实 用 工 程 数 学

数理统计及其应用

陈 凯　王玉孝

水 利 电 力 出 版 社

内 容 提 要

本书共分六章：1.实用统计方法；2.参数估计；3.假设检验；4.电力设备可靠性指标的现场统计方法；5.回归分析；6.马尔柯夫过程及其在电力系统可靠性问题中的应用。每章有小结和习题，并在书末附有习题答案，可便于读者自学解题。

本书可供从事电力、电机制造、自动控制和电子技术等专业的工程技术人员阅读，也可供有关专业的大专院校师生参考。

实用工程数学

数理统计及其应用

陈 凯 王玉孝

*

水利电力出版社出版

（北京三里河路6号）

新华书店北京发行所发行·各地新华书店经售

水利电力出版社印刷厂印刷

*

850×1168毫米 32开本 10.5印张 273千字

1986年6月第一版 1986年6月北京第一次印刷

印数0001—7700册 定价2.60元

书号 15143·5918

前　　言

近年来，我国电力工业迅速发展，工程实践不断向我们提出新课题，要求我们采用新技术、新方法。因此，过去不用或用得很少的数学分支，诸如：

（1）线性代数

（2）概率论

（3）数理统计

（4）图论

等在电力科技书刊和文献上已有了广泛的应用。

我国五十年代和六十年代由大专院校毕业多年来从事电力工业的工程技术人员在校学习期间多数没有接触到上述数学内容，为了实现四个现代化工程技术人员急需知识更新，迫切地需要有一套通俗易懂便于自学的实用工程数学方面的书籍，本书就是为了满足读者的这一需要而编写的。这一套书也可作为有关专业在校学生的数学参考书。

本书的特点是：

1）内容取舍以实际工作需要为依据，在保持数学体系的基础上，以应用广的内容为重点，其余内容予以删减。

2）注意基本概念和基本定理的叙述，附有较多例题；并尽可能地结合工程实际。

3）尽量作到深入浅出，文字通俗易懂。

4）每章有小结，借以帮助读者消化理解本章内容，并给学习某些章节的读者查阅有关内容提供方便。

5）每章均有习题，并附解答或提示，便于读者自学解题。

6）辟有专节或专章讨论有关数学在电力及自动控制工程中的应用，以便于读者了解如何应用所学数学理论来解决工程实际

问题。

这套书在编写过程中得到水利电力部电力科学研究院、原水利电力部工程数学及电子计算机应用学习班全体学员的大力支持和协助，在此表示衷心感谢。

限于我们的水平，书中难免存在缺点错误，热诚地希望读者批评指正。

编　者

1985年3月

目　　录

第一章　实用统计方法

数理统计是以概率论为理论基础，从实际观察数据出发，研究有关随机变量的分布及数字特征的。具体地说，数理统计的中心任务就是研究如何合理地搜集资料、整理数据、并利用这些数据来对有关随机变量的分布和数字特征等进行估计、分析和推断。数理统计的内容很多，最基本的是研究以下两个方面的问题：

（1）“估计”问题　由抽样取得的数据来推断母体的分布函数和数字特征；

（2）“检验”问题　根据所研究的问题的具体情况，针对母体提出某种“假设”，然后利用子样数据对所论“假设”进行检验，做出“接受”或“拒绝”假设的决定。

本章介绍如何整理由抽样得到的数据，以及利用这些数据推断母体分布和数字特征的简便实用方法。

第一节　母体和子样

在概率论中已经知道：很多实际问题（特别是自然现象和技术过程）中的随机现象可以用随机变量来描述。而要弄清一个随机变量，就必须知道它的分布，至少也要知道它的数字特征（期望和方差等）。例如，某钢铁厂某一天生产10000根16Mn型钢筋，按规定强度小于52kg/mm²的算作次品，怎样求这批钢筋的次品率p（也就是任取一根钢筋，它是次品的概率）呢？又如，灯泡厂生产灯泡，由于种种随机因素的影响，生产出来的灯泡的寿命是不同的，为了判断所生产灯泡的质量，怎样估计某月所生产的灯泡的平均寿命以及使用时数长短的相差程度呢？以上两个实际

