数理统计入门

陈德勤 编

近代数学知识丛书 JINDAISHUXUEZHISHICONGSHU

四川教育出版社

近代数学知识丛书

数理统计入门

陈 德 勤

四 川 教 育 出 版 社

一九八七年·成都

责任编辑　韩承训
封面装帧　邱云松

近代数学知识丛书　　数理统计入门

四川教育出版社出版　（成都盐道街三号）
四川省新华书店发行　自贡新华印刷厂印刷

开本787×960毫米 1/32　印张4.25 插页4 字数76千
1987年11月第一版　1987年11月第一次印刷
印数：1—1,480册

ISBN7—5408—0277—4/G·244
书号：7344·986　定价：0.94元

写在前面

从小学跨入中学，特别是步入高中以后，你将碰到许多完全陌生的东西，你将会发现数学天地是那样广阔，那样令人神往而又眼花缭乱，你可能会感到新奇，也可能会因无从下手而产生惶惑……其实，万事入门难，入门前或许不知所云，入了门便“不过如此”了.

这套知识小丛书将帮助你度过“入门难”这一关. 它用生动而浅显（有时还很有趣）的语言，准确而明晰的阐述，将近代数学中一些基本的概念、理论和运算一步一步展开，象一级级台阶，将你引入门去，使你并不感到突然和十分吃力. 许多地方还从生活现象入手，你读起来象是在聊家常，毫无枯燥之感. 所选的例题和练习也将帮助你加深理解. 当然，所谓入门，不过是给你一把小小的钥匙而已，一旦“入”得“门”去，那广阔的天地便由你自由驰骋，因而也就不再是这套书的任务了.

这套丛书所收书目见后勒口所列，大多为中学教学所涉及；少数关系不甚密，可辅你开拓视野，有兴趣者也不妨一读. 对于中学教师来说，这套丛

书也不失为良友，对提高教学质量或能助以一臂之力.

这本《数理统计入门》篇幅不大，语言也比较简练，每讲多从简单的实例出发，逐步引伸开去，浅显易懂，不难为初学者掌握. 当然，由于编者水平所限，缺点错误在所难免，敬请读者提出宝贵意见。

编　者

目 录

一 引 言

数理统计是一门应用数学，它于十九世纪末才开始形成一门独立的学科．数理统计所研究的问题可归结为两大类：一类是参数估计，另一类是假设检验．

数理统计是从实际考察资料出发，来研究随机变量的变化规律的一门科学．由于在研究中要大量应用概率论原理，因此，概率论是数理统计的理论基础．

数理统计在自然科学和社会科学中都有着十分广泛的应用．由于科学技术的迅速发展，特别是工业生产高速化、自动化，以及信息科学和电子计算机的发展，使数理统计的应用范围不断扩大．目前，在社会经济、人口调查、商品流通，以及物理、化学、天文、地理、气象、生物、医药等各个领域中，都要应用数理统计知识去解决一些理论问题和实际问题．

在数理统计中，不是对所要研究的全部对象一个一个地进行考察，而是只对其中的一部分，甚至是极少的一部分对象进行考察，考察的结果用统计

数字形式反映出来．数理统计研究的是如何根据这些由对一部分对象进行考察得到的统计数字，来分析和推断全部对象的情况．比如，某灯泡厂年产某种灯泡数百万只，若要检验这种灯泡的平均使用寿命（即每只灯泡大约可照明多少小时）是否合乎规定的标准，就要对灯泡使用情况进行实际考察．这有两种办法，一种是把灯泡一个一个地进行检验，这样做不仅浪费大量人力、物力，而且检验完毕后，该厂生产的这种灯泡也一个不剩，全都烧光了！这显然是行不通的．另一种方法是从该厂生产的这种灯泡中任意地抽取（注意：是任意抽取，决不能有意选取）极少的一部分，例如任意抽取几百只(约占总产量的万分之一)，对它们一个一个地进行详细考察，记录下每一只灯泡的使用寿命，然后运用数理统计知识，分析和推断出这种灯泡全体的平均使用寿命．这种检验方法才是切实可行的．

1·1 总体、个体和样本

总体、个体和样本是数理统计中最基本的概念．

我们所要研究的对象的全体称为**总体**，也叫**母体**；总体中的每一个元素都叫做**个体**．例如，我们要研究某厂产的灯泡的平均使用寿命，那么该厂产的全部灯泡就是总体，该厂产的每一只灯泡都是一个个体．又如，我们要研究四川省中学生健康情

况，那么四川省中学生的全体称为总体，四川省的每一个中学生都是一个个体.

