中国统计科学学报
Journal of Chinese Statistical Sciences
2023 年第 01 卷第 02 期
Vol. 01, No. 02, 2023

统计入门《统计思维:科学家入门导引》——书评

徐 慧

(长江大学地球科学学院,武汉 430199)

摘要:为了培养读者在面对数据时运用统计思维进行科学分析,在数据分析的基础之上做出更加科学合理的判断,作者 M. D. Edge 撰写了著作《统计思维:科学家入门导引》,其中包含了三个部分共十个章节的内容。主要涉及数据和概率等一些基础的概念,概率、随机变量以及随机变量的性质等概率论方面的基本理论知识,为后续更高级的统计学理论内容打下坚实的基础,重点讲述了简单线性回归的统计方法。阐述了点估计量的性质和区级估计与推断,分析了半参数估计与推断、参数估计与推断、贝叶斯估计与推断三种广泛使用的方法。

该著作的特点在于为统计学家和数据分析师的培养提供了概念框架,系统地讲述统计思维的理论框架,通过数据、随机变量和一些估计方法介绍几种重要的统计方法和思维,为了解统计方法是如何设计的提供了思路和方向,对读者已经知道如何应用技术,则增加其深度,激发读者的学习兴趣。

关键词:数据分析;统计思维;估计和推断;简单线性回归;书评

DOI:10.48014/jcss.20240402004

引用格式:徐慧. 统计入门《统计思维:科学家入门导引》——书评[J]. 中国统计科学学报,2023,1(2):9-14.

0 引言

《统计思维:科学家入门导引》[1]是由南加州大学定量和计算生物学系助理教授 M. D. Edge 编写的系统讲述统计思维的统计学入门的书籍。该书由上海财经大学数学学院冉启康教授翻译,由中国大型综合性专业出版社机械工业出版社出版。

统计学是一门收集、处理、分析、解释数据并从数据中得出结论的科学,它提供了一套各学科领域通用的数据分析方法,统计研究的是来自各领域的数据[2]。统计方法是适用于所有学科领域的通用的数据分析方法,只要有数据的地方就会用到统计方法[3]。目前,统计学在金融、经济、企业管理、工农业生产、军事、医疗、气象与自然灾害预报等方面都有重要的作用。不论是在工作、学习、健康、还是生活的方方面面,统计学都提供了有价值的信息,帮助我们理解和解读数据,从而更好地应对生活中的各种情况和挑战。

统计思维是指人们在认识客观事物的过程中,自觉运用数学和统计学的理论和方法,对客观事物和现象的数量特征及数量关系进行正确的描述和科学的分析,从而认识其本质,把握其发展变化的规律性的高层次的结合型思维方式[4]。统计思维是一种科学的用途广泛的思维方式,加强对统计学的认识,努力培养统计的思维方式,对于面对更加信息化的海量信息和更复杂的问题的时候不会束手无策,反而更加游刃有余,从而做出更加理性和科学的决策[5]。从统计的视角去认识问题时,要用统计思维去考虑问题、解决问题、指导决策。统计学中最基本的思维方式就是“调查研究”“实事求是”

通讯作者 Corresponding author:徐慧,1973195172@qq.com

收稿日期:2023-06-02; **录用日期**:2023-08-09; **发表日期**:2023-09-28

和“用数据说话”[6]。随着互联网快速发展，网络信息数据越来越丰富，大数据时代数据信息的规模巨大、结构更加复杂，所以现今统计思维的转变需要更加强大的数据挖掘、数据分析能力[7]。

因为传统的统计学有一个前提，读者在学习统计学之前必须掌握一定的大学数学的知识，比如微积分、概率论、线性代数，这就使得没有系统地学习过这些课程的读者望而却步。而《统计思维：科学家入门导引》一书不要求读者对微积分、线性代数和概率论有系统的理解，只需要掌握微积分的基本思想就足矣。本书从简单的线性回归开始，带领读者从零基础学习一种方法，建立估计和推断的整个理论框架，使读者了解统计方法是如何设计的，对在研究中应用统计学时隐含的哲学思维有所了解，对统计技术的优势和劣势有更清晰的认识。

