

概率论与数理统计

总体和样本

主讲人：郑旭玲



信息科学与技术学院



01

什么是 数理统计

一、什么是数理统计

- 概率论

已知随机变量的分布信息



$P\{X \leq a\}$ 、 $F(x)$ 、 $f(x)$ 、 $P \cdot j$ 、 EX 、 $DX \cdots \cdots$

- 数理统计

未知随机变量的分布



反复试验得到的观察值 x_1 、 x_2 、 \cdots 、 x_n

数理统计学是一门应用性很强的学科。

它使用概率论和数学的方法，研究怎样以有效的方式收集、整理和分析（称为**统计分析**）带有随机性的数据，以便对研究对象的性质、特点作出推断（称为**统计推断**）。

一、什么是数理统计

- 某国14岁男孩的平均身高
- 大西洋中鳕鱼的平均重量
- 某工厂生产的一批灯泡的平均寿命
- 观看某一电视节目的观众人数
-

由于大量随机现象必然呈现出规律性，只要对随机现象进行足够多次观察，被研究的规律性一定能清楚地呈现出来。

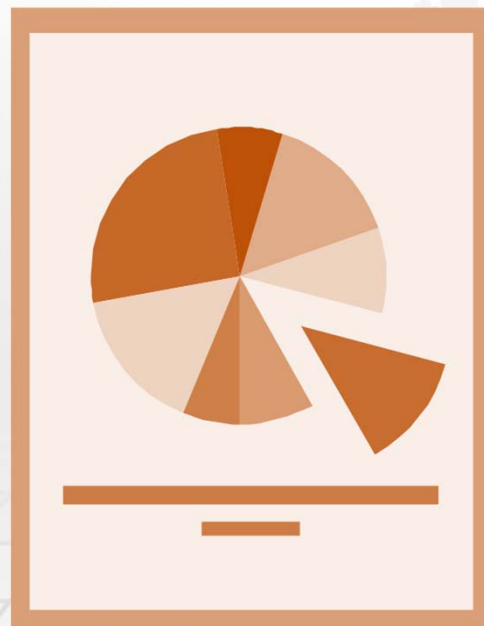
一、什么是数理统计

客观上，只允许我们对随机现象进行次数不多的观察试验，我们只能获得有限的局部观察资料。

在数理统计中，不是对所研究的对象全体（称为**总体**）进行观察，而是抽取其中的部分（称为**样本**）进行观察获得数据（**抽样**），并通过这些数据对总体进行推断。

一、什么是数理统计

数理统计方法具有“部分推断整体”的特征。





02

总 体

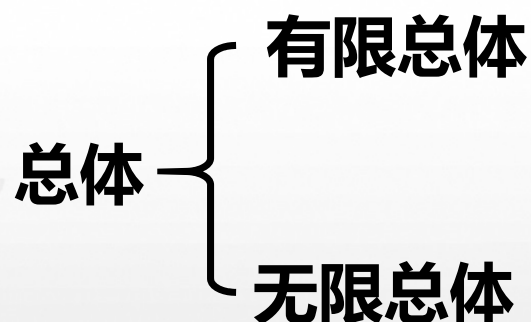
二、总体



定义

一个统计问题总有它明确的研究对象。

- 研究对象的全体称为**总体**；
- 总体中每个成员称为**个体**；
- 总体中所包含的个体的个数称为总体的**容量**。

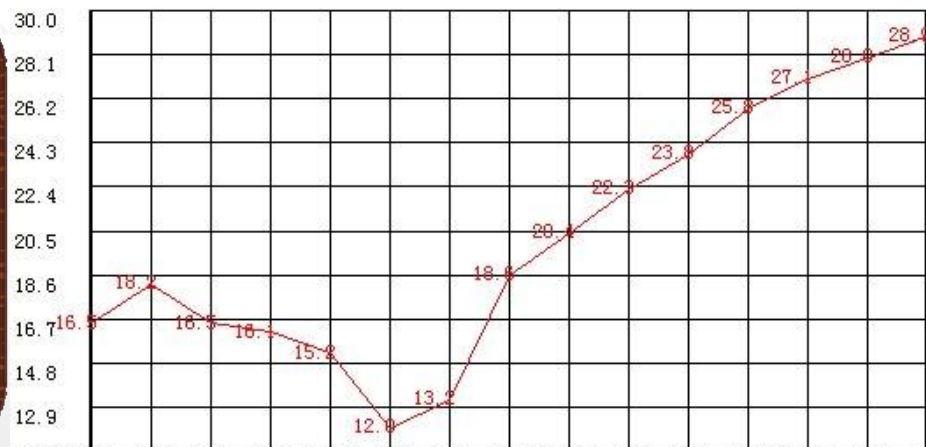
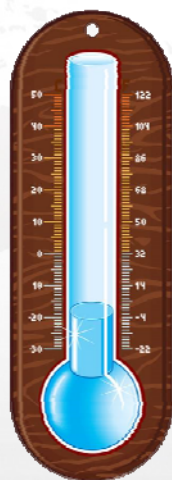


二、总体



...

