

第十七次全国统计科学讨论会文集

探索与创新

Discovery and Innovation

中国统计学会 编

16

中国统计出版社
China Statistics Press



探索与创新

Discovery and Innovation

中国统计学会 编

16

中国统计出版社
China Statistics Press



图书在版编目(CIP)数据

探索与创新(16):第十七次全国统计科学讨论会文集/
中国统计学会编.——北京:中国统计出版社,
2014.12

ISBN 978-7-5037-7371-6

I. ①探… II. ①中… III. ①统计学—学术会议—文集
IV. ①C8—53

中国版本图书馆 CIP 数据核字(2014)第 297834 号

探索与创新(16):第十七次全国统计科学讨论会文集

作 者/中国统计学会

责任编辑/张 赏

封面设计/李雪燕

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

电 话/邮购(010)63376907 书店(010)68783172

网 址/<http://csp.stats.gov.cn>

印 刷/北京京华虎彩印刷有限公司

经 销/新华书店

开 本/787×1092mm 1/16

字 数/910 千字

印 张/52.75

版 别/2014 年 12 月第 1 版

版 次/2014 年 12 月第 1 次印刷

定 价/133.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在

世界任何地区以任何文字翻印、拷贝、仿制或转载。

如有印装差错,由本社发行部调换。

《第十七次全国统计科学讨论会文集》

编委会

主 任：马建堂

副 主 任：徐一帆 李 强 鲜祖德

编 委：潘 璠 石方川 许亦频 孙学光 余根钱

主 编：潘 璠

副 主 编：许亦频

编 辑：吕 洁 陈妍君 邓 伟

王小舟 张 蒙 刘建军

王 谦 王淑君 张保娥

赵丽萍 关 怀 耿蕴洁

编辑说明

2013年10月，中国统计学会召开了第十七次全国统计科学讨论会。会议的主题是：大数据背景下的统计。本次大会共收到论文230余篇，内容涉及大数据背景下政府统计的变革及应对研究；大数据背景下的统计基本理论与方法研究；大数据环境下改进官方统计的建议研究；国民经济核算与统计评价研究；社会经济统计分析等方面。为增进学术交流，扩大对优秀科研成果的宣传，我们将部分参会论文汇编成书，作为《探索与创新》系列文集之十六，由中国统计出版社出版发行。

因篇幅的限制，我们按照参会论文的评审情况，将入选大会的邀请论文以全文形式刊出，并对入编的论文做了修改。受编辑力量 and 水平所限，所做的编选和删改未必精当，不妥之处敬请作者、读者批评指正。

编 者

2014年10月

目 录

中国统计学会第十七次全国统计科学讨论会学术综述	(1)
-------------------------------	-----

大数据与政府统计

大数据时代的统计挑战与应对	(17)
大数据视角下我国政府统计多级推断问题探讨	(23)
基层政府统计部门应用大数据的模式研究	(33)
大数据时代政府统计的挑战与变革	(44)
大数据时代的政府数据质量管理	(49)
大数据背景下对基本单位名录库管理工作机制的思考	(61)
大数据背景下政府信息寻租行为的博弈分析	(71)
“大数据时代”中国统计的应变	(77)
官方统计与大数据对比分析与思考	(85)
大数据环境下改进政府统计的建议研究	(97)
替代还是补充？	
——大数据及其在我国政府统计中应用的思考	(106)
大数据在住户调查中的应用论证及展望	(111)
大数据在居民消费价格指数调查代表规格品选定中的应用	(119)
刍议推进大数据环境下政府统计工作的创新研究	(125)
大数据对政府统计的影响及对策研究	(134)
工业统计数据标准化生产流程的改进研究	(145)
大数据环境下改进政府统计的建议研究	(156)
大数据背景下中国政府统计调查体系改革研究	(164)
大数据现象及理论对价格统计影响的研究	(171)

