

数理统计部分

数理统计是具有广泛应用的一个数学分支.它以概率论为理论基础,根据试验或观察得到的数据,来研究随机现象,对研究对象的客观规律作出种种合理的估计和判断。

数理统计主要内容

第六章 样本及抽样分布

第七章 参数估计

第八章 假设检验

引言

概率论（基础）讨论了如下问题：对随机现象进行研究，在数学上建立概率的公理化体系；引入基本概念、揭示常见各类随机现象的规律，总结为基本的随机模型和分布律，并研究它们的性质及数字特征；对大量随机因素综合影响的结果，以极限定理为内容作了介绍。这样对随机现象的研究，已有了基本的概念、思想方法和工具。但当**我们实际动手研究并解决一个实际问题时，会立即遇到下面的问题：**

（1）这个随机现象可以用什么样的分布律（分布函数）来刻画，这种分布的选择合理吗？

（2）所选用的分布的参数是多少？如何估计和确定这些参数？

我们对要研究的这个实际问题往往所知甚少，这样只能求助于观测，合理地取得一些数据，据此作出统计上的推断，回答上述问题，从而着手去解决问题。而这就是数理统计的基本且主要任务。



更准确地说 数理统计的主要内容是：

1. **实验设计和研究**，即研究如何更合理、更有效地抽取样本，从而获得观测数据和资料的方法。

2. **统计推断**：如何利用一定的数据资料，对所关心的问题，得出尽可能准确的统计结论：

(1) **估计**——从局部观测资料的统计特征，推断所观测对象的总体特征（包括总体分布与数字特征）；

(2) **假设检验**——依据抽样数据资料，对总体的某种假设做检验，从而决定对此假设是拒绝还是接受。



例 某钢筋厂日产某型号钢筋10000根，质量检验员每天只抽查50根的强度，于是提出以下问题：

(1) 如何从仅有的50根钢筋的强度数据去估计整批（1000根）钢筋的强度平均值？又如何估计这批钢筋强度偏离平均值的离散程度？ **-----参数估计**

(2) 若规定了这种型号钢筋的标准强度，从抽查得的50个强度数据如何判断整批钢筋的平均强度与规定标准有无差异？ **-----假设检验**



第六章

样本及抽样分布

§ 6. 1 总体与样本

§ 6. 2 抽样分布

§ 6.1 随机样本

总体与个体

1.总体 (Population) 研究对象的全体
试验的全部可能的观察值称为总体.

2.个体 总体中每个对象
总体中的每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.



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.

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3 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.

- [说明]** (1) 每个个体是某随机变量 X 的值, 一个总体对应一个随机变量 X ;
- (2) X 的分布函数与数字特征分别称为总体的分布函数与数字特征;
- (3) 今后将不区分总体和相应的随机变量, 笼统称为总体 X .

例如,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.

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。

当 n 次观察一经完成，我们就得到一组实数 x_1, x_2, \dots, x_n ，它们依次是随机变量 X_1, X_2, \dots, X_n 的观察值，称为样本值。

样本与样本值

1. 样本 从总体 X 中**随机地**抽取 n 个个体 X_1, X_2, \dots, X_n , 这样取得的 X_1, X_2, \dots, X_n 称为来自总体 X (分布函数 F 相同) 的一个**样本**;

2. 样本容量 样本中个体的数目 n ;

3. 样本值 X_1, X_2, \dots, X_n 的一组观察值 x_1, x_2, \dots, x_n ;

4. 简单随机样本 在总体中抽取样本的目的是为了对总体的分布规律进行各种分析推断, 这就要求抽取的样本能够反映总体的特点, 为此必须对随机抽取样本的方法提出如下**要求**:

(1) **独立性** 要求 X_1, X_2, \dots, X_n 是相互独立的随机变量;

(2) **代表性** 要求样本的每个 X_i ($i=1, 2, \dots, n$)与总体 X 具有相同的分布.

满足以上两个条件的样本称为**简单随机样本**, 简称**样本**.

[注]

- (1) 样本 X_1, X_2, \dots, X_n 相互独立, 且与总体 X 同分布;
- (2) 样本 X_1, X_2, \dots, X_n 可看成一个 n 维随机向量, 记为 (X_1, X_2, \dots, X_n) ; 样本值记为 (x_1, x_2, \dots, x_n) ;
- (3) 若总体 X 具有分布函数 $F(x)$, 概率密度 $f(x)$, 则样本 (X_1, X_2, \dots, X_n) 的分布函数及概率密度为:

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

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

- (4) 以后说到 “ X_1, \dots, X_n 是来自某总体的样本” 时, 若不特别说明, 就指简单随机样本.

小结

基本概念：**个体** **总体** $\left\{ \begin{array}{l} \text{有限总体} \\ \text{无限总体} \end{array} \right.$ **随机样本**

样本的性质

1. X_1, X_2, \dots, X_n 都与总体 X 同分布;
2. X_1, X_2, \dots, X_n 相互独立.

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; 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.