研究某厂生产的一批灯泡的质量



观察并记录某一地点每天的最高气温

二、总体

有些有限总体，容量很大，也可以把它当成无限总体来处理。



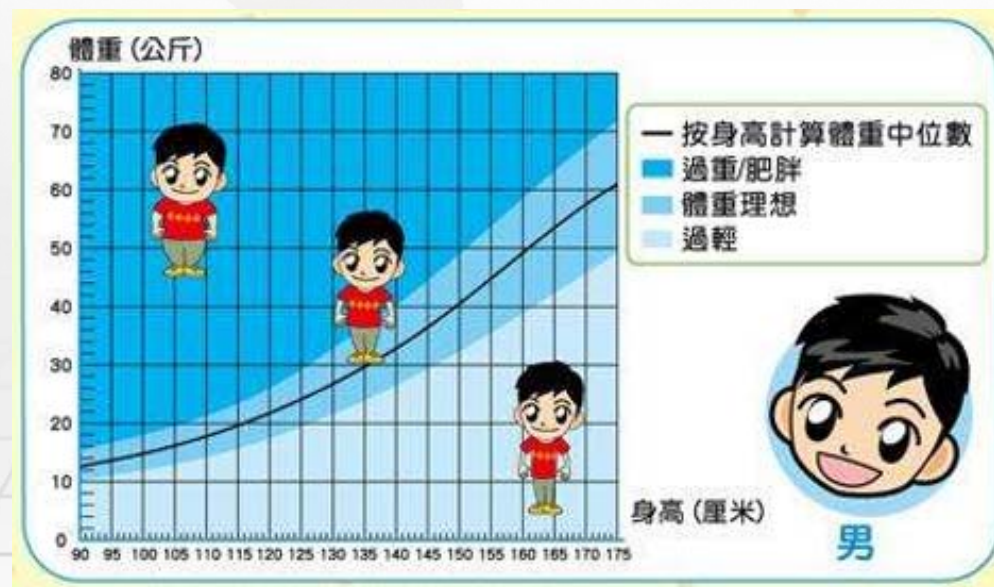
考察大西洋中鳕鱼的重量

总体容量取决于研究目的，也受人力、物理、时间等因素的限制。



二、总体

在数理统计中，往往注重考察研究对象的某一项或几项数量指标。





二、总体

该数量指标的所有可能取值
就是**总体**；
该数量指标的每个观察值
就是**个体**。

数量指标是一个随机变量 X ，
则总体就对应于这个随机变
量 X 。
 X 的分布和数字特征就是该
总体的分布和数字特征。

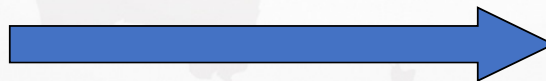
二、总体

例如，研究某批灯泡的寿命时，关心的数量指标就是寿命



总体

寿命 X 可用
指数分布来刻画

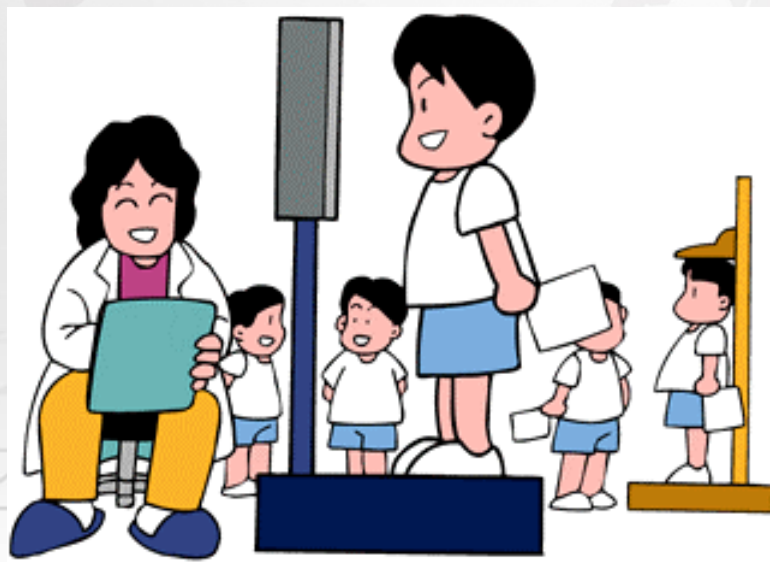


灯泡寿命
总体
是指数分布
总体

通常，用随机变量的记号或其分布函数表示总体，
即 总体 X 或 总体 $F(x)$ 。

二、总体

又如，研究某地区中学生营养状况时，
若关心的数量指标是身高和体重，分别用 X 和 Y 表示，
那么，此总体可表示为二维随机变量 (X, Y)
或联合分布函数 $F(x, y)$ 。