大数据理论与方法研究

基于云计算平台的权重迭代聚类算法研究	(183)
海量半结构化数据采集、存储及分析	
——基于实时空气质量数据处理的实践	(189)
大数据对统计学的挑战和机遇	(201)
大数据时代的“少数派报告”	
——基于数据分析的大数据影响研究	(209)
基于 Hadoop 的统计系统大数据平台框架研究	(218)
基于数据挖掘技术的大数据处理方法探索	(230)
大数据视域下的数据挖掘技术并行化研究	(241)
基于云服务的大数据挖掘研究	(250)

大数据的应用研究

大数据背景下完善我国现行 CPI 编制方法的构想	(261)
大数据对我国 CPI 调查的影响	(268)
“大数据”背景下利用扫描数据编制我国 CPI 问题研究	(275)
数据正态变换及在大数据分析中的应用	(289)
大数据技术下的购买力平价研究	
——基于分块自助面板单位根检验方法	(298)
大数据环境下生产型企业统计问题与对策研究	(309)
大数据环境下联网直报的现状与发展对策	(319)
携手“大数据”，传递“正能量”	
——大数据环境下改进官方统计的建议研究	(326)
大数据时代下淘宝 ISPI 给我国政府 CPI 带来的启示	(333)
基于大数据应用和投入产出技术的价格统计与预测	(342)
从 Data. gov 研究大数据在官方统计的应用	(350)
大数据在国民经济核算中的应用探讨	(358)

统计理论研究及在社会经济领域中的应用

贸易增加值率背后是企业增加值率

——针对中国加工贸易出口的讨论与测算分析 (373)

贸易开放影响非正规就业规模的理论机理与实证研究 (391)

物联网技术对国民经济核算发展的影响 (409)

要素禀赋、内在约束与中国经济增长质量 (419)

最终需求间接能耗的新核算模型研究

——基于投入产出方法的变形与转换 (433)

用分配法处理国家与省市级 GDP 匹配问题的研究 (448)

扩展双系统估计模型及其匹配性质研究 (455)

武汉市服务业优势效应分析

——基于动态偏离—份额 Esteban 模型 (468)

基于 MIDAS 方法的股市风险收益关系研究 (481)

我国利率市场化的结构效应及通货膨胀

——基于产权二元结构下利率弹性差异的分析 (492)

基于碳排投入产出表的旅游碳排研究 (509)

改善农业与农村统计全球战略：内容框架、实施情况及对中国的

启示 (528)

城乡住户一体化调查工作的实践与思考 (542)

CSD 系统对海关报关单数据质量控制的效用研究 (549)

国家统计能力比较及影响因素分析

——基于世界银行统计能力公告牌 (BBSC) 数据库的经验研究 ... (558)

群体综合评价中兼顾权威与共识的专家权重方法研究 (574)

行政记录：官方统计不可忽略的数据资源 (581)

Hedonic 法在 CPI 质量调整中的应用

——基于扫描数据改进数据支持的研究 (594)

消费者价格指数中虚拟消费的计价范围研究 (606)

基于 CARR—CVaR 模型的我国股市动态风险度量与分析 (617) 目

基于 RBF 神经网络的湖北省生态足迹预测研究 (629) 录

“碳排”基尼系数与中国省际二氧化碳排放的不均衡性 (642)

全球三大国家间投入产出数据库在增加值贸易核算方面的异同 (656)

中国对外贸易隐含碳的测算研究

——基于中国非竞争投入产出表的分析 (668)

贝塔 (β) 估计的新方法 (680)

居民消费信心内在影响因素研究 (707)

中国资源产出率估算分析 (718)

Block Bootstrap 面板单位根检验理论、方法及其应用 (728)

基于统计信息系统的网络交换技术研究 (739)

我国互连网络可持续发展水平统计测度与实证研究 (744)

基于多重抽样框的“三农”抽样设计与估计方法研究 (756)

我国公民权利意识的抽样调查方案设计 (768)

中国家庭分类贫困脆弱性研究

——基于多维框架对贫困的识别 (780)

基于全面调查数据的我国住宅价格指数编制方法研究 (791)

农民受教育程度与幸福感的实证研究 (804)

基于生态环境质量的 EKC 研究 (818)

基于随机控制模型的人力资本投资决策研究 (831)

