



哈爾濱工業大學

# 总体与样本





**数理统计：**是以概率论为理论基础，研究怎样用**有效**的方法去收集、整理、分析带随机影响的**数据**，以便对所研究的问题给出估计和推断，为决策提供依据和建议.

# 概率论与数理统计的不同



◆ 概率论、数理统计都是研究随机现象的统计规律性的数学分支，但两者**研究角度**不同.

**概率论** 从已知分布出发，研究随机变量 $X$ 的性质、规律、数字特征等等.

**数理统计** 研究对象 $X$ 的分布不知道或不完全知道，观察它的取值(采集数据)，通过分析数据来推断 $X$ 服从什么分布或确定未知参数.

# 数理统计研究问题的方法



例如，要检验某国产品牌轿车的耗油量，无论轿车总量是有限还是无限，从人力、物力和时间角度考虑，都不能对轿车逐一检验. 只能抽取一部分轿车进行检验，通过这一部分轿车的耗油量来推断这种轿车总体的耗油量. 以部分数据信息来推断整体相关信息，是数理统计研究问题的基本方法.



# 总体、个体



- 总体：研究对象的全体.

例如：要研究某大学学生的身高，**总体是该校的全体学生**. 每个大学生是**个体**. 每个大学生有许多**指标**，如身高，体重，年龄，…，我们仅研究“大学生的身高”**这项指标**.

- 个体：总体中每个成员.

# 总体分类



**有限总体：** 总体包含有限个个体.

如:考察某大学大一3000名男生的身高;

**无限总体：** 总体包含无限个个体.

如:测量一湖泊任一地点的深度.



► 在统计研究中，人们关心总体仅仅是关心其每个个体的一项(或几项)数量指标和该数量指标在总体中的分布情况.

总体的某项数量指标 $X$ ，对不同的个体，取值不同，这些数值满足一定的概率分布，因此数量指标 $X$ 是随机变量. 通常我们把总体和数量指标 $X$ 等同起来.



这样，总体就可以用一个随机变量 $X$ 及其分布来描述，称为一维总体.

本课程主要研究一维总体，多维总体是多元统计分析主要研究的对象.



# 样本



**样本**：按一定规则从总体中抽取的一部分个体.

**样本容量**：样本中所含个体的数目.

**抽样**：抽取样本的过程.

由于抽样的随机性，样本也具有随机性，通常容量为 $n$ 的样本用随机变量  $X_1, \dots, X_n$  表示.

# 简单随机样本



■ 若 $X_1, X_2, \dots, X_n$ 是相互独立的(独立性)且与总体 $X$ 有相同的分布(代表性),则称 $X_1, X_2, \dots, X_n$ 为来自总体 $X$ 的一个容量为 $n$ 的简单随机样本,简称为 $X$ 的一个样本.

获得简单随机样本的抽样称为简单随机抽样.

后面所说的样本均指简单随机样本.

■ 样本 $(X_1, X_2, \dots, X_n)$ 的每一个观察值 $(x_1, x_2, \dots, x_n)$ 称为样本值或样本的一次实现.





- 样本值的集合称为**样本空间**.
- 总体分布决定了样本取值的概率规律，因而可以由样本值去推断总体.
- ◆ 数理统计的主要任务之一就是**研究如何根据样本推断总体**.

若总体 $X$ 的分布函数为 $F(x)$ ,  $X_1, X_2, \dots, X_n$ 为来自 $X$ 的一个样本, 则样本 $(X_1, X_2, \dots, X_n)$ 的联合分布函数为

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





**例1** 加工某种零件时，每件需要的时间服从均值为  $1/\lambda$  的指数分布. 今以加工时间为零件的数量指标，任取  $n$  件零件构成一个容量为  $n$  的样本，求样本的分布.

**解** 设零件的加工时间为总体  $X$ , 则  $X \sim E(\lambda)$ , 其概率密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$



样本 $(X_1, X_2, \dots, X_n)$ 的概率密度为

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^n f(x_i) \quad f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases} \\ &= \begin{cases} \prod_{i=1}^n \lambda e^{-\lambda x_i}, & x_i > 0, (i = 1, 2, \dots, n). \\ 0, & \text{其他.} \end{cases} \\ &= \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n x_i}, & x_i > 0, \\ 0, & \text{其他.} \end{cases} \end{aligned}$$



谢 谢！