

# 大数据背景下统计思维及统计人才的培养

陈宜治

(浙江工商大学, 浙江 杭州 310018)

**摘 要:** 近年来, 大数据已经成为全球讨论的焦点。大数据正在日益影响着统计的教学、统计思维和统计人才的培养。该文以大数据为背景, 讨论了大数据对统计科学带来的变化, 以及在这样的背景下如何坚持统计学的本质特点、分析任务和目标, 加强统计思维能力的培养, 提高统计人才的培养质量。

**关键词:** 大数据; 统计思维; 辩证思维; 人才培养

**中图分类号:** C8

**文献标识码:** C

**文章编号:** 1672-4437 (2016) 04-0001-04

关于“大数据”(Big data), 麦肯锡全球研究所定义为, 一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合, 具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征。指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合, 是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产大数据, 指不用随机分析法(抽样调查)这样捷径, 而采用所有数据进行分析处理。IBM 提出了大数据的 5V 特点: Volume (大量)、Velocity (高速)、Variety (多样)、Value (价值)、Veracity (真实性)。大数据是新资源、新技术和新理念的综合体。从资源视角来看, 大数据是新资源, 体现了一种全新的资源观。

统计学, 作为认识社会现象和自然现象有力的武器之一, 是一门收集、整理和分析统计数据的方法论科学, 目的就是探索数据内在的数量规律性, 从而达到对客观事物的科学认识。因此, 培养统计思维能力, 为社会培养创新型统计学人才是时代赋予教育的责任和义务。

统计思维能力, 是人们在掌握基本统计理论基础, 运用统计方法对数据信息进行综合处理分析, 进而做出科学决策、解决实际问题的思想和能力。现代信息社会要求人们, 尤其是经济管理类人才必须掌握一定的数据信息搜集、处理和分析能力, 也就是必须具备较高的统计素养。缺乏必要的统计素养, 各项管理工作和科学研究将举步维艰。正如佐依等所言“近年来统计学教育对统计思想、统计素养、统计推理的研究愈加密切”。

大数据是一种新的数据形态, 它存在于社会生活的各个领域, 向人们提出了全新的问题。所以统计思维必须创新发展, 统计人才的培养必须注入新的内涵。

## 1 大数据给统计思维带来的变化

通过对思维的理性分析可以发现, 学习统计学与锻炼思维是相辅相成的。统计学能从多方面锻炼人的思维。收集数据、整理数据、分析数据都是统计思维的过程, 而且是较为纯净的思维过程。所以, 学习统计学是最好的思维训练, 而思维却又促进对

收稿日期: 2016-09-20

作者简介: 陈宜治 (1973-), 男, 安徽临泉人, 浙江工商大学教授、博士、硕士生导师。主要研究方向: 经济统计理论方法和统计教学。

统计学的理解和应用。但是,随着“大数据”这一概念的兴起,统计思维也要与时俱进,扩展新的思维模式。

### 1.1 拓展统计思维中数据的内涵

大数据正在悄悄改变传统统计思维中数据的内涵,将从以数字为基础的数据,向以信号、符号、文本、图片、声频、音频、超媒体等为基础的数据转化,数据不再是简单的“数+量”的关系,而是呈现多种多样的形态;将从以计算为基础的统计模型、指标、数字特征等,向以分析为基础的数据处理方式转变。因此,在统计学中积极主动地思维这些变化,统计思维能力才能得到锻炼和提升。

### 1.2 拓展统计思维中数据的形式与来源

拓展统计思维中数据的形式与来源,是指将传统统计思维中,结构化数据(存储在数据库里,可以用二维表结构来逻辑表达实现的),逐步向非结构化、半结构化数据(突破了关系数据库的结构定义,不易改变和数据定长限制)延展;从通过专门调查而获取数据的方式,向专门调查与筛选相结合、筛选更为重要的数据获取方式转化。大数据中数据内容密度低的特点,决定了对传统思维中的“总体”和“样本”的概念要重新梳理。传统统计学中,先拥有总体,从总体中按照一定的规则抽取样本,而大数据背景下,无法明确“总体”的范围,而是先有样本,再探索总体的外延。同时,要接受数据容量几何级数般地增长的事实,对数据的处理和分析要有思维上的准备。