中国统计学会第十七次全国统计科学讨论会学术综述^①

由中国统计学会主办,浙江工商大学、浙江省统计学会承办,阿里巴巴集团研究中心、浙江省高校人文社科重点研究基地(浙江工商大学统计学)和教育部人文社科重点基地(浙江工商大学现代商贸研究中心)协办的第十七次全国统计科学讨论会于2013年10月28日至29日在浙江省杭州市举办。本次会议的主题是“大数据背景下的统计”。中国统计学会常务副会长、国家统计局副局长徐一帆出席开幕式并讲话,他强调政府统计部门要主动迎接大数据时代所带来的挑战,而应对挑战的首要问题是搞清楚究竟什么是大数据,当务之急是着手制定并实施统一的数据标准,实现数据采集方式的变革。中国统计学会副会长兼秘书长、国家统计局科研所所长潘璠主持开幕式,浙江省政府办公厅副主任陆建强,浙江工商大学校长张仁寿,中国统计学会常务副秘书长许亦频等出席会议。来自全国有关院校、统计机构、政府部门的学者、官员和论文作者近200人参加了本次会议。

本次大会共收到论文236篇,经双向匿名评审,共有157篇论文入选,其中会议发言论文75篇。应本次会议的邀请,中国统计学会副会长、北京大学耿直教授,香港浸会大学数学系首席教授、系主任朱力行教授,阿里巴巴集团数据委员会会长车品觉,中国统计学会顾问、北京师范大学邱东教授,中国统计学会常务理事、中国科学院陈敏研究员,浙江工商大学副校长李金昌教授,北京大学统计科学中心联席主任陈松蹊教授,浙江财经大学副校长苏为华教授和国家统计局科研所所长潘璠在全体大会上做了特邀学术报告。中国统

^① 本文执笔:程开明、庄燕杰。

计学会副会长肖红叶教授、赵彦云教授、朱建平教授分别主持了三个时段的特邀报告。

本文选择具有代表性的学术观点进行综述介绍,供同行参考。

一、大数据背景下的政府统计

(一)政府统计面临的挑战及应对策略

随着互联网、物联网、无线传感器、云计算等快速发展,全球数据量出现爆炸式增长,人类社会进入大数据时代。目前大数据受到社会各方面的高度关注,大数据对经济社会发展产生越来越广泛的影响。大数据对政府统计的管理体系、生产方式都有不同程度的影响,民间数据产品的大量涌现对政府统计的主体地位和权威性也提出了挑战。国家统计局科研所所长潘璠在特邀报告中指出,“大数据”作为一种现象,是科学技术、社会生产力水平发展到一定程度的一种必然,大数据时代政府统计依然不能缺位,并从“什么是大数据、大数据将给政府统计工作带来什么影响、政府统计工作该怎样应对大数据带来的挑战”三个方面阐述了大数据与政府统计的关系。

“大数据”生产和使用的社会化,直接挑战当前官方统计生产和使用以及数据生态环境与运行模式。陕西省统计科学研究所郭秦川通过对官方统计与大数据现象、表现形式及特点、数据流程、数据使用、数据挖掘、数据技术等对比分析,对如何正确面对和正视“大数据”现象对官方统计的影响进行了思考,从统计运行平台、生态环境标准、专业化管理、挖掘途径与方法、人才队伍、问题解决等方面提出建设性建议。浙江工商大学统计学院程开明、陈龙认为大数据时代必然引致统计思维的转变,推动统计业务的变革,面对大数据的影响,有必要采取相应的统计应对策略。长春调查队刁鹏飞以大数据时代中国统计如何应变为切入点,阐述了大数据时代的显著特征和中国统计机构应对大数据的态度。江西省统计局科研所张启良、曾轶雄认为大数据对政府统计的唯一性、权威性产生极大的挑战,政府统计必须以变革顺应大数据时代,未来的政府统计将是传统统计和大数据利用的融合。北京市统计应用研究所刘陆芳从统计制度设计、数据采集、数据分析、数据发布四个方面列举了“大数据”时代对政府统计的冲击,对统计部门如何在竞争日益激烈的数据战争中脱颖而出进行了探讨,为建设更开放、更权威的统计提出参考建议。

