

第1章 绪论

§ 1.1 统计学的产生与发展

§ 1.2 统计学的研究对象和统计工作

§ 1.3 几个重要的基本概念

§ 1.1 统计学的产生与发展

- 一、统计学是统计实践活动的理论概括

- 二、各种统计学派简介

- 1. 国势学派

代表人物：H.康令 特点：实不符名

- 2. 政治算术学派

代表人物：威廉.配弟 特点：名不符实

- 3. 数理统计学派

以现代数学为工具，侧重统计学理论研究。

- 4. 贝叶斯统计学派

数理统计学的一个分支，强调先验信息的利用与研究

■ 三、统计学在中国

有半个世纪的争论，现强调“大统计”思想
描述统计与推断统计的有机结合

■ 四、统计与计算机

相辅相成，互相促进，共同发展

§ 1.2 统计学的研究对象和统计工作

■ 一、什么是统计学

统计学的涵义：统计工作、统计资料和统计学

- 统计工作是人们的统计实践，是主观反映客观的认识过程；
- 统计资料是统计工作的结果，统计工作与统计资料是过程与成果的关系。
- 统计学是统计工作经验的总结和概括；反之，统计学所阐述的理论和方法又是指导统计工作的原则和方法。
- 统计学和统计工作存在着理论和实践的依存关系。

统计学的定义：

统计学是门科学，它研究怎样以有效的方式，收集、整理、分析带随机性的数据，并在此基础上对所研究的问题作出统计性推断，直至对可作出的决策提供依据或建议

- 注意：** 统计学在中国半个多世纪的争论，
- 可以说其焦点是统计学定义的争论。

■ 二、统计工作

- 统计工作的第一阶段又称**统计调查**，统计调查就是根据一定的目的，通过科学有效的调查方法，搜集自然现象、社会现象的实际资料的活动，主要有统计调查方案的设计、实施等。
- **统计整理**是对调查来的大量统计资料加工、整理、汇总、列表的过程。它是统计工作过程的第二阶段，也即处于统计工作的中间环节，起着承前启后的作用。

统计分析是统计工作的第三阶段，它是将加工整理好的统计资料加以分析研究，采用各种分析方法，计算各种分析指标，来揭示事物的本质及其发展变化的规律性，做出定量的**有概率**依据的统计推断。

- **注意：**统计工作是围绕数据进行，故统计学又称数据的科学；

统计工作是有机整体，往往不能一蹴而就，有时需要反复多次。

■ 三、统计学的研究领域

只要有数据、信息需要处理的地方，就有统计学的用武之地。

- 1.空间与天文
- 2.地理与地质
- 3.医疗与卫生
- 4.生物与生命科学
- 5.经济与金融
- 6.社会与政治

§ 1.3 几个重要的基本概念

■ 一、统计总体、个体和样本

1. 统计总体

- 统计学中的总体，可以理解为有两层含义，
或指研究对象的全体，
或指研究对象一些指标对应随机变量的取值全体，总体的这种理解，也称为总体 X 和总体的精确定义。
总体的分布，也就是 X 的分布

■ 2、个体

个体也有两种含义，一是具体的研究对象即总体中的一个单元;其二是经概括抽象的，是与总体同分布的随机变量 X 或 X 的取值.

3、样本

- 是总体中抽取的 n 个个体的集合， n 称为样本容量，也称样本的大小。样本也有相应的两层含义。
- 设 X_1, X_2, \dots, X_n 是总体 X 的样本，如果 X_1, X_2, \dots, X_n 相互独立,且每个 X_i 都与总体有相同分布的，就称为总体的简单随机样本，简称简单样本或样本。

注意：要能根据上下文区别样本和样本观测值。

■ **例1.4** 为了解我国家庭年人均消费和年人均收入之间的关系， 从全国随机抽取5000个家庭作抽样调查， 从调查资料中整理出相应的人均消费与人均收入的数据。

(1) 从两种含义给出研究问题的总体及其大小；

(2) 给出两种含义下相应的样本及其大小；

解： (1)总体含义一是我国全部的家庭， 总体大小 $N = +\infty$ ；

- 设横坐标 X 为家庭年人均收，纵坐标 Y 为家庭年人均消费，总体含义二就是 (X, Y) 可能的取值全体，也即直角坐标平面第一象限的全部点，

(2) 样本含义一是被调查的5000个家庭，
样本大小 $n = 5000$

样本含义二可以这样理解，设 (x_i, y_i) 为被调查的第 i 个家庭的的年人均收与年人均消费，那么， $(x_1, y_1), \dots, (x_n, y_n)$ ， $n = 5000$ ，或表示5000个二维随机变量，或表示直角坐标平面第一象限的5000个点(如果得到了这些家庭的具体收入与消费的数值)。

二、统计量

■ 样本 X_1, X_2, \dots, X_n 的不含未知参数的函数称为统计量，统计量是特殊的随机变量。

■ 例如，设总体 $X \sim N(\mu, 4)$ ， (X_1, X_2, \dots, X_n) 是其样本，则

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \sum (X_i - \bar{X})^2$$

是统计量，前者称为样本均值，后者称为样本的离差平方和。而

$$\sum (X_i - \mu)^2$$

就不是统计量，因为它含有未知参数 μ 。

统计学中，还有一个重要统计量，是与离差平方和有关的样本方差（variance）

$$s^2 = \frac{1}{n} \sum (X_i - \bar{X})^2 f_i = \frac{1}{n} \sum X_i^2 f_i - \bar{X}^2$$

本教程把

$$\tilde{s}^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 f_i$$

- 称为修正的样本方差。