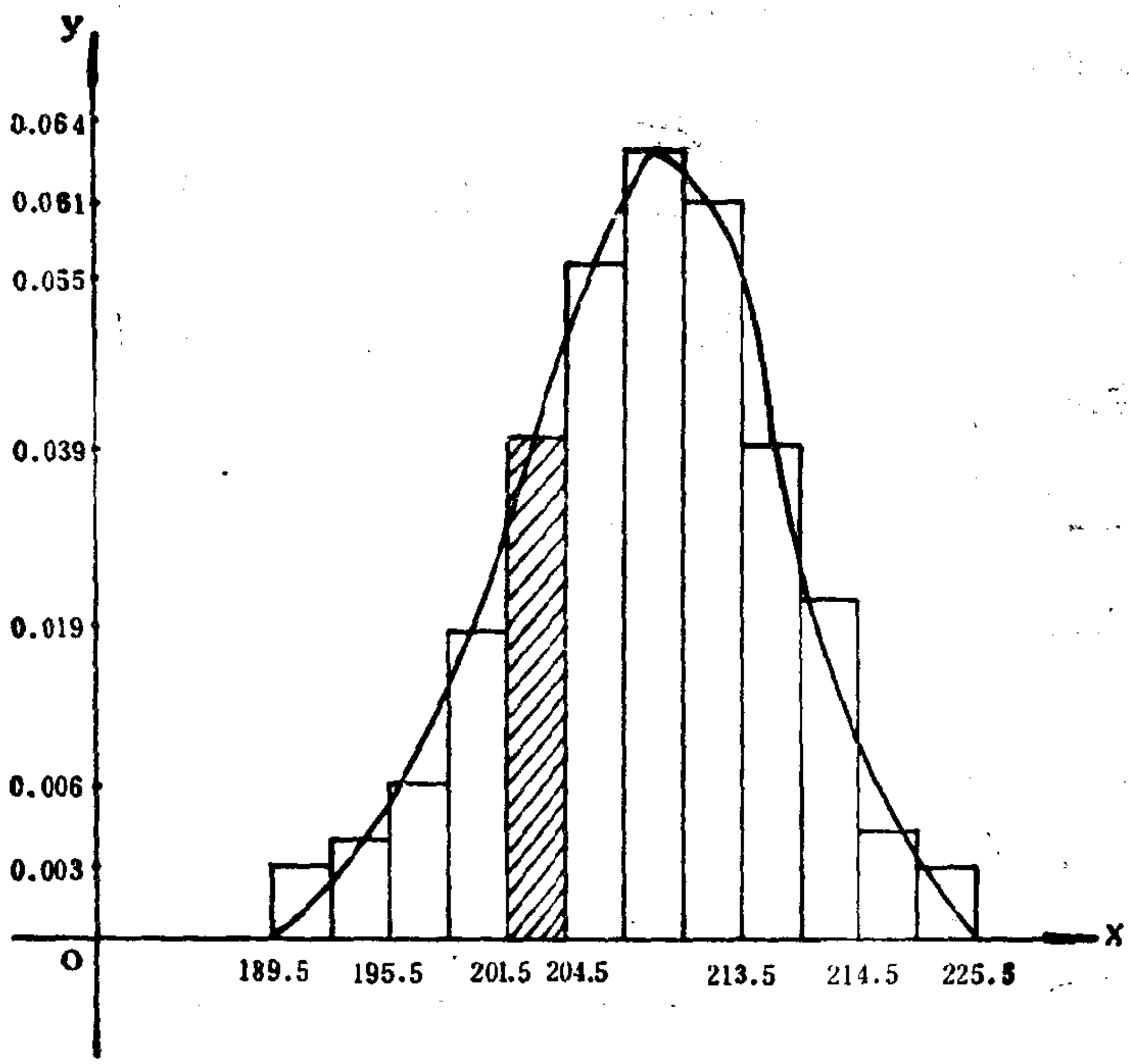

图1—1

积为

$$0.039\times(204.5-201.5)=0.117$$

是ξ取值落入“底边”的频率，它是ξ取值落入“底边”概率的近似值，而各组的频率之和等于1。故可借助频率分布直方图，形象地对总体分布作出估计，例如样本值落在204.5—213.5之间的频率是$0.167+0.192+0.183=0.542$，即在100只白炽灯泡中大约有54个光通量在204.5—213.5之间。

例3 为了研究玉米生长的变异情况，现从一堆玉米中抽出1000只，其长度从15厘米到29厘米不等，可写出这一千

只玉米的全部数据，既费时费力，又很难看出它们有多长、变异有多大，现将它们按组距为1厘米从15厘米到29厘米分成14组，得频率分布表如下：

表 1.3

分 组	频 数	频 率	频率/组距	累积频率
15—16	4	0.004	0.004	0.004
16—17	7	0.007	0.007	0.011
17—18	17	0.017	0.017	0.028
18—19	43	0.043	0.043	0.071
19—20	86	0.086	0.086	0.157
20—21	152	0.152	0.152	0.309
21—22	193	0.193	0.193	0.502
22—23	197	0.197	0.197	0.699
23—24	148	0.148	0.148	0.847
24—25	91	0.091	0.091	0.938
25—26	45	0.045	0.045	0.983
26—27	12	0.012	0.012	0.995
27—28	4	0.004	0.004	0.999
28—29	1	0.001	0.001	1.000