统计部门要善用“大数据”,才有可能在未来的数据版图中占据一席之

地。政府统计部门能“借力”大数据提高统计调查的准确性、透明性,北京市统计局马俊炯认为政府统计应当借助“大数据”所带来的有利条件,主动顺应数据社会化的趋势,加快推进政府统计的改革。国家统计局统计科研所丛雅静等认为有效应对大数据的挑战,政府统计部门要变革统计生产方式,在制度设计、数据收集、处理、分析和发布环节做出相应调整,综合运用各种大数据处理技术和分析手段,提高政府统计的有效性。同时,政府统计部门要完善统计立法,明确大数据背景下调查对象的权利义务,加强对数据的管理职能,积极开发相关软件,重视人才培养。为了应对大数据带来的挑战,广州市统计局陈婉清认为应从整合政府职能部门信息资源、建立共享数据中心,规范统计标准、统一指标口径,改善数据采集方法、建立企业调查数据库,完善统计数据发布制度、开放统计数据信息和成果等方面改进政府统计。

大数据之大,不仅在于容量大、类型多,更在于涵盖全体数据,赋予数字信息化的意义,同时能够反映以前仅靠样本无法揭示的细节信息。上海市闸北区统计局陈新光等认为,海量数据从简单的处理对象开始转变为一种基础性资源,如何更好地管理和利用大数据已经对官方统计提出极大的挑战,数据管理方式上的变革正在酝酿和时时发生。北京市统计局研究所谢黎认为合理融入大数据技术的同时,要充分理解“少数”数据的重要作用。深圳调查队王克林探讨了大数据在我国政府统计中使用的三个前提条件:大数据的数据质量足够可靠、高度网络化的信息资源能被充分共享、抽样调查的主体地位不会动摇,并展望了未来大数据在提高统计数据分析力度、降低调查成本、提高调查精度等方面的应用。北京市统计局王宇飞通过美国政府数据网站 Data.gov 的案例得出政府应用大数据的启示。

(二)统计基础建设

从2010年开始,国家统计局全面变革统计生产方式,大力推进以基本单位名录库为基础,企业一套表制度为核心,统一的数据采集处理软件系统为平台,联网直报系统为手段的四大工程建设,采取夯实基层统计基础,优化统计工作流程,加大与部门统计的协调力度等措施,以切实提高国家统计能力,提高数据质量,提高统计公信力。浙江省椒江区统计局的谢刚强认为基本单位名录库是政府管理与统计工作的基础信息库和管理工具,大数据背景下基本单位名录库建设存在着部门信息共享性差、单位代码难统一、名录库质量难保证、维护工作难落实、名录库信息难准确等问题,建议加强统计基层力量建设,促进行政部门间的合作联动和业务融合,加大对名录库发展方向的研

究和开发应用。浙江省路桥区统计局的贺书法等认为“企业一套表”联网直报作为大数据环境下确保统计数据质量的基础性工作,是统计方法制度的重大改革,但在制度制定、报表内容、指标提示和服务平台建设等方面还存在一些弊端,必须进一步加强基础性建设,夯实统计基础,建立健全应统尽统的统计方法和制度。国家统计局统计科研所孙娜娜、何强立足基层政府统计部门的实际发展现状,从数据采集、数据处理、统计舆情分析以及统计队伍建设等方面,对大数据在基层统计应用的整体模式进行了探讨。北京市统计应用研究所黄思宁结合我国政府统计的实际情况,借鉴国内外的研究经验,分析了大数据对于政府统计数据质量的冲击和影响,提出四条参考判断标准,归纳出掌握大数据去伪的三项基本技能以及相应的制度保障。

