

师大量诞生时的情况一样,需要统计学者对所认识客体科学领域的理论有相当的认识。

乘数据科学春风,播撒“数据+”种子

董 麓 天津财经大学

回顾半个世纪以来统计学的蓬勃发展,至少有两个方面起到了决定性作用。一是计算机科学的迅猛发展,使得与数据和统计相关的技术发生了革命性变化;二是各应用领域(如生物医药、金融等)快速发展,并对使用统计方法分析解决自身问题产生了强烈需求。迄今为止,这两方面依然是推动统计学发展的重要力量。

“大数据”时代的来临,标志着一门新兴科学,即数据科学的发展进入了一个新的时代。我们应该清醒地认识到,“大数据”只是一个状态描述,而数据科学将是有长久生命力的科学。统计学是数据科学的重要组成部分,数据科学的发展对统计学提出了前所未有的挑战,也提供了空前的发展机遇。它期待统计理论研究的创新,催生统计应用的发展,呼唤统计教育的改革。

提到我国的数据科学与统计学发展,就不能不提到肖红叶教授。早在20世纪80年代中期,他就提出了“计算机+数据+经济学”的发展思路并付诸于实践。从1987年开始,他先后主持创建了“经济信息管理”、“数据分析”、“数据工程”等本科专业和方向,成为国内高校开展“数据+”教育的探路者。

在“大数据”时代的今天,“数据+”的思想已经广泛深入人心,但“数据+”的系统教育与专业人才培养却远远落后于时代的需要。从统计学高等教育的角度来看,现阶段的一个重要任务就是大力推进“数据+”教育,要把“数据+”的根深植于各个应用型学科领域,并在这些领域开花结果。

大数据时代统计学应拥抱数据科学

方匡南 厦门大学

21世纪是信息爆炸的时代,随着计算机技术的飞速发展,极大地便利了数据的采集和存储,各个部门每天都积累了大量的数据,比如商业银行交易记录、超市的销售记录、政府统计中各中小企业的财务报表等等。同时这些数据的维度也越来越高,比如研究基因与癌症的关系涉及的基因有几万个,信用评级中有上千个自变量等等。数据来源多样化,有业务记录数据,有传感器数据,也有第三方数据,甚至是从网上爬取来的数据等。此外,数据的格式也越来越多样化,有结构化数

据,也有文字、图片、音频、视频等非结构化数据。这些海量多维、来源多样化、格式多样化的大数据给传统的统计分析带来了极大的挑战。

数据科学(data science)是一门研究数据的科学,但这个词在计算机圈子里接受程度较高,而在统计学圈子里接受程度不是很高。大数据时代,统计学应适应时代的发展,主动拥抱数据科学,必须做到如下两点:

第一,统计研究应坚持以实际需求驱动发展。统计的发展跟实际需求是息息相关的,比如20世纪初,为了研究如何提高农作物产量,提出了农田实验设计方法;再比如20世纪60年代为了研究工业品的质量,提出了可靠性统计方法;而到了21世纪初,随着互联网技术的发展,积累了海量的数据,也需要更多统计学家积极研究互联网大数据分析的方法。

第二,统计研究需要主动结合计算机。对于大数据的研究,对研究人员的计算机技术要求也更高了,统计学家不能把这些都推给计算机学科去做,更不能只守着统计的一亩三分地,这样只会越做越小,不断被边缘化,而是要主动学习一些必要的计算机技术或者与计算机研究人员合作,不断扩大统计研究的范围。只有这样,在大数据时代统计学才能发挥更大的作用,迎来新的发展机遇!

统计模型应用面临的挑战与转型

李宝瑜 山西财经大学

统计量化测度和统计模型是统计学的核心,从统计学诞生之日起,其生命力就在于能够对现实进行量化测度和用统计模型分析实际问题。没有量化测度,经济统计学就不会相对独立于经济学。没有模型应用,数理统计学就还是数学,就不会从数学中分离出来。目前社会已经发展到网络和大数据时代,传统和现代的统计模型无论在方法上还是在应用中都已经受到了挑战。

在方法上,传统的统计模型都是建立在利用样本“估计”总体的思想上的,因为不了解总体,所以每个模型都要对总体进行“假设”,然后来检验样本。但在大数据和网络时代,很多总体的信息量都能全面获得,这就有必要思考:抽样调查还需要吗?抽样的重要性是否减弱了?总体还需要通过某个样本来“估计”吗?进一步,总体还需要“假设”检验吗?时代的改变对历经几百年之久的经典统计方法提出了生存和发展的挑战。