1　内容解读

为了使每个实证学科的从业者学会数据分析，成为自信的数据分析师，作者通过数据、随机变量和一些估计方法介绍了几种重要的统计方法和思维，提供了统计学学生培养的基本框架，撰写了著作《统计思维：科学家入门导引》。该著作的前五章是基础，讲述数据、概率与随机变量，“插叙”之后转向更高级的理论——估计与推断，最后三章分别介绍了半参数估计与推断、参数估计与推断以及贝叶斯估计与推断三种广泛使用的方法。下面将对每个部分进行解读。

第一部分为基础，包括第 1 章到第 5 章。第 1 章介绍了一些富有启发性的问题，以及应该如何建立从数据进行推理的概念，提出了统计是利用数学思维来对数据进行推理的方法，并且引出了要重点学习的方法——简单线性回归。第 2 章主要介绍了关于统计软件 R 的部分功能，包括用 R 软件对鸢尾花数据集进行一些探索性数据分析，来模拟数据、分析数据和进行数据展示。第 3 章主要介绍了用直线概括数据的方法，提出了最佳拟合线的定义，阐述了最常用的数据汇总线是最小二乘线以及最小二乘线截距和斜率表达式的推导。第 4 章介绍了概率论的一些法则：概率是非负的；样本空间的概率为 1；两个互斥事件中至少一个发生的概率是两个事件发生的概率之和，介绍了条件概率、独立性、贝叶斯定理的定义，并分析了对于离散随机变量使用概率质量函数，对于连续随机变量使用概率密度函数。第 5 章首先用期望、方差和协方差的概念描述了随机变量的性质，并且提出了大数定律和中心极限定理两个重要的定理来刻画样本均值的分布，最后建立了一个描述两个随机变量之间的线性关系模型，并分析了这两个随机变量的性质。

第二部分为估计与推断，包括第 6 章和第 7 章。第 6 章阐述了点估计是试图使用抽样数据去确认与某些潜在过程或总体相关的参数值，以及点估计量的性质：偏差、方差、均方误差、有效性、稳健性。第 7 章解释了区间估计是尝试定义一个区间去量化点估计中的不确定性程度，并介绍了标准误差、置信区间、频率推断、备择假设和拒绝框架，阐述了假设检验和置信区间的关系以及零假设显著性检验及检验的滥用。

第三部分为估计与推断的三种广泛使用的方法，包括第 8 章，第 9 章以及第 10 章。第 8 章主要阐述了当模型只是部分参数受支配而不是完全受参数支配时使用半参数估计与推断，以及非参数和半参数方法的核心是经验分布函数，介绍了半参数点估计的矩方法，如何使用 bootstrap 进行半参数区间估计和使用 bootstrap 方法从经验分布函数中重抽样，最后介绍了使用置换检验的半参数假设检验。第 9 章阐述了当模型受参数支配的情况下使用参数估计与推断，介绍了参数估计最常用的方法——极大似然估计法以及简单线性回归模型的极大似然估计，使用直接方法和费希尔信息方法进行参数的区间估计，讨论了瓦尔德检验和似然比检验进行参数假设检验。第 10 章介绍了贝叶斯估计与推断，阐述了贝叶斯统计的主要研究对象时后验分布，介绍了如何选择一个先验分布，如何从后验分布中抽样，如何使用贝叶斯估计方法获得点估计量，如何使用可信区间进行贝叶斯区间估计，以及如何使用贝叶斯因子进行贝叶斯“假设检验”。

第四部分为附录、术语表和部分练习题的参考答案，主要介绍了该著作需要的基本数学知识，对第 2 章 R 语言进行拓展，列举了该著作中的专业术语。其中，附录 A 中的微积分的基本知识对于读者深入理解统计学中的概念和解决实际问题具有重

要的意义,为统计学提供了坚实的数学支撑和理论基础。而附录B是本书第2章关于R语言在一些功能介绍基础上的扩展,更加详细地介绍了R语言中可用的数据类型以及一些非常重要的参数,R语言提供了强大的数据处理、数据分析以及可视化等功能,对于统计学者进行更加高效、准确的统计分析具有十分重要的意义。附录C提供了部分练习的答案,通过对照练习的答案能够帮助读者深化对统计学概念和方法的理解,尤其是当读者在学习的过程中遇到困难时,参考答案能够提供正确的思路和解释,这有助于读者学会如何解决类似的问题,读者可以更好地将理论知识应用于实际数据分析中,提高解决实际问题的能力。附录最后的数学符号表和术语表提供了部分数学符号和统计术语的解释,帮助读者迅速理解这些符号和术语的含义,不仅方便读者进行查阅和学习,也有助于提升整体的学习体验和效率。

