

大数据与统计新思维

Big Data and Statistical New Thinking

桂春艳

Chunyan Gui

河北省塞罕坝机械林场千层林场

中国·河北 承德 068456

Qianceng Forest Farm of Saihanba Mechanical Forest Farm,
Chengde, Hebei, 068456, China

【摘要】随着现代互联网的普及,信息技术的飞速发展,中国现在已经迈入了大数据时代。大数据时代的特点是网络数据的信息越来越多且数量也越来越庞大。这也必将会影响到与数据、统计相关的经济学领域。因此本文首先解析了大数据的内涵,分析了大数据背景下统计思维的转变,最后就经济受到的大数据影响展开了讨论。

【Abstract】With the popularity of the modern Internet and the rapid development of information technology, China has now entered the era of big data. The characteristics of the big data era are that the information of network data is increasing and the number is increasing. So it will certainly affect the field of economics related to data and statistics. Therefore, this paper first analyzes the connotation of big data, analyzes the change of statistical thinking in the context of big data, and finally discusses the impact of big data on the economy.

【关键词】大数据;统计;新思维

【Keywords】big data; statistics; new thinking

【DOI】<http://dx.doi.org/10.26549/cjygl.v2i6.841>

1 引言

现代的社会是一个高速发展的社会,不仅科技发达,信息流通速度也在逐渐增快,而大数据就是这个高科技时代的产物。有人把数据比喻为蕴藏能量的资源,按照不同资源的性质对其进行了分类,当然不同资源的开发成本也不尽相同。与此类似,大数据并不在“大”,而在于“有用”。其价值含量和开发成本比数量更为重要。因此对大数据进行研究,可以从多维的角度来分析问题。

2 大数据内涵解析

2.1 统计学角度的大数据内涵

首先应确定的是,大数据与传统的数据的区别不仅仅是“量”大的区别,还在于“数据”本身的内涵。站在统计学的角度来解析大数据的内涵:大数据是建立在现代信息技术和新科技手段基础上的连续的、扩充的数据形式,其在记录能力和存储能力上远胜于传统数据,且摆脱了传统数据储存固定、数据不连续、储存有限和不可扩充等一系列缺点。总体可以概括为在短时间内无法通过目前的规范工具进行处理、管理、捕捉的所有数据集合^[1]。

2.2 传统的统计研究与大数据时代统计的区别

首先是数量上的明显差异,大数据其基本特点之一就是

数据量的庞大。其次,在研究对象方面也具有区别,传统的统计研究对象是以研究人员有意向的、收集的特定样本的数据,属于格式化数据。而现在大数据统计所面对的研究对象是随意的、动态的、容量大的多类型数据。另外,在数据的使用次数方面,传统的统计数据可能只限于此次的研究,而大数据的数据因为具有动态特征,可以经过筛选后用于多次研究。最后,具体使用中,传统统计的样本数据相对固化,如果抽检样本中出现没有预测的情形可能就不必重新收集进行研究。而大数据的动态扩充使得信息能够在一定技术支持下从多方面、多角度满足研究的需求。

3 大数据背景下统计思维的转变

3.1 创新数据梳理与分类方法

由于大数据的数量庞大,且储存的数据信息不断进行着动态扩充,所以当下大数据时代的难点在于对数据的搜索。且由于大数据无限量储存的特征容易造成数据混乱且储存样式多种多样的情况,这就要求研究人员在使用大数据时,能够准确分析和筛选所需的数据信息,找准数据特点进行有效的梳理和分类,通过不断创新以保证数据的高效利用率^[2]。

3.2 创新样本的采样方式

进入大数据时代后,传统的数据采集已不能满足当下的研究要求,虽然目前传统的采样办法依旧常见,但大数据的崛起必然带来采样调查工程的转变,促使新的采样办法产生。

3.3 因果关系与相关关系的演变

因果关系可以说来自于对经验的总结,而在大数据时代以前,研究者只需在了解因果关系的基础上,就可以进行研究展开,但随着大数据时代的改变,只要了解两者的相关关系,就能实现对未来的预测。但是目前为了研究的科学性与严谨性,要求不仅仅了解事物的因果关系,还要在此基础上建立起两者之间的相关关系。所以大数据时代的研究对两种关系的运用缺一不可^[9]。

