第一节　随机样本

一、总体与个体

二、随机样本的定义

三、小结

数理统计学与概率论是两个有密切关系的姊妹学科，大体上可以说，概率论是数理统计学的基础，而数理统计学是概率论的重要应用，它补充和丰富了概率论.数理统计利用试验或观测到的数据来研究随机现象。

数理统计学的研究内容：

- 如何有效地收集和整理带有随机性质的数据.
- 对数据进行分析（称为统计分析），以及对所研究的问题做出推断（称为统计推断）和预测.——**本课研究内容！**

数理统计学是随着概率论的发展而发展起来的，在19世纪中期以前就已出现若干重要的成果，其中特别重要的是高斯（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.数理统计学是在20世纪上半叶逐渐发展为一门成熟的学科，其中皮尔逊（K.Pearson）和费歇尔（R.A.Fisher）作出了重大的贡献.1946年，克拉默（H.Cramer）发表的《统计学的数学方法》是第一部严谨、比较系统的数理统计著作，它标志着数理统计学进入了比较成熟的阶段.

在数理统计问题中，常常研究的是有关对象的数量指标，例如某型号电子元件的寿命，某型号的次品率等（即使研究对象如人的血型不是数量，也可以将其转化为数量）。为此，考虑与这一数量指标相联系的随机试验，对这一数量指标进行试验或观察。

一、总体与个体

1、试验的全部可能的观察值称为总体（注意：这些观察值不一定均不相同，数目也不一定是有限的）。

2、每一个可能观察值称为个体。

实例1 在研究2 000名学生的年龄时，这些学生的年龄的全体就构成一个总体，每个学生的年龄就是个体，(所以一共有2000个个体，个体不一定全不相同)．

3.容量

总体中所包含的个体的个数称为总体的容量.

4.有限总体和无限总体

容量为有限的称为有限总体.

容量为无限的称为无限总体.

实例2 某工厂10月份生产的灯泡寿命所组成的总体中，个体的总数就是10月份生产的灯泡数，这是个有限总体；而该工厂生产的所有灯泡寿命所组成的总体是一个无限总体，它包括以往生产和今后生产的灯泡寿命.

实例3 在考察某大学一年级男生的身高这一试验中，若一年级男生共2 000人，每个男生的身高是一个可能观察值，所形成的总体中共含2 000个可能观察值，是一个有限总体.

实例4 考察某一湖泊中某种鱼的含汞量，所得总体也是有限总体.

有些有限总体，它的容量很大, 我们可以认为它是一个无限总体.

实例5 考察全国正在使用的某种型号灯泡的寿命所形成的总体, 由于可能观察值的个数很多,就可以认为是无限总体.

5. 总体分布

由于总体中的每一个个体是随机试验的一个观测值，因此它是某一随机变量的取值，这样一个总体实质上对应于一个随机变量.相应地，对总体的研究就是对一个随机变量的研究.

随机变量的分布和数字特征我们叫做总体的分布和数字特征。

约定：**不区分总体与相应的随机变量X，统称为总体X.**

实例6 在2 000名大学一年级学生的年龄中，年龄指标值为 “15”， “16”， “17”， “18”， “19”， “20” 的依次有 9，21，132，1207，588，43 名，在总体中所占比率依次为

$$\frac{9}{2000},\quad \frac{21}{2000},\quad \frac{132}{2000},\quad \frac{1207}{2000},\quad \frac{588}{2000},\quad \frac{43}{2000},$$

即学生年龄的取值有一定的分布.

实例7 我们检验自生产线出来的零件是次品还是正品，以0表示产品是正品，以1表示产品为次品. 设出现次品的频率为 p（常数），那么总体是由一些“0”和一些“1”所组成，这一总体对应于一个具有参数为p的（0-1）分布：

$$P\{X=x\}=p^x(1-p)^{1-x},\quad x=0,1$$

的随机变量.

6. 简单随机样本

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判断的.

被抽出的部分个体叫做总体的一个样本.

所谓从总体抽取一个个体, 就是对总体X 进行一次观察并记录其结果.

在相同的条件下对总体X进行n次独立重复的观察，将n次观察按顺序记为 $X_1, X_2, \ldots, X_n$ 。这是n个独立同分布的随机变量，称为来自总体的一个**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 $X_1, X_2, \ldots, X_n$ 的观察值，称为样本值 .

二、随机样本的定义

1.样本的定义

设 X 是具有分布函数 F 的随机变量，若 X_1, $X_2, \cdots, X_n$ 是具有同一分布函数 F、相互独立的随机变量，则称 $X_1, X_2, \cdots, X_n$ 为从分布函数 F (或总体 F、或总体 X) 得到的容量为 n 的简单随机样本，简称样本.

它们的观察值 $x_1, x_2, \cdots, x_n$ 称为样本值，又称为X 的 n 个独立的观察值.

2. 简单随机抽样的定义

获得简单随机样本的抽样方法称为简单随机抽样.

若 $X_1, X_2, \cdots, X_n$ 为 F 的一个样本, 则 $X_1, X_2, \cdots, X_n$ 相互独立, 且它们的分布函数都是 F, 所以 $(X_1, X_2, \cdots, X_n)$ 的分布函数为

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

又若 X 具有概率密度 f, 则 $(X_1, X_2, \cdots, X_n)$ 的概率密度为

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

三、小结

基本概念: 个体 总体 $\begin{cases}\text{有限总体}\\ \text{无限总体}\end{cases}$ 随机样本

说明1 一个总体对应一个随机变量X, 以后将不区分总体和相应的随机变量, 统称为总体X.

说明2 在实际中遇到的总体往往是有限总体, 它对应一个离散型随机变量;当总体中包含的个体的个数很大时, 在理论上可认为它是一个无限总体.

作业：5、6（1）（2）