大数据时代给经济活动参与者带来新的机遇和挑战,传统生产型企业若能加强企业自身与相关数据信息的统计分析,以客观数据为基础做出科学决策,有利于帮助企业在市场经济的大海中找到方向。浙江省铁路投资集团有限公司的余健尔等对生产型企业的数据现状与特点进行了总结与概括,深入分析了企业在统计工作中具体存在的问题,提出相应的改进措施。随着ERP、PLM等信息化系统在工业企业的部署完成以及新产品研发速度和设计效率的大幅提升,浩如烟海的企业产品信息对工业重点指标统计的准确性和快捷性均提出新的挑战。北京调查总队李珊珊认为通过建立设有“主键”的产品链数据库、应用数据透视表和函数实现海量数据计算和产品信息关联等手段,可以推行具有自动计算功能的工业统计标准化生产流程,不仅能满足大数据对工业指标统计准确性和快捷性的要求,而且有利于减轻企业负担,实现从数据源到企业的事前控制,使工业数据生产链的质量控制形成一个从原始数据整合、企业计算统计指标到政府审核评估发布统计数据的完整闭环。

农业与农村统计是各国经济统计的重要组成部分,联合国统计委员会第41届会议通过的《改善农业与农村统计全球战略》,建立了收集农业统计数据的概念框架,提出农业与农村统计的三个支柱,对世界各国的农业和农村统计具有指导性意义。东北财经大学徐强、李文肖全面分析了“全球战略”的主要内容和现实意义,对全球战略的行动计划及实施情况进行了介绍,并结合中国农业统计的现状,分析了中国与全球战略的差距。广东商学院段军山等将世界银行统计能力公告牌数据库中中国统计能力得分与世界、亚洲各发展中经济体以及一些发达国家做了比较分析,发现中国统计能力综合得分虽从2011年起有明显提高,处于BBSC数据库的中上水平,但仍低于其他金砖国家及美国等发达国家,这与我国统计体系不健全、统计投入力度不够以及发

布统计信息不及时有关。

(三) 统计调查的变革与发展

科学、合理的调查体系设计是保障各项调查工作顺利开展并保证调查数据质量的前提。大数据时代客观上要求进行统计调查体系的改革,以充分挖掘和利用大数据在政府统计中的潜力。暨南大学经济学院的陈光慧、刘建平从思想意识、思维方式、工作作风、调查方法、调查手段、调查内容、调查项目、调查指标,以及数据发布与后续分析等多个角度对中国政府统计调查体系提出一系列改革设想,以便将传统的调查方法与新式的电子化记录数据结合起来,整合利用各类大数据资源,更加准确、及时、全面和连续地产生各类宏观时间序列数据和微观纵向数据,推动中国统计管理体制的改革与发展。

住户调查是重要的官方统计内容,大数据现象的出现可以推动住户调查继续完善。广东调查总队谢韩涛认为,随着社会的进步及技术的发展,大数据对住户调查的辅助方式也在变化,并推动住户调查方式转变,同时也不断影响人们对待数据的思维和观念,但目前大数据仍只能辅助住户调查,而不能取代住户调查。为满足统筹城乡发展和改善收入分配格局对民生统计数据的需要,我国从2013年起把城镇住户调查和农村住户调查合二为一,实行城乡一体化住户调查。路桥区统计局课题组根据路桥区城乡一体化住户调查初阶段的实践,分析了大规模铺开可能遇到的困难和问题,并就如何改进和完善调查工作提出了指标设置、调查内容、记账平台和调查员报酬等方面的对策建议。

近年来,国外许多政府统计机构积极应用现代化信息技术,利用行政记录、企业财务会计资料来采集数据,努力减轻调查对象的负担。暨南大学经济学院方越峦、刘建平认为行政记录是官方统计不可忽略的数据资源,借鉴国外提高行政记录在官方统计数据来源比例的经验,在深入分析行政数据代替或推算统计调查数据的理论、现实依据和技术支撑条件,比较国内外官方统计利用行政记录异同的基础上,提出我国行政记录应用于官方统计的路径和建议。

随着社会经济改革的不断深化,诸如农民工的迁移和农地的流转等新问题不断涌现,编制单一且完全覆盖的“三农”调查抽样框已成为抽样实践中的难题。华南理工大学经贸学院贺建风副教授在构建多重抽样框的前提下,按照对“户”和对“地”抽样两种情形分别给出基于名录框和区域框结合的多阶段抽样设计方案,为政府统计部门开展基于多重抽样框的“三农”抽样提供了参考。多级推断问题已成为当前在政府统计中推广抽样调查面临的一个难题,面对“大数据”时代的海量辅助信息,中国人民大学统计学院孙婕、金勇进