### 1.3 拓展统计思维中数据的复杂性

大数据时代中,面对的数据量在成倍增长,数据的复杂性也在发生变化。要适应这些变化,思维就要跟着改变,同时要充分把握数据的复杂性。大数据背景下,统计思维中的数据由样本数据的随机性(可由抽样分布描述)、误差构成的多样性(可一定程度测算与分解)、观察变量的多元性(可根据需要加以控制)逐步向数据总体的非确定性、数据表现的非标准性、数据含义的非单一性、数据产生的非独立性、数据真伪的难辨性转变。

## 2 统计思维从变化中寻求不变

无论数据形态多么复杂,数据以什么样的内涵呈现,也无论数据来自哪里,容量有多么庞大,永远改变不了统计学的基本真理,那就是“统计学,是一门收集、整理和分析统计数据的方法论科学,目的就是探索数据内在的数量规律性,以达到对客观事物的科学认识。”

统计思维作为一种认知世界的思维方式,它与哲学、数学一样具有同等重要的基础性地位。现代统计通过对客观事物的观察,收集、筛选相关的信息,去除、澄清混淆关系,使得事物的真相“原形毕露”成为可能。所以,统计观念对现代文明的贡献是非常巨大的。现代文明从追寻绝对真理到探求相对真理的华丽转身,正是从现代统计观念那里获得灵感的。因此,统计分析的最终目的就是“探求真相,发现真理”。无论过去、现在或者将来,统计分析的三大任务就是“发现规律,探索因果,预见未来”。这些都是统计思维坚持不变的事实和方向。

统计学只有以实践为目的,走出一条应用的道路,才能充分发挥统计学作为方法论的各种功能。形成具体的统计思维后,需要付诸真正的实践才是学习统计的最终目的。同时,传播统计思想,培养统计思维能力,其首要目的,是为社会培养创新型统计学人才。我们秉承的统计人才培养目标是:培养具有系统统计思维能力、能开展科学数据分析、准确剖析现象本质的统计人才。

## 3 大数据下统计思维能力的培养

大数据背景下,统计思维的培养应首先完成三个转变:从纯粹数学到概率思维的转变,发展符号意识;从确定性研究到随机性研究的转变;从演绎归纳到统计思维模式的转变,建立基于数据分析的逻辑思维意识。因此,如何完成这三个转变,如何提高统计学的教学质量,如何培养统计思维能力等都是值得研究和探讨的问题。有教的方面,也有学的方面;有理论上的认识,也有实践上的探索。在统计思维的培养中,教学扮演着非常重要的角色,而教学又是一个多环节综合统一的复杂过程。本节只从广义的课堂教学的角度,结合统计学的显著特点:抽象性、随机性和应用广泛性,着力开展在大学阶段统计学教学的一些新方法的研究,同时从培养学生的创造性思维能力,独立思考、独立实践的能力和调节他们自身学习心理等方面作一些努力和探索,旨在提高统计学教学水平和学生利用统计思维解决问题、分析问题的能力。更新课堂教学理念,创新课堂教学方法,提升教师课堂教学能力,提高学生自主学习、实践能力和创新能力,也是社会发展对高等院校提出的迫切要求。

### 3.1 培养系统性思维能力

作为一门方法论科学,统计学既适用于社会学,也适用于自然科学,这就要求其内容体系完整

科学。如果把注重内容的教学仅仅理解为给学生灌输大量的知识,那就未免过于肤浅了。高校教育的特点还要求教师将教学重心放在培养学生的思维能力和科学的思维方法上。只有在教学过程中注重统计学教学的系统性,才能使学生在从大格局上把握该门课程的框架,在系统整体性上抓住统计学的理论方法。统计学的一次讲课里,需要有系统的引入概念,推证定理,导出方法,分析范例等。因为统计学本身整体内容的系统性就很强,章与章之间都有着深刻的联系。每部分内容又自然一体。这就要求我们在备课与讲授中,不论是一节课还是一章内容,都要从整体上加以处理,注意部分与部分之间的有机联系。只有如此,才能做到少而精,使学生每节课都能听得明白。

同时,我们要遵循人们循序渐进的认知过程来组织教学。统计学教学要由浅入深,由近及远,由简到繁引导学生由不知到知,由知之甚少到知之甚多,由知之片面到知之全面。完全深入领会所涉及的概念和原理,通过循序渐进的认识过程,同学们会对教材中专业性的词语不再感到陌生和晦涩,而会有种“原来如此”的愉悦和快乐。