据表1.3的数据，绘出频率直方图。

解 在平面上以玉米长度x作横轴，以频率/组距y作纵轴，由表1.3，作出下面直方图（图1—2）。

与例2类似，我们可以借助频率分布直方图形象地对总体分布作出估计。

由于直方图1—2中每个长方形的底宽为1，故而每个小区间内玉米长的频率恰好等于小区间上的长方形的面积，直方图的总面积为1，有了这个直方图，就可以估计玉米长度落在某一区间内的概率是多少。例如落在区间［17,18］内频率为0.017，它就是概率的近似值。

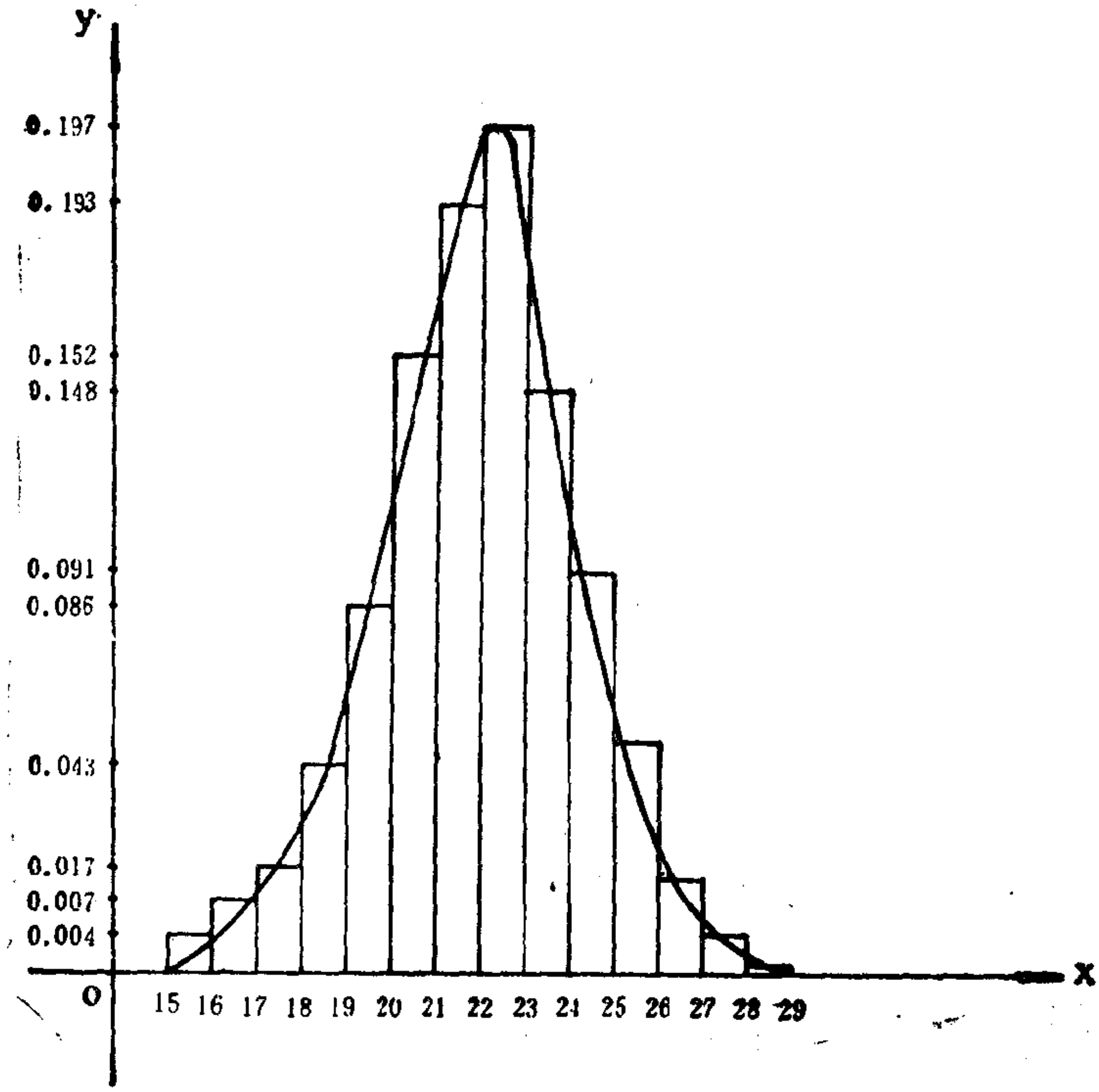

图1—2

（3）分布密度曲线的画法

设总体ξ 是连续型随机变量，其分布密度 $p(x)$ 未知，是否能根据总体 ξ 的样本值$x_1, x_2, \cdots, x_n$去估计 $p(x)$？下面来讨论这个问题。

我们知道根据样本值$x_1, x_2, \cdots, x_n$可以作出频率分布直方图，由贝努利大数定律可知，当n相当大时，第i组的频率f_i可以近似地表示ξ取值落入$[x_i', x'_{i+1}]$内的概率，即

 贝努利大数定律？