问题的研究都可以归结为对于随机变量的研究。对于第一个问题，若考虑下面的随机试验："从该钢铁厂某一天生产的10000根16Mn型钢筋中任取一根，观察它是否为次品"，并定义随机变量

$$\xi=\begin{cases}1 & \text{当取到的一根钢筋为次品时}\\ 0 & \text{当取到的一根钢筋为正品时}\end{cases}$$

则求这批钢筋的次品率p的问题即等价于求随机变量ξ的分布。对于第二个问题，若考虑下面的随机试验："从该灯泡厂某月所生产的灯泡中任取一只测验它的寿命τ"，则τ为一随机变量。而估计"该厂某月所生产的灯泡的平均寿命以及使用时数长短的相差程度"的问题就是要估计它的期望和方差。求描述某随机现象的随机变量ξ的分布和数字特征是数理统计的主要课题。

对于第一个问题，要求ξ的分布就必须把该钢铁厂某一天生产的10000根钢筋作为研究对象；对于第二个问题，要研究随机变量τ就必须把该灯泡厂某月所生产的灯泡作为研究对象。

一、母体和个体

在数理统计中，把对某一问题的研究对象的全体称为母体，把组成母体的每个基本单元称为个体。在第一个问题中，该钢铁厂某一天生产的10000根钢筋的全体是母体，其中的每一根钢筋就是个体；在第二个问题中，该灯泡厂某月所生产的灯泡的全体是母体，其中的每一个灯泡就是个体。

于是，为了研究某个随机变量就必须研究相应的母体，而这个随机变量却又表示相应母体的某一数量特征。因此，以后对母体和表示母体的某一数量特征的随机变量不予区分。例如，说"母体ξ服从什么分布"指的就是表示该母体的某一数量特征的随机变量ξ服从什么分布；说"母体ξ的平均值"指的就是随机变量ξ的平均值等。

用什么方法来研究某个随机变量的分布及其数字特征呢？一个重要的方法是抽样法。这个方法的基本思想是：从相应的母体

中抽取一些个体来进行试验观察，由观察到的数据对母体进行估计或推断。由于在工业生产和科学研究等领域里，对母体的每一个个体进行普遍试验观察常常是办不到的。例如，要检验某批产品的质量而必须进行的破坏性试验就是这样。在上面的例子中，要检验钢筋的抗拉强度或者灯泡的寿命就需要进行破坏性试验。若对每一个个体都进行试验，则试验完成之时，也就是产品全部报废之时。所以根本就不能逐个检验，而且检验的个数还应适当地少。即使在有的问题中不须进行破坏性试验，例如，观察某个城市居民的平均身高或体重，也不可能进行普遍观察，因为那样做不仅耗费的人力物力太多，而且时间上也不允许。怎么办呢？只好进行抽样观察。

二、子样

在数理统计中，通常把从母体中抽取的若干个受试验观察的个体称为母体的子样。抽取的个体数目 n 称为子样的容量，并把这样的子样称为该母体的容量为 n 的子样。

1.简单随机子样

抽样本身就是在做随机试验，因此抽样观察的结果是随机的。抽样的方法有各种各样，其中最常用的也是最基本的抽样方法是简单随机抽样。这种抽样方法要求满足以下两条：1）在抽取子样的每个个体时，母体中的每一个个体都有同等概率被取到；2）在抽取子样的每一个个体时，母体中的个体成分并不改变。设有母体 ξ，设想用简单随机抽样的方法从中抽取容量为 n 的一个子样。则子样中的第一个个体的观察结果是随机变量，用 ξ_1 表示；第二个个体的观察结果是随机变量，用 ξ_2 表示；……；第 n 个个体的观察结果也是随机变量，用 ξ_n 表示。把这 n 个观察结果放在一起构成了母体的一个容量为 n 的子样，它可以用 n 维随机向量（ξ_1，ξ_2，……，ξ_n）来表示，也可以一般地记作 ξ_1，ξ_2，……，ξ_n。由于采用的是简单随机抽样，那么，根据它的两条抽样要求，可见 ξ_i 可以取得 ξ 能取得的每一个值，而且取值的概率规律也和 ξ 一样，所以 ξ_i（$i=1,2,\cdots\cdots n$）与母体 ξ 有相同的分