大学生已积累一定数量的文化知识,生理上和心理上也逐渐走上成熟。分析解决问题的欲望和能力都大大增强了。他们不但需要了解结果,更渴望了解过程和获得结果的推理方法。社会要求他们在实际工作中有较强的分析解决问题的能力,而提高分析解决问题的能力是以提高科学思维能力为前导的。这些都是系统性思维能力的体现。

### 3.2 培养辩证思维能力

统计学是以数据为主线,深刻地揭示了自然界各种数量之间内在的必然联系。大数据背景下,数据错乱复杂,无论从数据的搜集还是数据的整理,直到数据的分析和解读,都需要甄别、思辨等辩证思维能力。统计学中的基本理论体系,数据处理的绝妙方法,不仅反映了自然界内在的精美和谐,也反映了唯物主义和辩证法的胜利。这些辩证思维过程的学习和研究,无形中提高了学生对统计学学习的兴趣和能力,建立了对周围事物辩证认识的正确世界观和人生观。

### 3.3 培养逻辑思维能力

逻辑思维能力又称抽象思维,是舍弃认识对象及具体对象,通过语言表述反映客观事物本质和内在规律性的思维。概念的引入、公式的推导、定理的证明都体现了逻辑思维。由于统计学本身又是逻辑的科学,对培养学生的逻辑思维无疑是最重要的途径之一。在小数据世界中,人们往往执着于现象

背后的因果关系,试图通过有限样本数据来剖析其中的内在机理。而在大数据时代,人们可以通过大数据技术挖掘出事物之间隐蔽的相关关系,可以帮助我们捕捉现在和预测未来,建立在相关关系分析基础上的预测正是大数据的核心议题。通过关注线性的相关关系,以及复杂的非线性相关关系,可以帮助人们看到很多以前不曾注意的联系,还可以掌握以前无法理解的复杂技术和社会动态,也就是说,在大数据时代,思维方式要从因果思维转向相关思维。因此,要充分利用好大数据背景,培养学生的逻辑思维能力。

## 4 大数据下统计人才的培养

大数据背景下,结合统计学的特点,应该注重以下三种统计人才的培养:1. 数据科学家:能运用统计分析、机器学习、分布式处理等技术,从大量数据中提取对业务有意义的信息,以易懂的形式传达给决策者,并创造出新的数据运用服务的人才,精通统计建模以及如何构建与定制高级算法,熟练掌握软件工程技能;2. 数据工程师:能深刻理解统计学的软件工程师,既能提供可建模数据所需平台的人。他们的核心价值在于建造清晰创建数据管道,充分了解文件系统,对算法有相当好的理解,能运用基本数据模型,但对具体应用领域略微欠缺;3. 数据分析师:能洞悉方程式现实意义的人。他们知道如何提出正确的问题,非常善于数据分析、数据可视化和数据呈现。他们是数据提取、模式识别以及从大量数据中洞察问题方面的能手。懂业务与管理流程,系统掌握统计分析方法与相关软件应用,擅长数据分析结果的呈现。基于此,构建统计人才培养模式。推行“思学相通、知行合一”的育人理念,提倡思想与能力、理论与实践、应用与研究的“三结合”,以思促学、练用结合、赛证辅助、学术引导的“四联动”教学方法,环环相扣,相辅相成,构成了新型的统计学教学方法体系,形成了师生互动、教学相长的统计素养培养机制,为培养高质量统计人才提供了强大的方法保障。

### 4.1 以“双核”为课程教学的基本目标

统计素养的内容极其丰富,我们在教学过程中发现基本统计思想和综合统计技能是它的两个核心,称之为“双核”。著名统计学家陈希孺院士曾经指出,统计素养除了知识技能和解决问题的修养外,也应包括思想观念的修养。

基本统计思想就是统计的世界观与方法论,包括统计的目的、内容与方法。统计学课程的教学目

的,不是仅仅让学生记住一些公式、掌握一些概念,而是能够真正领悟统计学方法的思想与灵魂,能够用独特的统计思维模式去思考、解决实际问题。也就是说,通过基本统计思想的培养,能够让经济管理类人才在面对具体问题时具有理性的统计思维和科学的定量方法。