3.4 统计与云技术的融合

大数据时代,由于数据的量过于庞大,所以对数据的收集技术和统计技术提出了新的要求。虽然当前的数据收集技术和统计数据技术随着科技的发展也达到了一定技术高度,但对于大数据时代而言,还需要进一步提升,因此在进行数据的统计和分析时,应充分利用新的技术手段融合云技术对数据进行处理运用^[10]。

4 大数据对经济学的影响

4.1 研究对象成为了主体

传统的研究经济过程中,受到数据收集能力的限制主要采用抽样的方式进行,其做法是通过对整体的部分进行抽样研究,这样也使得最终的研究受到所抽取样本的影响,具有一定的风险误差。而大数据时代的到来使得研究对象成为了主体,其数据来源更据科学性,也在一定程度上影响了研究结果。

另外还需注意研究主体时的方法运用:一般对主体进行研究分析时,可以利用统计分析方法中的分组分析法,即依据统计分析目的要求,将研究总体依照标志进行划分,然后再整理,最后进行分析、观察以发现事物之间的内在联系和规律性。运用分组分析是在遇到统计整体单位具有较多特征,需要分析总体数量的关系和数量的特征以及总体内部分组分析时的一种有效研究办法^[9]。

4.2 研究超脱了假设检验的限制

传统经济学研究时通常会依据所研究的内容提出多个假设,然后对假设利用数据模型进行检验。但大数据时代的到来,为研究提供了充足的数据和变量,只要运用合理的人工智能手段对数据进行有效的挖掘和开发,就可以得出大量的结论和变量,而不是依据小范围数据统计提出假想假设,传统研究的假设假想对于当下的大数据背景而言,就显得缺乏严谨性、科学性。除此之外,大数据的应用还带来了处理结果的多样化和完备的变量条件,这些都使得研究愈加具有价值^[9]。

4.3 给计量经济学造成了冲击

计量经济学是以一定的经济理论和统计资料为基础,运

用数学、统计学方法与电脑技术,以建立经济计量模型为主要手段,定量分析研究是具有随机性特性的经济变量关系的一门经济学学科。曾经诺贝尔经济学奖中的一半以上人都是计量经济学家,但是大数据时代的到来改变了这一状况。

举例说明,在研究自变量 X 与因变量 Y 之间的关系时,检验 X 的回归系数为 t,如果相伴概率没有超过 0.05,则表示两个变量之间有所关联,那么两个变量没有关系的概率仅为 5%。2012 年,中国网民数达到的高度为 5.64 亿人,在研究过程中,假设中国网民平均受教育程度(X)与上网时长(Y)之间关系,那么 5%也就是 2820 万人,这样的情况下,庞大的基数致使我们不能忽视这 5%。

4.4 改变了统计调查

综上,大数据的到来使得统计调查方式发生了巨大的变化。传统方式在处理统计学中遇到异常点时,往往都会直接丢弃,也对最终的研究结果造成一定影响。但是运用大数据则能改变这种情况,其数据量大,所以使得异常点能够得到应有的重视,且计算到其对最终结果的影响。这也是在统计数据过程中,人们重视非结构化数据和原始数据的原因,因为经过加工的“二手数据”,始终会对最终的统计结果造成或多或少的偏差影响。

5 结语

大数据的到来,为统计带来了革新,既创新了研究方法,也精准了最终的统计结果。同时给新时代的经济带来了巨大影响,也为研究最终的准确率提供了保证。因此大数据时代的到来也对进行调查和研究的人们也提出了新的挑战。因此这就要求在大数据背景下,研究人员必须要具有新的统计学思维,在进行分析时,及时运用一些现代的高科技技术和手段来处理大数据,才能更好地进行具有严谨性和科学性的统计及经济的研究。

参考文献

- [1]李秀丽.论大数据与统计新思维[J].中国商论,2018(05):5-6.
- [2]邱丽萍.基于大数据与统计新思维的研究[J].经贸实践,2017(18):246+248.
- [3]李伟.浅谈大数据与统计新思维[J].财经界(学术版),2016(04):293.
- [4]陶雪娇,胡晓峰,刘洋.大数据研究综述[J].系统仿真学报,2013(8):142-144.
- [5]国务院“促进大数据发展行动纲要”[Z].电子政务,2015,(09).
- [6]国家发改委办公厅关于做好制定“互联网+”行动计划有关工作的通知 [EB/OL].(2015-03-31)[2015-04-25].<http://www.jushihuichina.com/a/xinwenjijiao/xingyezixun/2015/0331/258.html>.