布，根据它的第2）条要求，可见$\xi_1,\xi_2,\cdots\cdots,\xi_n$是相互独立的。今后把满足这两条的子样（$\xi_1,\xi_2,\cdots\cdots,\xi_n$）称为母体$\xi$的一个简单随机子样，简称随机子样。由此可见，采用简单随机抽样的方法抽取容量为n的子样就是在做n次独立重复试验。

在具体问题中，怎样实现简单随机抽样呢？为了保证它的第1）条要求，在抽样时常常采取一些具体措施，以防抽取中的人为偏见，例如先将母体中的各个个体搅拌均匀后再抽取，或者先将母体中的每一个个体编号，用抽签的方法决定抽到几号个体。为了保证它的第2）条要求，可分下列情况区别处理：对于无穷母体（母体中个体的数目为无穷），只须一个一个地抽取，它的第2）条总是满足的；对于有限母体（母体中的个体数目为有限），则须进行有放回地抽取才能保证抽取每一个个体时母体中的个体成分不变。在实际问题中，做有放回的抽取常常是不可能的。例如，抽样进行破坏性试验的情况就是这样。对于这种情况，如果母体中个体的总数N远远大于子样的容量n（例如当$N/n\geqslant10$）时，可以近似地看成为无穷母体，因为此时从母体中去掉n个个体造成的母体中个体成分的改变微乎其微，可以忽略不计。值得注意的是，对于无穷母体或者近似按无穷母体处理的情况，一个一个地抽取n个个体进行试验观察与一下子抽取n个个体同时进行试验观察是没有区别的，故在实际问题中常常采取后面一种方法。例如，当对灯泡厂某月生产的灯泡进行寿命试验时，若灯泡的总数N与子样容量n之比$N/n\geqslant10$时，我们就是一下子从这批灯泡中随机地抽取n个灯泡同时进行寿命试验。

2.子样观察值

设想的抽样观察是随机试验，观察结果可以一般地用（ξ_1，ξ_2，……，ξ_n）来描述，但一次具体的抽样观察结果却是n个确定的观察值，记作$x_1,x_2,\cdots\cdots,x_n$，其中x_i是随机变量$\xi_i(i=1,2,\cdots\cdots,n)$的一个观察值。（$x_1$，$x_2$，……，$x_n$）是$n$维随机向量（$\xi_1$，$\xi_2$，……，$\xi_n$）的一个观察值。叫做随机子样观察值，简称子样。

母体、个体、子样是数理统计中的三个最基本的术语，它们的含义已在上边给出。对于子样这一术语，要注意区分随机子样和子样观察值。

三、子样的数字特征和母体的数字特征

在实践中遇到的母体，往往不但不知道它的分布，而且也不知道它的数字特征；有时知道它的分布类型也不知道它的数字特征。子样对于母体有代表性，它的数字特征对母体的数字特征也有代表性，尤其子样的容量相当大时更是这样。这就使我们有可能利用子样来估计母体的数字特征。

利用子样来估计母体的数字特征是数理统计的一个重要课题，这不仅由于母体数字特征本身在实际问题中的重要性，而且由于母体数字特征的确定对于确定母体分布也是很重要的。于是必须研究这样的问题：如何利用子样的数字特征估计母体的数字特征？这就是利用子样估计母体的数字特征问题。这里主要介绍有关利用子样的数学期望和方差来估计母体的数学期望和方差的问题。其它数字特征的估计问题也附带地提一下。更详细的将在第二章讲。

