# 浅析大数据时代统计学的发展

来源：网络 作者：静默星光 更新时间：2024-02-09

*大数据时代以迅雷不及掩耳之势席卷世界，在全球范围内掀起了前所未有的数据革命浪潮。相对于政府单位的统计数据来说，大数据主要利用的是多层次、多样化的数据采集方式，整合了多种数据的开发优势，并且利用现代科学技术手段和高速处理以及信息架构数据等资源...*

大数据时代以迅雷不及掩耳之势席卷世界，在全球范围内掀起了前所未有的数据革命浪潮。相对于政府单位的统计数据来说，大数据主要利用的是多层次、多样化的数据采集方式，整合了多种数据的开发优势，并且利用现代科学技术手段和高速处理以及信息架构数据等资源，兼具极高的使用价值和判断决策能力。一方面，统计调查数据的多样化发展趋势和电子商务产业的不断发展，为统计数据的使用方式和生产方式制造了不小的麻烦，不断地挑战者政府部门数据管理系统和统计数据的概念。另一方面，信息技术、网络发展以及空间信息技术的不断进步，为统计生产力的升级发展提供了广阔的视角和空间。数据量急剧增长的电子化、信息化和产业化数据，都成为了统计数据发展的重要来源。种类不断增多的“大数据”资源，正在成为政府统计部门利用研究的重要领域。

一、大数据与统计学的区别

统计知识在大数据的利用研究中有多样化的应用形式，主要是对“大数据”进行肢解，对爆炸增长的数据信息进行搜索、分类以及整合主要依赖于统计学。因此，大数据的相关研究在一定程度上运用了统计学的知识。但是，大数据的使用尚未被统计学这门学科充分利用，这主要是因为大数据的运用方式，使用模式和统计学之间存在着重要差异。统计学主要利用的是样本统计资源，样本主要在根据既定的概率标准从总体中抽样调查，但是随机抽样调查是带有成本属性的，例如消耗时间、资本投入的成本等。在样本数量逐渐增加的情况下，样本估计的误差范围是伴随着总体样本数量的增大而逐渐增加的，这是样本统计学不能忽视的缺点。大数据时代最具代表性的就是海量的信息数据化以及即时电子商务信息，大数据在整体上呈现出“总体样本数据化”的趋势，这样的特征恰好可以补充样本统计的弊端。大数据环境下的整体样本统计即使可以囊括全部的样本容量，但是因为很多情况下数据具有非结构性和半数据化的特征，而且大量的数据资源呈现的是重视尾部分布的状态，方差、标准差等标准化的方法变得毫无意义，整体依靠性和不稳定性经常会超越经典时间内的时间序列的整体假设性，所以概率论的应用范围呈现狭窄化的发展趋势。因此，统计学在利用大数据进行样本统计的过程中，可以对整体上的数据资源进行融合和选择，这和样本统计中的数据化处理技术存在异曲同工之妙。

二、大数据时代统计学教育的发展

1.全面培养人才素质

统计学专业的学生需要具备良好与人交往能力。统计学的学生很多都是理科出身的学生，不善于交际。但是在日常的工作中，有数据经验的科学家应该经常和每个部门的工作人员交流，协同工作。怎么样才能让颇具专业性的数据分析结果让普通的老百姓也可以读懂，让每个部门的工作人员都能无障碍地理解，这是不容易做到的。要训练自己的交往能力和沟通技能，主动地参加演讲活动是不错的渠道，演讲活动锻炼了演讲者的自信，在整个演讲的过程中，能否清晰地表达自己的思想以及给人以信服力是至关重要的。需要培养数据常识，广其见闻。数据科学家经常面对各种各样的海量数据，并需要从这些数据中挖掘出有价值的信息，这就需要数据科学家具有强烈的数据敏感性。对数据的敏感程度的训练不是一蹴而就的，要经过长时间的积累和数据分析工作的磨练，同时也可以根据阅读数据分析材料积累阅历，提升对数据资源的敏感程度。

2.培养应用型人才

大数据时代培养的数据科学家需要两方面的基本素质，第一是概念性，也就前面所说的数据科学家需要掌握的基本素养和专业知识;第二是实践性，也就是本文中我们提及的应用型人才，也就是实际操作中处理数据的能力。在高校开展大数据分析研究生学科，最大的问题是没有可用的数据，这就需要高效与大数据企业合作，进行研究生的联合培养，注重学生的实际操作能力，这里面涉及到我们的应用统计学专业硕士的双导师培养制度，一名校内导师一名校外导师，校内导师注重学生的概念性，校外导师注重学生的实践性，学生通过在校外导师单位的实习，从而熟悉并且掌握实际工作中所需要的技能。

3.促进统计与数学、计算机学科合作

“大数据”时代需要的海量数据分析资源仅仅凭借统计学科单一学科的发展是不能满足发展需求的，大数据的数据结构性特征已经抛弃了传统意义上的数据分析模式的非智能化框架，而且数据分析需要利用新型的数据运算方式以及计算机技能分析，这也是进行数据分析工作的拦路虎。所以，数据科学家的成长仅仅依靠单一的统计学科知识的学习是远远不够的，其需要的是数学、计算机和统计学三门学科融合发展，紧密结合。三门学科之间交叉发展，融会贯通，这样既可以发挥学科的优势资源，同时也能弥补其他学科的弊端。

三、结语

数据信息的爆炸式增长使我们在使用统计数据处理信息时需要更多的数据资源，更有甚者，在很多情况下可以利用全面化的数据，数据资源不再是制约统计分析的唯一因素，大数据前提下的统计学效用和粘合度预测的准确程度不断提升，而且可以发现诸多在样本统计基础上未能显现的细节。统计学关键优势就是“见微知著”，也是统计学在数据环境下的约束性妥协。在海量数据汹涌袭来的年代，充分发挥统计学的优势，和大数据资源整合发展，实现“以小见大”和“由繁入简”的有效结合。

本DOCX文档由 www.zciku.com/中词库网 生成，海量范文文档任你选，，为你的工作锦上添花,祝你一臂之力！