2 研究发现和主要贡献

《统计思维:科学家入门导引》是一部致力于培养科学领域中统计思维的书籍。在当今大数据驱动的研究环境中,统计学不仅是数据分析的工具,更是一种科学的思考方式。作者以其自身的学习和教学经验为基础,打造了这本旨在帮助初学者建立统计思维的指南。该书的研究发现和主要贡献如下:

(1)提供了统计学学习的基本框架。重点解读一种统计方法,即简单线性回归,聚焦于最关键的统计学基础的概念和方法论。该书前五章以数据的统计描述、R软件交互、概率论基础和随机变量等核心概念为起点,为读者打开统计学的大门。在此基础之上进一步介绍估计与推断的理论,包括半参数估计与推断、参数估计与推断和贝叶斯估计与推断等高级的统计学理论。书籍内容从最基础的简单线性回归方法开始,逐步深入,使得即使是零基础的读者也能够跟随学习并掌握统计方法。这种由浅入深的结构设计为读者提供了一个清晰的学习方向,有助于建立完整的统计知识体系。

(2)结合了实际案例和R软件教学。本书通过将理论与实际相结合,每个章节都提供了丰富的练习题,这些问题穿插在本书的正文中,是论述的一部分,提供了实践、关键原理的证明、重要的推论。这种互动的学习方式不仅能够提高学者的参与度,还能帮助学者们在实践中深化对统计学概念的理解。此外,书中还详细地介绍了R软件的部分功能、R软件中可用的数据类型、一些非常重要的函数以及R软件的编程实例和丰富的编程练习。R软件不仅提供了强大的数据处理分析能力、统计分析功能、图形绘制功能,还提供了大量的统计学习方法和算法。R软件作为统计学的一个重要的工具,其功能之大,应用范围广泛,对于统计研究和实际应用都有着深远的影响。不论是学者还是从事统计研究工作的专业人员,学习和掌握R软件都是提高统计分析能力的重要途径。

(3)分析了在面对不同的统计问题时,如何选择合适的统计分析方法。非参数和半参数统计方法假定模型的性质不能由有限数量的参数来刻画。非参数和半参数方法的核心是经验分布函数,当数据是来自同一分布的独立观测值的情况下,经验分布函数是真实累积分布函数的一致估计,使用非参数和半参数方法。为了研究点估计的抽样分布,使用bootstrap方法从经验函数中重抽样。对于假设检验,可以使用基于bootstrap方法的置信区间,也可以进行置换检验,置换检验可以用来进行独立性检验。如果假设数据是由完全参数模型生成的是合理的,则使用极大似然法来估计和推断,极大似然估计是一种在参数估计与推断中常用的方法,极大似然估计量是通过确定参数使似然函数最大化而得到的,可以通过微积分或数值方法来求得。当面对不确定性和复杂性较高的问题时,贝叶斯方法能够提供一种系统的框架来更新和表达参数的不确定性。贝叶斯统计的主要研究对象是后验分布,即在给定观测数值条件下估计得到参数分布,描述了给定先验分布和观测数据条件下相关参数的不确定性。贝叶斯点估计和区间估计是后验估计的特征,它们度量了后验分布的中心趋势和参数以特定概率落入的区间。贝叶斯假设检验的一个工具是贝叶斯因子,它比较了在一对不同假设下观测数据的概率。