1. 子样的数字特征

设$(x_1, x_2, \cdots, x_n)$是母体ξ的一个子样观察值。它可以被看作是只取这n个值的离散型随机变量η的所有可能取值。若从中任取一个，则取得每一个值的概率都是$\frac{1}{n}$。因此有

$$M\eta=\sum_{i=1}^{n} x_i\frac{1}{n}=\frac{1}{n}\sum_{i=1}^{n} x_i$$

$$D\eta=\sum_{i=1}^{n}(x_i-M\eta)^2\frac{1}{n}=\frac{1}{n}\sum_{i=1}^{n}(x_i-M\eta)^2$$

定义 1.1 设母体ξ的容量为n的子样观察值$x_1, x_2, \cdots\cdots, x_n$的相应随机子样为$\xi_1, \xi_2, \cdots\cdots, \xi_n$，我们称算数平均值

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i \tag{1.1.1}$$

为母体ξ的子样平均值。称

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad S^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2 \qquad (1.1.2)$$

为母体ξ的子样方差。

显然，$\bar{\xi}$和S^2是随机变量，而$\bar{x}$是$\bar{\xi}$的一个观察值，s^2是S^2的一个观察值。

子样的数字特征，除平均值和方差外，还有k阶矩。

定义 1.2 称

$$\mu_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k$$

$$M_k=\frac{1}{n}\sum_{i=1}^{n}\xi_i^k \qquad (k=0,\ 1,2,\cdots) \qquad (1.1.3)$$

为子样的k阶原点矩。称

$$c_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k$$

$$C_k=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^k \quad (k=0,\ 1,2,\cdots) \qquad (1.1.4)$$

为子样的k阶中心矩。

显然，M_k和C_k都是随机变量，而μ_k和c_k分别为M_k和C_k的一个观察值，并且$\mu_1=\bar{x}$，$M_1=\bar{\xi}$，$c_2=s^2$，$C_2=S^2$。

2.母体的数字特征

今后不管母体ξ服从什么分布，总设

$$M\xi=a,\ D\xi=\sigma^2。$$

在大数定律中已经证明，当n充分大时，可以用子样的平均值作为母体数学期望的估计值，即当n充分大时

$$a\approx\bar{x} \qquad (1.1.5)$$

类似地可以证明当n充分大时

$$\sigma^2\approx s^2 \qquad (1.1.6)$$

即当n充分大时可以用子样的方差作为母体方差的估计值。

通常也把$\bar{\xi}$记作$\hat{a}(\xi_1,\ \xi_2,\ \cdots\cdots,\ \xi_n)$，并称为母体期望值$a$的估计量；把$\bar{\xi}$的一个观察值$\bar{x}$记作$\hat{a}(x_1,x_2,\cdots,x_n)$，并称为母体的期望值$a$的估计值。类似地，记$S^2$为$\hat{\sigma}^2(\xi_1,\xi_2,\cdots\cdots,$

ξ_n)，称为σ^2的估计量，记s^2为$\hat{\sigma}^2(x_1, x_2, \cdots, x_n)$，称为$\sigma^2$的估计值。

类似地，常把子样的M_k和μ_k分别作为母体的$m_k=M\xi^k$（$k=0, 1, 2, \cdots$）的估计量和估计值；把子样的C_k和c_k分别作为母体的$C_k=M(\xi-a)^k$（$k=0, 1, 2, \cdots$）的估计量和估计值。

第二节　数据整理及对母体分布的估计

设$x_1, x_2, \cdots\cdots, x_n$是母体$\xi$的一个子样观察值。这是一批数据，要利用这批数据来对母体分布作出初步估计。为了清楚地说明问题，下面我们通过一个具体例子来叙述处理问题的方法和步骤。

例 1　从某厂生产的220V、25W的白炽灯泡中随机地抽取了120个，测得其光通量的数据如下（单位：lm）：

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	221	218
218	190*	219	211	208	199	214	207	207	214	206	217
214	201	212	213	211	212	217	206	210	216	204	221
208	209	214	214	199	204	211	201	216	211	209	208
209	202	211	207	202	206	206	216	206	213	206	207
200	198	200	202	203	208	216	206	222	213	209	219