提出从改进估计的角度解决多级推断问题,讨论了借助辅助信息改进估计的方法和建立统计模型进行推估的方法,并探讨了大数据背景下政府统计数据收集的一些建议。中国人民大学应用统计科学研究中心戴明锋、金勇进将调查总体划分为两个子总体:大学生子总体和居民家庭子总体,设计了一套分层多阶段系统 pps 调查方案,给出确定各阶段样本量的原则和方法,讨论了调查中出现无回答问题的补救措施,给出主要估计量及其精度的估计,针对调查中复杂样本的方差估计,提出使用刀切法进行估计。国家统计局统计科学研究所陶然针对双系统估计模型理论与其在人口普查实践应用中的不足,通过增加两项模型基本假设,结合前人研究提出构建适用于一般普查类型的扩展双系统估计模型,验证了有关学者的研究假设,并探讨了基于双系统估计模型的普查涵盖误差测量。

(四)大数据背景下的价格指数编制

6 作为官方重要统计指标的居民消费价格指数是衡量通胀程度、影响经济主体决策的重要指标之一。大数据背景下,官方价格统计作为公共价格统计数据信息主渠道的地位面临着前所未有的挑战,传统的价格统计方法不可避免地受到诸多冲击和影响。顺义调查队王娟、刘浩然阐述了大数据的来临对政府价格统计的冲击和影响,包括对数据采集方式、调查方法、工作及时性三个方面的冲击,提出可通过提升信息化建设、开放数据、加强数据信息隐私管理等手段提升政府价格统计水平。盐城调查队的虞华等详细分析了现行 CPI 统计制度和方法,认为要适当对现行制度做出改进,以尽早实现居民消费价格调查工作的科学化、制度化、规范化、标准化,适应大数据时代要求。河南科技大学的常志勇在界定大数据特征的基础上,认为大数据时代对我国 CPI 调查的影响主要包括:可以同时计算多种 CPI;有效减少计算误差;改变数据收集方法和数据质量的评估方法;对统计体系的影响等四个方面,初步设想大数据时代的 CPI 调查的流程。

北京调查总队吴琼通过比较淘宝 ISPI 和政府 CPI 的数据来源和权数确定两方面,探讨大数据时代下淘宝 ISPI 给政府 CPI 带来的启示:一是数据来源方面,政府 CPI 要借鉴淘宝 ISPI 的“全”,而不适用淘宝 ISPI 的“混”,数据“全”可提高代表性,但不分品牌和规格的“混”将会使数据缺乏可比性。二是权数确定方面,如果能将淘宝 ISPI 权数的季节性融入政府 CPI 的权数确定方法中,则可大大提高政府 CPI 的时效性。对于经济统计来说,扫描数据及其所代表的“大数据”意义,绝不仅仅是统计源头数据的信息化,更为重要的是

能为宏观经济测度提供新的条件。河南大学经济学院乔晗立足我国于现行CPI调查的实际情况,结合各国的研究现状,提出一个利用扫描数据编制CPI的思路,并对我国在这一领域所面临的挑战进行了阐述。北京师范大学国民核算研究院胡雪梅在对比分析调整CPI质量变化偏差的直接质量调整法和间接质量调整法的利弊后,选取Hedonic法对质量变化偏差进行调整,认为可利用扫描数据为质量调整提供数据支持,并提出运用扫描数据进行Hedonic质量调整的两条途径:一是扫描数据包含交易价格、交易数量、交易时间和地点等信息,为Hedonic特征价格指数中虚拟价格的计算提供数据支持;二是为质量调整因子的计算提供数据支持。

