

大数据时代下对统计工作的思考*

岳大波

(成都信息工程学院 成都 610103)

摘 要 大数据时代是推动经济社会变革的重要力量。对传统统计工作的方式、方法以及社会重要性将产生重大的影响。本文从大数据的视角下全方位对统计工作的变化进行了探讨,并结合统计工作实际提出大数据背景下统计工作应该做相应的变革,对促进统计工作能效具有重要意义。

关键词 大数据 统计信息化 数据管理

中图分类号 C829.23

文献标识码 A

文章编号 141202-6881

The Thinking of Statistical Work under the Era of Big Data

Yue Dabo

(Chengdu Information Engineering College Chengdu 610103)

Abstract The age of big data is an important force to promote the economic and social change. Method of traditional statistics work mode, as well as the importance of will produce major effect. From the data perspective change a full range of statistical work are discussed, and combined with the practical statistical work put forward under the background of big data statistical work should make corresponding change, it has important significance for promoting the statistical work efficiency.

Keywords Big data Statistical informatization Data management

一、前言

随着网络的广泛应用与发展,数据量几乎成指数级增长,这就促进大数据时代的到来。大数据是商业智能化的一个新的发展阶段,商业智能这一概念早在 20 世纪 50 年代就被 IBM 公司提出,大数据时代通过对大数据的深度分析和价值发掘将有力提高社会生产效率,从而推动经济持续增长,为提升国家的竞争力起到重要的作用。

二、大数据时代对统计工作提出的挑战

1、大数据时代对统计数据的需求层次提出挑战

传统统计技术帮助人们通过观察小部分的样本情况从而描述与推测总体情况,帮助人们更好地了解中观和宏观数据。统计数据反映的是总体趋势,是无法对应到具体的个体情况的,也就难以获得更加有价值的信息。在数据采集技术推动下,人们可掌握的数据可以通过全局性的宏观数据,到中观层面的数据,最后回到微观层面上来,这既是关注个体的行为模式。“大数据时代”的来临使人类第一次有机会和条件在非常多的领域以及非常深入的层次获得和使用全面

数据、完整数据和系统数据。这样的系统数据可以更加偏向个体情况,更加符合个体的感知,产生宏观数据无法实现的价值。

2、大数据时代对统计实证的研究流程提出挑战

“大数据”采用的是自下而上的数据挖掘模式,不需要预先设定研究目的或方法,以数据为先导,从大量实际产生数据中通过数据挖掘技术找到数据之间的关系并建立模型,以至于形成理论和新的认知。然而传统的统计的实证研究范式,是先从经济理论或社会经验出发,并且根据理论设定指标,最后利用统计数据进行实证分析。“大数据”的研究范式已经被广泛运用到当前的数据处理中,对统计实证的研究提出更高要求。

3、大数据时代对传统统计数据搜集流程提出挑战

依据统计研究的目的和研究对象的特点,来设定统计指标、统计报表以及调查、汇总方法等制度,通过调查、访问将生产行为转化为可以量化的统计数据。“大数据”下的统计数据搜集需要面对的是已经存在的或通过一定手段可以获取的大量数据,并且弄清楚是通过什么方式产生的这些已经存在的相关

* 注:本论文为国家统计局统计信息技术与数据挖掘重点开放实验室开放课题的部分研究成果,项目编号:2011YBAN13

数据,这些数据是又从那些领域产生的,并且会对研究目的产生什么影响,进而确定采用什么样的数据来反映统计目标。

4、大数据时代对统计工作中数据采集智能化提出了挑战

传统统计是通过结构化的报表和统一的计量方式将被调查对象的行为转化为可用的数据,这其中需要被调查者的参与。然而一旦调查对象不予配合,或者采用撒谎(虚报、瞒报、乱报)的方式消极配合,统计数据的质量就难以控制,会造成调查误差较大,并且矫正这些调查误差的成本将会很高。不同于传统统计,在“大数据”时代,数据来源于信息技术记录下的原始数据,这些数据仅仅依赖于测量方法(如 GPS 定位测量、超市收银管理系统、ETC 电子收费系统),而不需要调查对象长期、认真的配合。一旦测量技术成熟,大量数据的传输中,想要篡改数据变得非常困难,数据质量将会大大提高。