(4)《统计思维:科学家入门导引》这本著作强调了统计思维的重要性,从简单的线性回归开始,

从零基础出发学习一种方法,用数学、模拟、思想实验和例子的组合来探索尽可能多的基本问题,建立估计和推断的整个理论框架。这本著作的研究成果具有以下几个方面的重要意义和贡献。首先,该著作的重点是只说明一种统计方法,即简单线性回归。通过从零开始学习一种方法,考虑在这一背景下的估计和推断的整个概念框架,获得适用于其他背景的工具、理解和直觉。其次,该著作要求的数学水平较低,统计学中许多丰富的思想都是用数学来表达的,所以需要对微积分的主要思想有一定了解。其中,微积分不仅为统计学提供了一套严密的理论体系,还为数据分析、模型建立和理论推断提供了坚实的数学基础。此外,该著作提供了许多关于 R 软件的练习。通过实际的案例来学习如何进行探索性分析,帮助读者理解数据分析的过程与步骤,以及如何从数据中提取有价值的信息,不仅提供了理论知识,还提供了实践操作的机会,有助于读者更好地掌握统计学的应用。总的来说,该著作的主要贡献在于重点解读了一种统计方法,即简单线性回归,以及使人们了解统计方法是如何设计的。其特色和创新之处在于对数学的要求非常低,并且没有全面介绍统计学,甚至不要求读者对微积分、线性代数和概率论有系统的理解,而是通过数据、随机变量和一些估计方法介绍几种重要的统计方法和思维,提供了统计学学生培养的基本框架,对于统计学学者具有重要的参考和指导价值。

3 与同类书的比较评论

在统计学书籍的海洋中,《统计思维:科学家入门导引》无疑是一颗璀璨的明珠。这本书以其独特的视角、深入浅出的讲解并且结合了大量实用的练习,赢得了众多读者的喜爱。然而,将其与同类书籍进行比较时,又会有哪些独特之处呢?以下是《统计思维:科学家入门导引》与部分同类书籍之间的比较评论。

David A. Freedman 撰写的《统计模型理论和实践》[8],这本书以线性模型为核心,深入讲解了广义最小二乘、两步最小二乘模型以及二分变量的 probit 和 logit 模型等应用。同时,还包括关于研究设计、二分变量回归及矩阵代数的背景知识,为读者提供了更加全面的统计模型知识体系。所以《统计模型理论和实践》这本书适合统计学高年级本科生、研究生以及相关领域的研究人员,可以帮助他们深入理解和掌握统计模型的应用技巧和方法。而《统计思维:科学家入门导引》更注重于统计学的基础思维和入门指导。这本书没有系统地介绍统计学的内容,而是通过数据、随机变量和一些估计方法介绍几种重要的统计方法和思维,提供了统计学学生培养的基本框架,更适合统计学初学者。

田霞撰写的《统计学入门很简单:日常生活中的统计学》[9]这本书注重统计学在生活中的实际应用,介绍了概率的基础、统计的基础和统计的进阶,每章在介绍过预备知识后,都结合相应的统计知识给出了具体案例,指导读者一步步进行运算,最后得出结论。而《统计思维:科学家入门导引》这本书通过数据、随机变量和一些估计方法介绍了几种重要的统计方法和思维,旨在提供一个统计学学生培养的基本框架,帮助初学者建立统计思维。

威廉·M·门登霍尔撰写的《统计学(原书第 6 版)》[10]详细介绍了统计学的基本理论、方法和应用,内容涵盖了描述性统计、概率论、推断统计、回归分析、方差分析、非参数统计等多个方面。这本书采用了较为专业的语言和表述方式,对统计学的理论和方法进行了深入的探讨和解释,是一本更为全面和系统的统计学教材。而《统计思维:科学家入门导引》并没有全面和系统地介绍统计学,而是重点说明了一种统计方法,即简单线性回归,帮助读者用数学、模拟、思想实验和例子的组合来探索尽可能多的基本问题,建立估计和推断的整个理论框架。

罗伯特·古尔德等撰写的《统计学基础:透过数据看世界》[11]是一本以数据分析为基础的面向统计学、数据科学的统计学著作,通过大量的案例、示例、图、表等,让读者能够直观地理解统计学的原理和方法。而《统计思维:科学家入门导引》结合了实际案例和 R 软件教学,利用 R 语言进行一些探索性数据分析,来模拟数据、分析数据和进行可视化展示。

涌井良幸等撰写的《统计学入门很简单 看得懂的极简统计学》[12]围绕搜集数据、分析数据以及得

出统计结论这一主线，全面、系统地讲解了有关统计学的基础知识。内容涉及统计原理、基本方法及发展应用，以及参数估计、假设检验、线性回归、实际应用等几个方面的有关知识。而《统计思维：科学家入门导引》结合了实际案例，通过将理论与实际相结合，提供了一个系统学习的框架，有助于读者建立起对统计学的整体认识和基本框架。