综合统计技能就是综合运用各种统计理论与方法,从定性与定量的结合中去发现、分析并解决现实问题的能力,是统计思想与专业技能相结合的能力。统计学课程作为统计学科最基本的原理性课程,包括了一整套关于数据搜集、整理、分析、检验和预测的方法。如果能够让学生把统计方法与相关领域的具体现象很好地结合起来,必将极大地提高他们处理和解决实际问题的综合能力。

#### 4.2 以“四联动”为课程教学的基本方法

如何在教学过程中系统培养学生的“双核”统计素养呢?我们在实践中逐步总结出了“四联动”的教学方法:

4.2.1 以思促学——理念驱动。通过设计科学先进的课程内容体系,将课堂教学打造成一种兴趣盎然的活动,使学生在统计思想的熏陶下掌握课程知识,并根据统计学方法论的特点培养学生的研究思维能力。通过课堂教学与网络教学、实验教学、案例教学相结合,基本教学方法与启发式、探究式、讨论式、参与式教学方法相结合,把统计思想的培育贯穿于整个教学活动。我们强调统计科研与统计教学结合,注重把统计科研成果转化为教学内容;强调统计方法与学生专业相结合,注重不同专业对统计方法的不同要求。

4.2.2 练用结合——实践推动。思而后行,学以致用。统计学十分强调应用性,实验教学和实践教学十分重要。统计学只有以实践为目的,走出一条应用的道路,才能发挥统计学作为方法论的各种功能。实践,不仅包括让同学们独立完成各自相对内容的联系和作业,而且要强化他们“写、议”的能力。学习每一章内容之后,留出一节习题课的时间来进行讨论和提问,彼此相互交流;并且要求每一独立内容学习完毕交出一份相关的学习体会、心得或思索、探究,这是一个方面。另外方面,计算机强大的计算能力使实践如虎添翼,显得比以前任何时候都更具有威力和渗透力。一些复杂的统计学问题,现在可以依靠计算机,对各种理论进行试验,对建立的统计学或者数学模型进行模拟。因此,计算机的引进改变了统计学实践方式,为统计学教

与学提供了新的方法。利用计算机技术收集整理和分析统计数据,运用专业的知识理解数据的内在规律性,学会常用的统计软件,excel、SAS、SPSS等等。

4.2.3 赛证辅助——技能带动。统计方法的应用就是转化为一系列技能,包括数据收集、数据处理、数据分析以及统计软件应用等技能。通过创办大学生统计调查方案设计大赛,鼓励学生结合各自的专业特点组成团队参赛,以赛促练,以练促学,促进教与学共同提高。“统调大赛”的研究报告,因其调查方案的严密性、研究主题的针对性和数据处理的科学性而成为“挑战杯”获奖作品的摇篮。统计软件(SPSS)考证系统,进一步提高了学生运用软件进行数据处理的能力。

4.2.4 学术引导——创新促动。在不断改进课堂教学、实验教学和实践教学的同时,我们还积极鼓励学生开展学术创新活动。我们一方面组织学生参与教师的课题研究,通过教师的带领来培养和激发学生的科研兴趣;另一方面引导学生参加“挑战杯”和其他学科竞赛,培养学生的自主创新研究能力。

大数据背景下,统计思维和统计人才的培养是一个不断完善不断提升的永无止境的过程。“总有一天,统计思维就像人的读写能力一样,成为一个有效率公民必备的能力。”<sup>①</sup>

---

注释: 本文得助于浙江财经大学李金昌教授的教诲和悉心指导,文中很多思想来源于李教授的授课和演讲内容,在此致上诚挚谢意!

#### 参考文献:

- [1] 李金昌, 苏为华. 统计学(第4版)[M]. 北京: 机械工业出版社, 2014.
- [2] 李金昌. 统计思想研究[M]. 北京: 中国统计出版社, 2009.
- [3] C. R. 劳. 统计与真理: 怎样运用偶然性[M]. 北京: 科学出版社, 2004.
- [4] David S. Moore. 统计学的世界[M]. 郑惟厚译. 北京: 中信出版社, 2003.
- [5] 萨尔斯伯格. 女士品茶[M]. 邱东译. 北京: 中国统计出版社, 2004.
- [6] 戈尼克等. 漫画统计学入门[M]. 梁杰等译. 辽宁: 辽宁教育出版社, 2002.
- [7] 李金昌. 应用抽样技术[M]. 北京: 科学出版社, 2010.
- [8] 维克托. 大数据时代[M]. 周涛译. 杭州: 浙江人民出版社, 2012.