5、对统计工作中的数据分析提出挑战

传统的统计分析是根据统计制度设计的要求以及对研究对象的认知,对采集的结构化统计数据进行分析,对未来的发展情况进行统计意义上的预测。“大数据”背景下的数据分析,面临的是大量存储于各处(包括“云端”)的非结构化或半结构化数据环境,首先要将这些无法识别和运算的信息转化为结构化的数据,针对结构化的海量数据分析挖掘数据之间的内在关系,寻找更多价值信息。在这一背景下,数据分析就变成了统计部门一个关键性的环节,需要专业化的数据挖掘与处理技术。

三、“大数据”时代加速统计改革

统计应当借助于“大数据”所带来的有利条件,加快推进统计工作的改革,主动顺应数据社会化的趋势。当前可以从推进统计业务信息化、优化统计数据管理、提高数据挖掘分析能力、完善统计数据发布等方面着力。

1、扎实推进统计工作信息化

“大数据”时代统计工作中调查对象行为的数据化以及调查业务的数据化对统计工作提出数据化的要求。在数据化管理时代,企业更重视自身的业务模式的数据化。加快推进统计工作信息化进程,帮助社会形成生产、物流、交易等环节的数据化,并将这些数据通过互联网系统搜集、存储起来。推行的“联网直报”等四大工程也显示,网络技术对人工的替代导致地方统计机构职能做出调整,由以前的催报、加工汇总、审核等职能转化为质量控制,促进统计工作效能的提高。

2、统计数据由数据搜集过度到数据管理

调查业务的数据化主要是针对统计调查的质量控制。目前“联网直报”采取 IP 定位控制、修改痕迹保留等方式就是这一理

念的现实应用。比如为了保证每个采价员按照“三定原则”进行采价,可以对每台采价器进行数字化定位,其采价的时间、地点、数据情况可以即时传到服务器,即可进行监测。未来的统计数据采集可能就是联网的测量仪来进行,统计业务的数据化管理显得尤为重要。

3、大力提高数据挖掘分析能力

“大数据”的概念来源于互联网,其要求的数据分析已经远非目前的统计数据处理技术能够完全实现。我们所要面对的大量、复杂的行政记录、商业交易数据实际上是结构化的海量数据,从这些海量数据中提取隐含的关系、模式和趋势等信息供统计使用。为了在未来的数据竞争中统计发挥更大的社会功能,就必须在数据挖掘方面作出更大努力。

4、完善统计数据发布

统计最终是为用户服务的,发布对用户有价值的统计数据才是统计存在的意义所在。海量的“大数据”对不需要的用户也是信息负担,因此,我们需要提高数据发布的针对性,发挥数据的最大信息价值。在发布数据的形式上,以文字、表格的方式发布统计数据已经跟不上用户的数据需求,也落后于商业调查的步伐。因此,我们在数据发布方面还需要进一步加大创新力度,完善各种发布渠道,充分发挥数据的价值。

四、结论

大数据时代的到来,对传统的统计工作是机遇也是挑战,顺应这种潮流,并借助于“大数据”的力量变挑战为机遇,变数据搜集为数据管理,加快提高统计分析能力,使统计工作在大数据时代中发挥更大的作用。

参考文献

- [1]许小乐,“大数据”与政府统计改革[J].调研世界 2013 年第 5 期.
- [2][英]维克托·迈尔·舍恩伯格、肯尼斯·库克耶著,周涛译.大数据时代[M].P32 2012 年,浙江人民出版社.
- [3]韩家炜、Micheline Kamber.数据挖掘:概念与技术[M].机械工业出版社,2006 年.
- [4]李红宇.信息时代的统计发展趋势探讨[J].北方经济,2003(05).
- [5]程芙蓉.推进我国统计信息化建设的思考[J].行政事业资产与财务 2011(02).

作者简介

岳大波(1978~),工作单位:成都信息工程学院统计学院工作,职称:讲师,学历:硕士,研究方向:统计信息化方向。