4 结束语

《统计思维：科学家入门导引》作为一本统计学入门的书籍，没有全面介绍统计学，而是通过数据、随机变量和一些常用的估计和推断方法介绍了几种重要的统计方法和思维，提供了学习统计学的概念框架。该著作为学者学好统计学，利用统计思维去考虑问题并解决实际问题提供了重要的方法支撑和理论基础。同时，该著作不仅是一个引导初学者踏入统计世界的宝贵指南，也是那些希望提升和巩固自己统计分析技能的资深人士的实用手册。此外，尽管书中的一些统计方法可能需要读者具备一定的数学基础，但该书作者提供了充足的解释，有助于读者逐步理解并掌握复杂的统计学工具，进一步促进读者对统计思维的理解和运用。

利益冲突：作者声明无利益冲突。

参考文献(References)

[1] [美]M. D. 埃奇(M. D. Edge). 统计思维：科学家入门导引[M]. 冉启康，译. 北京：机械工业出版社，2023.

[2] 贾俊平. 统计学[M]. 北京：中国人民大学出版社，2021.

[3] 林侠. 统计学原理与实务[M]. 北京：中清华大学出版社. 2020.

[4] 崔青云. 论统计思维及培养[J]. 山西煤炭管理干部学院学报，2009(3)：2.

[5] 张林云. 统计思维的特点与培养初探[J]. 统计与管理，2014(06)：14-15.
https://doi.org/10.3969/j.issn.1674-537X.2014.06.002

[6] 郭思亮，宋晓蕾. 基于统计思维的地摊经济时空属性分析[J]. 齐鲁师范学院学报，2020，35(5)：80-85.
https://doi.org/10.3969/j.issn.1008-2816.2020.05.013

[7] 郭鑫雨. 大数据时代下统计思维研究[J]. 数字通信世界，2018(2)：230-231.
https://doi.org/10.3969/J.ISSN.1672-7274.2018.02.186

[8] [美]David A. Freedman. 统计模型理论和实践[M]. 吴喜之，译. 北京：机械工业出版社，2020.

[9] 田霞. 统计学入门很简单：日常生活中的统计学[M]. 北京：中国纺织出版社，2023.

[10] [美]威廉·M·门登霍尔(William M. Mendenhall)，特里·L辛西奇(Terry L. Sincich). 统计学(原书第6版)[M]. 关静，等，译. 北京：机械工业出版社，2018.

[11] [美]罗伯特·古尔德(Robert Gould)，丽贝卡·王(Rebecca Wong)，科琳·莱恩(Colleen Ryan). 统计学基础：透过数据看世界[M]. 田金方，译. 北京：机械工业出版社，2023.

[12] [日]涌井良幸，涌井贞美. 统计学入门很简单 看得懂的极简统计学[M]. 刘楚姮，译. 北京：人民邮电出版社，2019.

Statistical Thinking from *Scratch*: *An Introductory Guide for Scientists*-Book Review

XU Hui

(School of Geosciences, Yangtze University, Wuhan 430100, China)

Abstract: In order to train readers to make scientific analysis with statistical thinking in the face of data and to make more scientific and reasonable judgments based on data analysis, the author M. D. Edge has written the book Statistical Thinking from Scratch: A Primer for Scientists, which contains ten chapters in three parts. It mainly involves some basic concepts such as data and probability, basic theoretical knowledge of probability theory such as probability, random variables and the properties of random variables, which lays a solid foundation for the subsequent more advanced statistical theory content, with an emphasis on the statistical method of simple linear regression. The properties of point estimators and regional estimation and inference are described, and three widely used methods of semi-parameter estimation and inference, parameter estimation and inference, Bayes estimation and inference are analyzed. The feature of this work is that it provides a conceptual framework for the training of statisticians and data analysts, systematically describes the theoretical framework of statistical thinking, introduces several important statistical methods and thinking through data, random variables and some estimation methods, provides ideas and directions for understanding how statistical methods are designed, and adds depth to the techniques that readers already know how to apply, stimulate the reader's interest in learning.

Keywords: Data analysis; statistical thinking; estimation and inference; simple linear regression; book review
DOI: 10. 48014/jcss. 20240402004
Citation: XU Hui. Statistical thinking from *Scratch*: *An Introductory Guide for Scientists*-book review[J]. Journal of Chinese Statistical Sciences, 2023, 1(2): 9-14.