浙江工商大学统计学院的洪兴建教授针对我国住宅价格指数引发的争论,通过分析不同类型价格指数的特性和特征价格模型的特点,提出应该基于半对数模型编制杰文斯价格指数,其优点是价格指数满足所有检验、单推测和双推测的结果是等同的、可以方便地推算同比和年度价格指数等。厦门大学经济学院的付荣在梳理CPI编制目的及对应框架的基础上,参照国际统计资料讨论了纳入CPI的虚拟消费(自有住房和金融服务)的计价范围,认为需要在我国自有住房服务的估算租金的成本法中完善计价范围,添加尚未列出的费用项目;金融服务可效仿自有住房服务的处理方法,以支出法GDP的居民消费为基准口径,逐步将金融服务费用纳入我国CPI中。CPI调查中代表规格品的选定对数据的科学性和准确性有决定性作用,借助大数据技术辅助规格品的选取,能提高准确度和时效。南京调查队纪习尚以电子商务交易大数据为例,探讨了大数据技术在CPI调查代表规格品选定中的自动聚类、种类确定、动态预警等方面的应用。

二、大数据背景下的统计基本理论与方法

大数据对海量数据进行存储、整合、处理和分析,可视为一种新的数据分析方法,这种基于数据关系的内在本质决定了大数据与统计学之间的必然关系,大数据对统计学的发展既提出挑战又提供了机遇。大数据是涵盖了多个学科领域的、多源的、混合的数据集,当代的存储和计算能力难以应付这种洪水般的数据流,需要探索新的、适应大数据的处理方法。北京大学耿直教授在特邀报告中介绍了大数据背景下的抽样与分析方法,并针对大数据的来源问题探讨对多源数据进行整合的方法,针对大数据流的问题探索如何从海量数据流中抽取足以满足统计目的的数据,针对大数据的高维问题讨论了解析大数据和高维数

据降维的方法。北京大学陈松蹊教授认为统计学之所以成为一门学科,是因为其能对接受伪命题和拒绝真命题的概率给予度量,大数据的出现对统计学及实证科学在研究方法和计算上提出一系列的挑战,包括随机抽样、因果分析和数据维度等方面,但也带来很多机会,譬如新的数据分析技术、发现隐含的关系与规律,以及利用大数据提高官方统计的精度性和可信性等。

北京师范大学国民核算研究院邱东教授从“除了机遇还有挑战、大数据能否覆盖全世界、信息和噪声、统计与数据科学、大数据时代统计重心的转移、大数据与统计道德”等六个方面对大数据对统计学的挑战给出独到的看法。大数据正在改变着人们的行为与思维,以数据为研究对象的统计学该怎么办?基于对大数据的理解,浙江工商大学李金昌教授认为统计思维需要发生三个方面的改变:一是对数据的认识要改变,二是数据收集思维要改变,三是数据分析思维要改变。关于数据分析思维,又要在分析过程、实证分析方式、推断分析等方面发生变化。围绕这些变化,必须研究应对大数据的对策,以促使统计学科跟上时代的步伐。

8 随着信息科技的进步和发展,美国对大数据时代的国家战略计划激发起全世界科技界,工业界,商界的研究和应用浪潮。香港浸会大学朱力行教授在特邀报告中描述了中国信息学界对大数据领域的研究和发展的一些观点,以及统计学界的一些同仁的思考,提出两者想法:一是重分析轻建模,大数据的结构极其复杂,基于模型的统计分析在很多情况难以得到有意义的结果,探索性数据分析也许更有意义;二以局部事全局,现有统计方法不可能直接解决大数据问题,可先得到局部解,进而通过探寻局部解之间的关系,力求接近全局解。中国科学院陈敏研究员从统计是资源、统计是科学、统计是产业三个方面介绍了大数据时代为统计科学带来的难得机遇和前所未有的挑战,并通过金融资产选择与组合投资、轿车二手车定价系统的实际案例展示了大数据的现实应用。阿里巴巴车品觉会长认为大数据产业发展至今,互联网、移动互联网以及电子商务等领域的相关人士对于数据的本质是什么、数据的价值是什么、数据的出口是什么及同业的趋势向何处聚焦等问题往往未有清晰明确的认识,根据自身的大数据行业经验和理解,逐一阐述了自己的见解。武汉大学游士兵教授等认为,大数据对统计学的挑战体现在大样本标准的调整、样本选取标准和形式的重新确定、统计软件有待升级和开发及实质性统计方法的大数据化等方面;大数据对统计学的机遇体现在统计质量的提高、统计成本的下降、统计学科体系的延伸、统计学作用领域的扩大及统计学家地位的提升等方面。