

前言

随着数据时代及人工智能时代的到来，数据越来越多、越来越重要。数据已成为生产要素，也是一种资源。作为专门研究数据的学科——统计学的春天来了。

大家对“统计”一词并不陌生，在实际生活中也经常用到，如班长“统计”一下明天春游的人数，足球比赛结束后的技术“统计”等，但这都只是简单的汇总，并不需要统计知识。高中时所学课程里就有统计模块，其中最熟知的直方图不仅是数据的简单汇总，而且在画直方图时，首先要考虑区间的划分，因为它影响到直方图的形状；直方图画好后，还要对此进行解释，以了解数据所包含的信息与知识。

按照《不列颠百科全书》的定义：统计是研究数据的科学与艺术。故可知凡是研究数据包含的内在规律，由数据产生价值、产生知识，以及企事业单位或政府各部门基于数据制定相关政策与规划等，都离不开统计。正如我国著名教育学家马寅初（1882—1982）所说：学者不能离开统计而研究，政治家不能离开统计而执政，企业家不能离开统计而执业，军事家不能离开统计而谋划。世界著名统计学家 C. R. Rao（拉奥，1920—2023）曾在其《统计与真理》一书前言中写道：在理性的基础上，一切判断都是统计学（All judgements are, in their rationale, statistics）。后来他本人又说道：所有获取知识的方法本质上都是统计学（All methods of acquiring knowledge are essentially statistics）。也正如 D. Salsburg（萨尔斯伯格）博士撰写的一本统计科普读物《女士品茶》（*The Lady Tasting Tea*）的副标题：20 世纪统计怎样变革了科学（How Statistics Revolutionized Science in the Twentieth Century）。

统计方法的产生与应用有着悠久的历史。如果从数据分析角度看，J. Graunt（格兰特，1620—1674）于 1662 年撰写的《关于死亡公报的自然与政治观察》一书，对 1604—1662 年间伦敦教会每周一期的“死亡公报”数据进行了统计初步分析；如果从随机模拟角度看，历史上最早的应该是 G. Buffon（比丰，1707—1788）于 1777 年进行的“Buffon 投针”；如果从统计方法角度看，可以追溯到 1805 年由 C. F. Gauss（高斯，1777—1855）提出的最小二乘法，以及 1802 年由 P.-S. Laplace（拉普拉斯，1749—1827）在法国人口抽样调查时提出的等概率估计法；如果从统计理论角度看，20 世纪初由 K. Pearson（皮尔逊，1857—1936）和

R. A. Fisher(费希尔, 1890—1962) 提出的系列理论与方法, 奠定了现代统计学基础与框架。

撰写本书的目的在于把统计学基础理论与方法介绍给理科统计学专业本科生, 并要求学生已掌握数学分析、概率论的基础知识。第一章介绍一些基本概念, 如样本、参数、统计量、抽样分布, 以及一些常用的抽样分布, 如 χ^2 分布、 t 分布、 F 分布、 Γ 分布, 还有充分与完全统计量等; 第二章讲述参数点估计准则与方法, 如无偏估计、一致最小方差无偏估计, 相合估计、相合渐近正态估计、最大似然估计, 还有 Bayes(贝叶斯) 估计、minimax 估计等; 第三章讲述统计推断中的假设检验理论, 内容包括显著性检验思想、 p 值、似然比检验等, 进一步介绍置信区间的概念及求取方法等。第四章讲述多元模型的统计推断理论, 如常用多元分布、多元正态总体的参数估计和假设检验, 也对多重检验的相关准则和方法进行介绍。第五章介绍最小二乘估计的性质以及线性模型的估计、检验、模型选择, 并讨论 logistic 回归模型。第六章给出总体密度、总体分布和分位数的非参数估计方法, 以及针对总体分布、独立性、均值的非参数检验方法。第七章介绍 bootstrap 方法在方差估计、置信区间构造、线性模型中的应用。附录部分介绍概率收敛的四种定义、中心极限定理、Slutsky(斯卢茨基) 定理、Delta 方法等内容。

在本书编写过程中, 参考了多位老师撰写的相关书籍, 这些书籍为我们提供了极其丰富的材料, 非常感谢他们; “101 计划” 统计工作小组在邵启满教授的带领下, 设计了本书的基本框架, 非常感谢他们的顶层设计及对南开统计的信任; 最后要感谢高等教育出版社的编辑, 从本书撰写之初就对本书给予了极大的关心与帮助。

由于编者无论是专业还是文字水平所限, 编写之中难免会有不妥之处, 敬请读者批评指正。

编 者

2024 年 2 月于南开园