

论大数据时代统计数据处理思维的转变

田浩

云南开放大学 云南昆明 650223

【摘要】 随着互联网时代的到来,使得人类拥有的数据量达到了前所未有的程度,大数据时代到来对统计思维提出了新的挑战和要求。本文在大数据与传统数据对比分析的基础上,提出在大数据时代背景下,如何转变统计数据处理思维的几点意见和看法,期望对提高我国统计工作有所帮助。

【关键词】 大数据时代; 数据处理思维; 转变

我们知道,统计是通过研究现象的数量及其变化反映现象质的一门学科,其研究的基础在于数据的“大量性”,没有足够的“量”就无法准确推断出现象的“质”。在传统统计工作过程中,我们通常通过搜集—整理—分析三个工作步骤完成数据处理工作。在数据匮乏时代,这种数据处理模式不但是必要的、也是必需的。但是,随着“大数据”时代的到来,我们惊讶地发现,这一传统数据处理思维受到了前所未有的挑战。

一、大数据与传统数据对比分析

大数据(Big data),是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。它具有海量规模性、快速流转性、类型多样性和低价值密度性四大特点。

从以上定义及特点我们可以看出:大数据与我们认识的传统数据相比较,其数据来源、数据类型及量化表现形式都发生了根本性的变化,具体体现在以下几个方面:

1. 数据来源

传统数据具有明确的目的性,数据提供者大多身份确定且特征可明显识别,事后还可以跟踪核对;但大数据来源于物联网,不是为了明确目的而产生,它自动记录下人们一切可记录的信号,在这种情况下,识别数据提供者身份变得十分困难、很难追溯数据来源。

2. 数据类型

传统数据属结构型数据,即定量数据加少量定性数据,格式化且有固定的标准,一般用统计指标、统计图表就可以表现;大数据则是非结构型数据,包括一切可记录、可存储的信号,多样化且无固定标准,难以用常规统计方法予以表现。不同的网络信息系统,其数据识别方式就有差别,不需要预先设定就能够自动包容各种各样的海量数据。

3. 量化表现形式

传统数据量化处理经过上百年的发展,已经形成一整套完整方法,量化结果可直接用于各种运算分析;但大数据中非结构化数据如何提取、量化以及实现与结构化数据的对接,却是我们面对一个崭新问题。我们很难套用传统的统计公式、定律去量化非结构化数据,更谈不上对非结构化数据进行分析处理。

二、大数据统计数据思维的改变

综上所述,大数据杂乱无章且不规整,它与传统意义数据存在着很大差异,但时代的发展决定了我们不能回避、拒绝大数据,只能接纳、包容。我们需要顺应时代需求、转变思维,将统计研究的对象范围从结构型数据扩展到全部数据中,重新

定义数据及分类,改革和创新分析方法。秉承“只有未被利用的数据,没有无用的数据”这一理念,从全新的角度去探索搜集、整理及分析数据方法。下面,从传统的统计工作过程中“搜集—整理—分析”三个步骤入手,阐述如何转变统计数据处理模式。

1. 数据搜集思维转变

数据搜集是进行统计分析的基础和前提。没有黏土,何以做砖?传统的数据搜集模式,是根据统计研究的目的和任务,精心设计调查方案,应用各种调查方法,有意识搜集所需数据。这样的搜集模式得到的数据,很容易受到调查者主观臆断影响,调查者往往会根据不同程度偏好,选择自认为适合的调查对象展开调查,获取自认为有用的信息,而不自觉地屏蔽、撇弃自认为无用的信息。这样的搜集模式带有很强的主观性而导致信息失真。现在,我们拥有了大数据,就等于拥有了超大量数据可供选择,这样的数据是客观存在的,不再是我们主观选择结果。因此,我们的数据搜集思维应由“搜集数据”向“利用数据”转变,所做工作也不再是带有很强的主观性去搜集数据,而是对客观存在的海量数据进行归纳总结。

2. 数据整理思维转变

由于大数据存在来源与种类多样性、数据变化快速性等特点,我们不得不面临如何甄别数据真伪以及如何提炼和利用数据这一问题,这是数据整理阶段所必须解决的。现代大数据更新速度飞快。今天的大数据,明天就不再是大数据,此时有用的信息,几秒后就可能变成完全无用的信息!如果我们不能及时提取有用数据,放任数据库自动更新,就有可能失去很多宝贵的数据信息。因此,现代大数据的特点迫使我们不得不转变数据整理思维,即由传统思维的“如何筛选出有用信息”转变为大数据时代下的“如何删除无用信息”。如果说以前有针对性地获得数据叫做数据整理,那么今后有选择地删除数据才称得上是真正意义的整理。因此,大数据时代的数据整理思维将更多的是一个从已有的超大量数据中剔除无用信息的过程。

3. 数据分析思维转变

随着上述搜集、整理思维的变化,现代大数据分析思维模式也必将随之转变,我们要学会利用各种软件工具,从大数据中挖掘有用信息,并在这个过程中发展、创新大数据分析新方法,具体从以下三方面进行转变。

(1) 分析过程转变

在传统思维下,统计分析过程是“定性—定量—再定性”,第一个定性是初步确定事物性质,从而找到今后定量分析方向,这一过程主要依靠经验判断,在过去数据短缺、分析手段落后的情况下,这是不可或缺的手段。但在海量数据的存在、分析速度大为提高的今天,初步确定事物性质变得越来越简单,我

们所要做的仅只是从各种“海量数据”中找出那些真正有用的数据信息,探讨其数量特征及变化规律,作为判断或决策的结论依据。因此,大数据时代数据分析过程完全可以简化为“定量—定性”,从量的结果及其变化直接确定事物性质。

(2) 分析思路转变

传统的统计分析思路,一般是先根据研究目的提出某种假设,然后通过数据收集与分析去验证该假设是否成立,即我们常说的“假设—验证”。这种思路下的验证往往由于受到假设的局限、所需数据的缺失而得不出真正符合实际的结论,在实际过程中也会经常出现实证分析纯粹就是为了凑合假设的假象。现在,我们有了大数据,就可以不再提出假设,直接从海量数据中去寻找关系、发现规律,然后再加以总结、形成结论。也就是说,大数据时代数据分析思路已经由传统的“假设—验证”转变为“发现—总结”。

(3) 分析方法转变

传统的量化分析大多数采取抽样-推断法完成,即先从总体中抽取部分样本,再通过样本推断出总体特征。其推断是否正确,完全取决于所抽取的样本数据,不可避免存在一定程度的误差。现在,大数据面对的是全部数据,不需要抽取样本进行推断,只需对得到的所有数据进行计量和归纳即可完成。不仅如此,还可以根据全部数据的分布情况来判断其中出现某类情况的可能性有多大。这一数据分析思维方法的转变。可以大大简化数据分析过程,避免了总体和样本之间误差的出现。

总之,大数据时代的到来,给传统的统计思维带来了前所未有的挑战和机遇。我们必须清楚认识到这一变化,及时转变统计思维模式,探索出新的数据搜集、处理及分析方法,方能应对大数据时代对统计的要求,使统计这门学科在大数据时代焕发出新的活力,得到更好的发展和应用。

参考文献

- [1] 李金昌. 大数据与统计新思维. 统计研究. 2014年第1期.
- [2] 柴世琿. 试论大数据统计的新思维. 中国高新技术企业. 2015年第9期.
- [3] 陈宇秋. 小议大数据与统计新思维. 人力资源. 2016年第2期.