第六章　数理统计的基本概念

1.	总体、样本与统计量
2.	常用的分布

电子科技大学数学科学学院 杜鸿飞　hongfeidu@qq.com

第6章1节　总体、样本与统计量

一. 引言

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.

引例1~5

二、总体与个体

总体：是指研究对象的全体所组成的集合.

个体：是指组成总体的每个对象即元素.

例如，我们要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.

我们关心的是**总体的的一项或几项数量指标** X .
如上例，我们只考虑男生的身高和体重，不考虑男生的视力、成绩等.

二、总体与个体

由于数量指标 X 往往是一个随机变量，具有一定的分布.

例如，我们考察电子元件的寿命时，则寿命 X为其一个数量指标，且 X 是服从指数分布的随机变量.

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　体视为随机变量

三、样本

为了研究总体的性质，乍看起来，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时，不可能将每一发炮弹都拿来作试验.

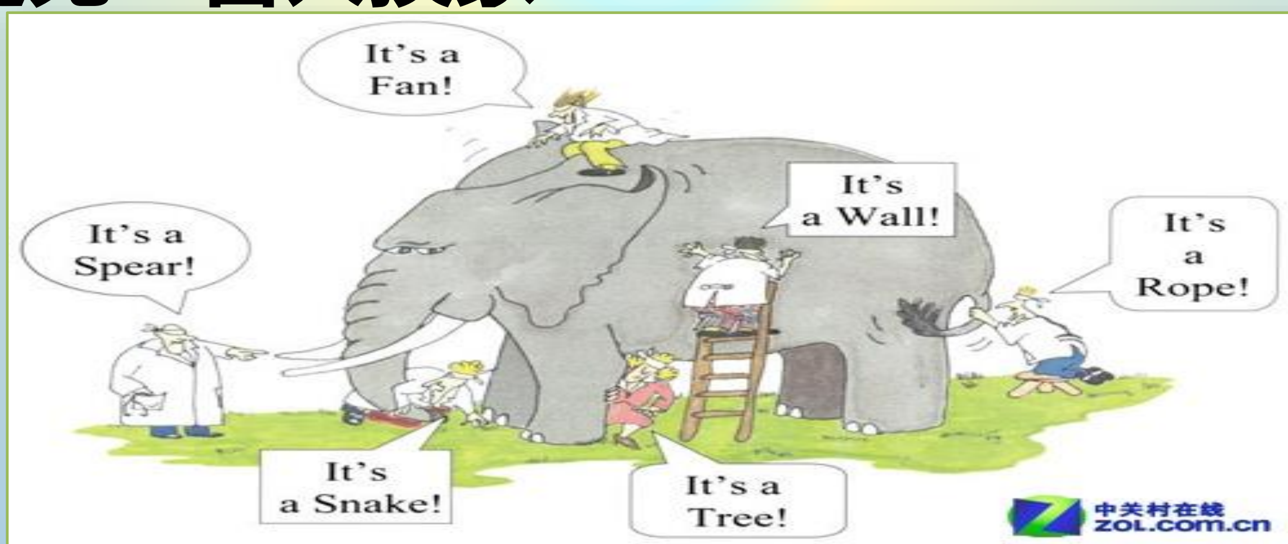
第6章1节　总体、样本与统计量

三、样本

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**.

我们希望“一叶知秋”、“管斑窥豹”，但又要避免“盲人摸象”.

抽样方法



推断方法

三、样本

样本： 按照一定的规则从总体中抽取的一部分个体；

抽样： 抽取样本的过程；

样本容量： 样本中个体的数目 n.

规则？　过程？　数目？

TIPS

从民意测验看抽样

三、样本

为了使样本具有代表性，抽样必须是随机的.

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 $X_1, X_2, \cdots, X_n$ 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.

注意大小写的区分!

三、样本

若总体 X 的分布函数为 $F(x)$，$X_1, X_2, \cdots, X_n$ 为来自$F(x)$的一组样本，则对 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 有

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

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

若总体 X 的分布未知，如何根据样本 X_1，X_2，$\cdots$，X_n 推断总体的分布或参数？

四、统计量

统计量：完全由样本决定，不包含未知参数的函数$g(X_1, X_2, \dots, X_n)$.

TIPS　统计量的判断

对于相应的样本值$(x_1, x_2, \dots, x_n)$，$g(x_1, x_2, \dots, x_n)$称为**统计量的值**，简称**统计值**.

第6章1节　总体、样本与统计量

四、统计量

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)

第6章1节　总体、样本与统计量

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***.

常见统计量

思考1:

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 (据定义)

常见统计量

思考2：如何用样本矩(统计量)推断总体特征?

用样本矩估计总体矩

$$E(X_i)=E(X)=\mu \qquad E(\overline{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\mu$$

可用样本均值估计μ——其原理是什么?

独立同分布大数定律

几个重要关系式

$$A_1 = \overline{X}$$

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

思考： 设 $E(X)=\mu$ ， $D(X)=\sigma^2$

则 $E(A_1) = \mu$

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

$\overline{X}$, S^2, A_k, M_k ⟺ $\overline{x}$, s^2, a_k, m_k

统计量　　统计值

常见统计量

思考3：中心极限定理在这里有何作用?

1. 正态分布是最常见的分布，本书统计推断部分将以正态分布为主进行分析

2. 独立同分布随机变量之和近似为正态分布，故：**样本容量足够大时，样本矩(统计量)可近似为正态分布**