从总体中任意地抽出来的一部分个体所组成的集合称为**样本**，也叫**子样**.

样本中含有的个体数目叫做**样本的容量**. 例如，从灯泡厂产的全部灯泡中任意抽取一百只，这一百只灯泡就是全部灯泡的一个样本，这个样本的容量等于一百.

1·2 随机抽样

在数理统计中，我们研究的是如何根据样本的特性来分析和推断总体的相应的特性，因而在选取样本时，要力求使样本能真正反映总体的情况. 为此，通常采用**随机抽样**的方法. 所谓随机抽样，就是丝毫不凭主观臆断，而是以对每一个个体都以完全平等的方式进行抽样，也就是要保证总体中的每一个个体被抽到的机会都完全相同. 注意，后面提到的抽样都是指的随机抽样，提到的样本都是指通过随机抽样得到的样本. 下面举两个例子说明随机抽样这一重要概念. 其一：把某电视机厂三月份生产的全部电视机都一一编上号码，然后用抽签的方法抽出一百个号码，并根据这些号码找出相应的一百台电视机. 这种抽取样本的方法就是随机抽样. 详细考察这样抽出来的一百台电视机的质量，利用数理统计的方法，就可以分析和推断出该电视机厂

三月份生产的全部电视机的质量如何．其二：某灯泡厂八月份生产了某种灯泡一千箱，每箱装灯泡一百只，如要从中抽取容量为一百的一个样本，该怎样抽取呢？如果象前一个例子那样，把十万只灯泡一一编上号后，再用抽签的方法找出相应的一百个灯泡，那实在太麻烦了．我们可以这样进行：先将箱子一一编上号码，用抽签的方法从中随机地抽取一百个箱子；然后又用抽签的方法从这一百个箱子中各抽取一个灯泡．这种抽取样本的方法也是随机抽样．详细考察这一百只灯泡的使用寿命，利用数理统计的方法，也可以分析和推断出该灯泡厂八月份生产的这种灯泡的平均使用寿命．在全国进行国产电视机质量评比时，参加评比的各厂都必须用随机抽样的方法抽取一定数量的电视机，而决不允许有意选择质量最好的产品去参加评比．

有人可能会说，进行随机抽样太麻烦了，既要对全部产品一一编号，又要进行抽签．其实不然．比如编号，象电视机这样的大型或贵重产品，本身就逐一编了号的；而象灯泡之类的普通产品，也都要装箱打包，箱上或包上一般也都有编号，里面的产品放置也是有规律的，即相当于给每件产品编了号．显然，编号这件事并没有给我们增加新的麻烦．至于抽签，也勿须真去施行，数学家们为此专门编制了随机数表，只要查一下表，就相当于进行了若干次抽签．

随机数表是把从0到9这十个数字随机排列而成

的表，其用途极为广泛，甚至连解密码都可能用到它. 本书末附有随机数表，下面介绍它的使用方法.

第一步，决定页码：闭上眼睛把铅笔放在随机数表上，若笔尖最接近的是奇数，则决定在第一页，否则取第二页.

第二步，决定起点：闭上眼睛把铅笔放在随机数表上，把01到40中最接近笔尖的二位数作为行数；再用同样的方法确定列数.

第三步，前进方向：从确定的页、行、列出发，若取的是一位或二位数，则往右取，如果到达右端，则移到下一行的左端继续往右取；若取的是三位以上的数，则往下取，如果到达下端，则移到下一列的上端继续往下取.

例如，要从学号为1到50的50名学生中随机地抽取7人，可以这样进行：先确定页码、行数和列数，比如已确定从第2页第8行第34列起，从表上查出对应数字是9；然后向右取二位数得到：95、41、92、85、15、44、48、20、06、26、16、50、46……再去掉其中大于50及重复的数字，剩下的数字中前7个就是被抽到的学生的学号，即41、15、44、48、20、06、26.