二、总体

数理统计中，总体这个基本概念的要旨：**总体就是一个概率分布。**

总体分布一般是未知，或只知道近似服从某种分布，但包含未知参数。



03

样 本

三、样本



定义

为推断总体分布及各种特征，按一定规则从总体中抽取若干个体进行观察试验，以获得有关总体的信息，这一抽取过程称为“**抽样**”。

被抽取的部分个体称为总体的一个**样本**。

样本中所包含的个体数目称为**样本容量**。

三、样本

例如，从某批灯泡中抽5只进行
寿命试验



样本容量为5

抽到哪5只是随机的

从总体中抽取样本容量为 n 的样本，
即得到 n 个随机变量，
按试验顺序记为 X_1, X_2, \dots, X_n

三、样本



简单随机抽样的特点：

1. 代表性：

X_1, X_2, \dots, X_n 中每一个与所考察的总体有相同的分布。

2. 独立性：

X_1, X_2, \dots, X_n 是相互独立的随机变量。

由简单随机抽样得到的样本称为**简单随机样本**，可以用与总体**独立同分布**的 n 个相互独立的随机变量 X_1, X_2, \dots, X_n 表示。除特别说明外，通常提到的样本都是指简单随机样本。



三、样本

取定一组样本 X_1, X_2, \dots, X_n ，分别对它们进行观察，就得到 n 个具体的数 x_1, x_2, \dots, x_n ，称为样本 X_1, X_2, \dots, X_n 的**观察值**，简称**样本值**。

三、样本

若 总体的分布函数为 $F(x)$ 、概率密度函数为 $f(x)$,

X_1, X_2, \dots, X_n 是 F 的一个样本 ,

则 X_1, X_2, \dots, X_n 相互独立且同分布 , 该样本的联合分布函数为:

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

该样本的联合概率密度函数为:

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

三、样本

实际操作中，怎样获得简单随机样本？

放回



不放回



三、样本

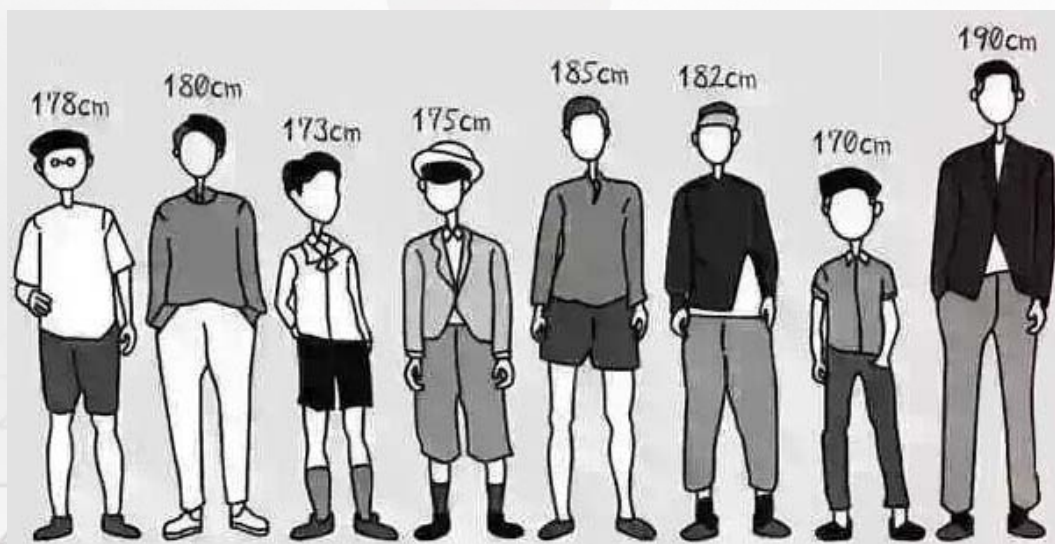
- 对于有限总体，采用放回抽样，就能得到简单随机样本。
 - 当个体的总数 N 比样本容量 n 大得多时，实际中可将不放回抽样近似地当做放回抽样来处理
- 对于无限总体，总是用不放回抽样。

04

总体、样本、 样本值的关系

➤ 四、总体、样本、样本值的关系

事实上，抽样后得到的是具体的样本值，而不是样本，即只能观察到随机变量取的值，而见不到随机变量。



➤ 四、总体、样本、样本值的关系



数理统计是从手中已有的资料——样本值，去推断总体的情况——总体分布 $F(x)$ 的性质。样本是联系二者的桥梁

总体分布决定了样本取值的概率规律，也就是样本取到样本值的规律，因而可以由样本值去推断总体。



小结



数理统计



总体和样本（简单随机样本）



总体、样本、样本值的关系

The background of the slide is divided into three horizontal sections. The top section is white with a faint, light gray world map. The middle section is a solid dark red color, featuring a faint, darker red world map. The bottom section is white with a light gray grid pattern that recedes into the distance.

谢 谢 大 家