第 六 章 数理统计的基本概念

第一节 总体、样本与统计量

1、总体、样本的概念

2、统计量

3、小结、思考

数 理 统 计 的 引 入

某厂生产的一批产品中次品率为 p 。
从中抽取10件产品装箱。

1）没有次品的概率

2）平均有几件次品

3）为以 0.95的概率保证箱中有10件正品，
箱中至少要装多少件产品。

所有这些问题的关键是 p 是已知的！

如何获取 p ？ ⟶ 数理统计的任务！

一个很自然的想法：

首先从这批产品中随机抽取几件产品进行检验。

其次利用概率论的知识处理实测数据。

1）如何估计次品率 p？ 参数估计问题

2）如果以 $p<0.01$ 为出厂的标准，这批产品能否出厂？

假设检验问题

数理统计:

是以概率论为理论基础，　研究如何以有效的方式收集和整理受随机因素影响的数据(称为试验设计)，　研究如何合理地分析这些数据从而作出科学的推断（称为统计推断)。

这两部分有密切联系，实际应用中更应前后兼顾。我们将主要介绍统计推断方面的内容。

§1 总体、样本与统计量

一、总体与个体

总体 研究对象的单位元素所组成的集合。

个体 组成总体的每个单位元素。

例如： 我们考察本校男生的身高和体重情况，则将本校的所有男生视为一个总体，而每一位男生就是一个个体。

在实际问题中，我们关心的是总体的一项或几项数量指标值的分布情况。

例如:考察电子元件的寿命时，我们感兴趣的是总体中有多少个寿命在500h以上。有多少个寿命在10h以下，从总体中任取一个个体，其寿命在10h以下的概率有多大？... 这样要研究的总体实质上是某个概率分布，因此我们将总体定义为一个随机变量 X 。

我们以后就把总体和数量指标 X 等同起来。

所谓总体分布就是指数量指标 X 的分布。

总 体 是 随 机 变 量。

二、样本

为了研究总体的性质，看起来最好是把每个个体都加以观测研究，但这往往是不必要的，有时甚至是不可能的。例如，研究一批炮弹的杀伤力，不可能将每一发炮弹都拿来作试验。

一般，我们是从总体中抽取一部分，比如说 n 个进行观测，再根据这 n 个观测值去推断总体的性质。

样本是按照一定的规则从总体中抽取的一部分个体

抽样就是抽取样本的过程

样本容量是样本中个体的数目 n

为了使样本具有代表性，抽样必须是随机的。并且抽取一个个体后总体成分不变。

我们称与总体同分布且相互独立的样本为简单随机样本，简称样本

样本是一组随机变量 记为 $X_1, X_2, \cdots, X_n$

我们对样本进行一次观测时，其具体数值记为x_1，$x_2, \cdots, x_n$ 称为样本观测值，简称样本值

若总体 X 的分布函数为$F(x)$， $X_1, X_2, \cdots, X_n$为来自 $F(x)$ 的一个样本，则对 n 维随机变量 $(X_1, X_2, \cdots, X_n)$

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

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

三、统计量

统计量是完全由样本决定，不包含未知参数的函数 $g(X1, X2, \ldots, Xn)$。统计量的分布称为抽样分布。

例 6.1.1 设总体 $X \sim B(1, p)$，其中 p 是未知参数，$(X_1, X_2, \ldots, X_5)$ 是来自 X 的简单随机样本，指出：

$$X_1 + X_2, \quad \max_{1 \le i \le 5} X_i, \quad X_5 + 2p, \quad (X_5 - X_1)^2$$

哪些是统计量，为什么？

$(X1, X2, \ldots, X5)$ 的联合分布律？

$$\prod_{i=1}^{5} P\{X_i = x_i\} = \prod_{i=1}^{5} p^{x_i}(1-p)^{1-x_i} = p^{\sum\limits_{i=1}^{5} x_i}(1-p)^{5-\sum\limits_{i=1}^{5} x_i} \qquad x_i = 0, 1$$

对于相应的样本值$(x_1, x_2, \ldots, x_n)$，

$g(x_1, x_2, \ldots, x_n)$称为统计量的值，简称**统计值**

总 体 是 随 机 变 量。

样 本 是 一 组 随 机 变 量

统计量 是 随机变量(或向量）

常见统计量：

样本均值：

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

样本方差：

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

样本 k 阶原点矩：

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$$

样本 k 阶中心矩：

$$M_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$$

统称样本矩。

思考： 样本矩是不是随机变量？

总体矩（即第四章中定义的矩）呢？

样本矩 是 随机变量， 总体矩 是 数值。

几个重要关系式：

$$\boxed{A_1 = \overline{X}} \qquad \boxed{M_2 = \frac{n-1}{n} S^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = A_2 - A_1^2}$$

思考： 设 $E(X) = \mu$ ， $D(X) = \sigma^2$

则 $E(A_1) = \mu$

又若 $E(S^2) = \sigma^2$ ，则 $E(M_2) = \dfrac{n-1}{n}\sigma^2$

$\boxed{X,\ S^2,\ A_k,\ M_k}$ $\Longleftrightarrow$ $\boxed{x,\ s^2,\ a_k,\ m_k}$

统计量 统计值

注意到样本$X_1, X_2, \cdots, X_n$ 是相互独立同分布的随机变量

由中心极限定理，当n 充分大的时候

样本均值

$\overline{X}$ 服从正态分布或近似服从正态分布。