1·3 样本平均数

从总体中抽取一个容量为n的样本，考察由这

个样本中的n个个体得到的n个数据：

x_1，x_2，…，x_n，这n个数的算术平均数称为**样本平均数**，记作$\bar{x}$①：

$$\bar{x}=\frac{1}{n}(x_1+x_2+\cdots+x_n),$$

或简记为

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \text{②}.$$

总体中所有个体的平均数叫做**总体平均数**．以后我们将会看到，对由随机抽样得到的样本，样本平均数可以看作就是总体平均数．

【例1】 从某厂生产的某种灯泡中随机抽取20只，考察得到其使用寿命如下（百小时）：

31 32 30 31 28 29 33 30 29

28 29 33 32 34 32 30 34 29

33 33 计算样本平均数．

解：$$\bar{x}=\frac{1}{20}\sum_{i=1}^{20}x_i=\frac{1}{20}(31+32+\cdots+33)$$

$$=31\text{（百小时）}.$$

即样本平均数为3100小时，于是可将该厂生产的灯泡总体平均数看作是3100小时．这仅仅是一个估计，我们在后面将对此进一步进行研究．

注① $\bar{x}$读作"x bar"（x拔）。

② Σ是求和符号，读作"sigma"（西格玛）．

1·4 样本方差

样本中各数据与样本平均数的差的平方的平均数叫做**样本方差**，记作S^2：

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

或简记为

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

样本方差反映了一个样本中的数据偏离样本平均数的大小，样本方差越大，说明样本的波动越大.

当样本容量很大时，样本方差就近似地反映了总体的波动性，也就是样本方差很接近总体方差. 因此，在样本容量很大时，通常用样本方差去估计总体方差.

样本方差的算术平方根

$$S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

叫做**样本标准差**.

【例2】 从某厂甲、乙两车间生产的某种灯泡中分别随机地各抽取20只，考察得到其使用寿命

如下（百小时）：

甲：	31	32	30	31	28	29	33	30
	29	28	29	33	32	34	32	30
	34	29	33	33				
乙：	25	37	32	40	24	29	31	30
	31	32	29	31	35	27	30	34
	30	32	30	31				

分别计算它们的样本方差.

解：$\bar{x}_{甲}=\frac{1}{20}(31+32+\cdots+33)$

$=31$（百小时），

$S^2_{甲}=\frac{1}{20}\left[(31-31)^2+(32-31)^2+\cdots+(33-31)^2\right]$

$=3.7$〔（百小时）^{2}〕，

$\bar{x}_{乙}=\frac{1}{20}\left(25+37+\cdots\cdots+31\right)$

$=31$（百小时），

$S^2_{乙}=\frac{1}{20}\left[(25-31)^2+(37-31)^2+\cdots+(31-31)^2\right]$

$=12.9$〔（百小时）^{2}〕.

由此可见，虽然两个样本的样本平均数相同，但样本乙比样本甲的波动性大，从而可以估计乙车间生产的灯泡使用寿命比甲车间的波动性大，即乙车间产的灯泡质量比甲车间的差.

【例3】 求证：

$$S^2=\frac{1}{n}\left(\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right)$$

证：$\because \overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$,

$$\therefore\quad S^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

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

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

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

$$=\frac{1}{n}\left(\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right).$$

有时利用例3给出的等式计算样本方 差S^2更方便.

1·5 频率分布

某灯泡厂为了了解所生产灯泡的平均使用寿命

这一质量指标，随机地抽取了一个容量为100的样本，考察后得到其使用寿命如下（百小时）：

25	33	28	36	32	35	27	22	28
30	35	27	31	33	29	29	24	32
38	29	30	16	33	34	21	35	31
35	30	26	27	37	27	30	28	26
35	39	28	29	22	33	34	40	28
32	33	31	34	30	29	25	27	30
31	19	30	37	32	21	31	38	38
32	44	32	25	33	29	34	33	31
23	34	34	41	27	28	26	29	32
24	27	31	29	32	23	30	36	32
27	30	34	28	24	31	37	28	35
29								

由这些数据，我们可以算出样本平均数、样本方差．现在我们还想了解使用寿命在某一范围内（例如2950小时～3250小时）的灯泡所占的比例大小，为此，可以按照下面的步骤对这组数据进行适当的处理：

（1） 计算最大值与最小值的差．由上列数据，找出其最大值是44，其最小值是16，它们的差是

$$44-16=28（百小时），$$

（2）决定组距与组数．组距可以根据最大值与最小值之差的大小适当地选择，当样本容量在100以内时，常分成5～12组．在上列数据中，如果取组