问该厂生产的这种灯泡的光通量服从什么分布？其数学期望和方差各等于多少？

所列120个灯泡的光通量，是该厂生产的这种灯泡的光通量（母体）ξ的一个$n=120$的子样观察值，它对于母体的特征应具有代表性，即子样的分布应该反映母体的分布，子样的数学期望和方差应当反映母体的数学期望和方差。

一、数据整理

为了求出子样的分布，进而推断母体的分布，可按下列步骤对所给数据进行整理。

（1）找出子样的最大值和最小值，确定子样的变化范围　在例1中，通过观察得

$$\max\{x_1, x_2, \cdots\cdots, x_{120}\}=226$$

$$\min\{x_1, x_2, \cdots\cdots, x_{120}\}=190$$

一般当该厂生产的这种灯泡的个数相当多时，可以认为光通量能够取得一个包含（190，226）在内的某个区间上的任何值，故可以认为它的光通量ξ是连续型随机变量，从而可以取从190到226之间的任何值。

一般地，对于连续母体ξ的一个确定的子样x_1，x_2，……，x_n有

$$\min\{x_1, x_2, \cdots\cdots, x_n\}=a$$

$$\max\{x_1, x_2, \cdots\cdots, x_n\}=b$$

则ξ可以取区间$[a, b]$内的任何值，而且还可能取得包含$[a, b]$在内的某个区间上的任何值。

（2）确定分组组数k　对于连续型随机变量ξ，由于$P\{\xi=x\}=0$，故只能考虑它的取值落在某一区间上的概率，对应地可以用子样落在一些区间上的频率来表征母体ξ落在这些区间上的概率。因此，一般把子样的取值范围分成k个相等的小区间，划分的原则之一是使得在每一个小区间上至少有一个x_i（$i=1$，……，n）落入。相应地，将x_1，x_2，……，x_n分成k组。分组的组数k，可根据子样容量n的大小按下列经验规律来确定：

当$n\leq 20$时，取$k=5\sim6$；当$n=40\sim60$时，取$k=6\sim8$；

当$n=60\sim100$时，取$k=8\sim10$；$n=100\sim500$时，取$k=10\sim20$。

n和k的这种关系不是完全定死的，允许有一定的灵活性。在例1中$n=120$，k取13。

（3）定组距$\Delta x=\dfrac{b-a}{k}$ 一般Δx取一个简单的数，在例1中

$$\Delta x=\frac{226-190}{13}=2.7692\cdots\cdots$$

可取$\Delta x=3$。此时，$3\times 13=39$，比$226-190=36$大3。一般母体的取值范围比$[a, b]$大，多出的数可以分在$[a, b]$的两端。可以均分，也可以某端多分一些。在例1中取$(a-0.5, b+2.5)$，即（189.5，228.5）。把（189.5，228.5）分成13个区间，其分点为189.5，192.5，195.5，198.5，……，228.5。分点比原始数据多取一位小数，这样可以保证子样中每一个数都落在所分的区间内，而不至于落在端点上，保证了子样的每个数都只属于一

表 1.1

组限	组中值 $\overline{x}_i$	频数 m_i	累计频数	频率 $f_i=\frac{m_i}{n}$	累计频率 $F^*(x_i')$
189.5～192.5	191	1	1	0.0083	0.0083
192.5～195.5	194	2	3	0.0167	0.0250
195.5～198.5	197	3	6	0.0250	0.0500
198.5～201.5	200	7	13	0.0583	0.1083
201.5～204.5	203	14	27	0.1167	0.2250
204.5～207.5	206	20	47	0.1667	0.3917
207.5～210.5	209	23	70	0.1917	0.5834
210.5～213.5	212	22	92	0.1833	0.7667
213.5～216.5	215	14	106	0.1167	0.8834
216.5～219.5	218	8	114	0.0667	0.9500
219.5～222.5	221	3	117	0.0250	0.9750
222.5～225.5	224	2	119	0.0167	0.9917
225.5～228.5	227	1	120	0.0083	1.0000
		120		1.0000	