

浅谈大数据时代统计学方法的创新与发展

云 昊

(呼和浩特市统计局, 内蒙古 呼和浩特 010010)

摘要: 统计学是一门通过信息的收集、整理、分析和描述统计过程等科学技术手段, 能够推断和测量对象信息本质以及预测未来发展趋势的综合性基础科学技术。近年来, 随着云技术、物联网等前沿技术的出现, 现代信息科学技术水平显著提高, 引发了新信息时代中大规模的信息处理需求, 对数据处理质量提出了更高的要求。传统的统计技术和方法已经无法适应新信息时代的需求, 因此人们不得不寻找适应新时代要求的新科学技术, 对数据的科学研究和创新变得更加重要。

在大数据分析时代, 许多组织正在广泛应用计算机技术进行统计工作。然而, 在大数据分析时代如何创新和完善统计学方法是统计学领域需要反思的一个重要问题。这意味着统计学需要与新技术相结合, 探索新的数据处理和分析方法, 以更好地满足大规模数据处理的需求。这可能涉及到使用先进的机器学习和数据挖掘技术, 发展更灵活、高效的统计模型, 以及更好地理解 and 解释大规模数据的特征和趋势。统计学需要在大数据分析时代进行创新和完善, 以适应新技术和新需求。这需要统计学领域的研究者和专家与其他学科合作, 探索新的方法和技术, 以使统计学继续发展并在新信息时代中发挥重要作用。

关键词: 大数据时代; 数据处理; 统计学

通常而言, 我们认为有关统计学理论的研究起源于古希腊的亚里士多德时代, 发展至今已有了 2300 多年的历史。统计学通过运用大量的数学以及其他部分学科的专业知识, 在社会科学和自然科学的各个领域被广泛运用。随着网络技术大幅提升、前沿科技不断涌现, 传统统计学从某种意义上而言已无法满足大数据时代对于数据更新迭代的速度要求, 因此便要在原有的基础上进行一定的创新与再发展。现如今最为势不可挡的便是大数据革命的席卷, 它对各领域的技术发展都或多或少的产生影响与冲击。因此, 本文对于大数据时代下对统计学的创新与发展进行一定的研究与探讨。

一、大数据时代统计学方法的创新

(一) 统计学的计算方式发生转变

利用同一台电子计算机进行大量数据的分析和数据处理, 是以往统计学在计算机中主要的应用过程, 其数据的保存能力则局限于计算机硬盘的大小。而在新的大数据时代, 由于科技的进步和发展, 以往这种简单的分类和存放模式已经开始被颠覆, 大数据的保存方式变得越来越简单, 保存性能大大增强, 数据的分类也不再局限于同一部电脑, 而且还能够利用云端储存技术, 以及大数据分析相关技术等把人们所需要的海量数据, 融入数据整理和分类的工作中。除此之外, 多台计算机联合, 把它们硬盘中保存的大数据在同一个计算机系统中加以分析和处理, 即通过多个存储设备为一个计算机系统的正常运行提供保障支持。

(二) 研究问题的方式发生变革

1. 数据预处理

随着网络的普及与发展, 数据的搜集显得更为方便。通过互联网的使用, 我们可以搜集到更多有关国内外的相关资料、书籍

和数据等。同时, 由于互联网的应用能够使人们运用更简便的方式获取更为广泛的数据, 故而大大提高了查找数据的速度, 并使得数据的质量发生质的突破。通过大数据应用对所获取数据进行预处理, 可以提升大数据技术的使用价值。在数据清洗、不完全数据填补以及数据纠正三大阶段有效利用大数据技术, 确保数据的有效性, 可以提升数据预处理的效率。同时, 有效将统计学与大数据技术进行结合, 可以及时更新数据库并将其相连接, 为数据的预处理工作提供有效帮助。

2. 数据抽样

在统计学中, 支持样本应包含所有的数据。而在大数据时代, 采用随机抽取混合数据样本的方式, 可以获取具有代表性的统计数据, 这种方式无须将所有数据都纳入统计范围, 大大提升了数据抽样的效率, 这一技术的实现主要依赖统计学相关知识, 从而保证其科学性、有效性, 并对数据进行严密控制。此外, 在大数据时代, 利用大数据技术可以缩短数据收集时间, 效率成倍增长, 在短时间内保证处理过程统一、简单, 这有效地解决了在数据收集过程中需面临时间周期过长、数据处理难以按照统一标准执行这一难题。

3. 数据整合与分析

通过大数据的相应变革, 让数据处理手段不再简单, 变得复杂而多元。压缩、转移、传输数据等手段提高了数据处理的效率, 且结果也更为精准。大数据技术在统计学中的运用, 有效地减少甚至是清除了复杂运算步骤中所产生的误差。另外, 在大数据时代还可以采用数据动态法, 以较为简单地构建数据模型。在现代大数据处理技术的帮助下, 大型数据库也能够被划分成若干小组, 将数据按照某种关系加以存储, 以便于大数据的检索和使用。

(三) 应用范围的不断扩大

随着数据来源的多样化, 利用采集到的数据可以衍生出很多新的数据。数据因此变得更为复杂, 能够通过某一个领域的数据对其他领域也产生影响, 不再局限于单一的领域。同样, 我国教育部在统计学学科设置方面新增“统计学+”交叉学科, 不断拓展统计学与其他领域的关系, 进而扩大了统计学的应用范围。

三、大数据时代统计学方法的发展

(一) 拓展统计学的研究体系

将大数据看作一个具有创新性、统一性、时代性的技术框架, 现如今, 统计学仅属于其中组成部分的一种, 将统计学融入到技术创新和学科研究中, 可以为统计学的发展提供全新的思路。在大数据时代背景下, 面对并存的机遇与挑战, 科研学者更需要在遵循可持续发展的基础上用辩证的眼光来对待其发展过程。从科研角度来看, 除严格遵守“以小见大”的计算原则, 获取较为完整的统计结果外, 更应以大数据思维为核心, 不断调整、优化数据收集、整理、分析与描述等方面的问题。

(二) 提出新的分类方法

传统上统计学在分类数据信息时, 需要先明确定义分类内容, 再进行有效分类, 并且收集过程需要按照定义的标签有针对性地进行。这种方式虽然可以更便捷地寻找所需内容, 但整个信息都

被固定在已有的框架当中,实际操作较为单一。在数据的预处理过程中,分类和梳理是其重要操作环节,人为因素会直接决定计算结果。在大数据的影响下,数据分析的形式以及数据来源都被大大拓宽,具有开放性、多源性、丰富性等特征,因此只需设定好信息类型就可以按照其关系和标识进行分类,在不同领域提取所需有效信息。值得注意的是,在研究期间所有信息数据的需求条件不是一成不变的,虽然获取数据的渠道和储存方式有很多,但不能按照统一标准来设定要求。在保留传统分类优势的基础上,将大数据的优势融入其中,不断进行创新优化,以满足新时代环境的要求。

(三) 重新整合统计学方法

包容性和融合性是大数据环境中的关键特征之一。传统统计学的研究方法在大数据环境下可能会遇到一些困难,无法进行创新。大数据在收集大量数据信息后,结合数据的变化规律,可以总结分析不同时间的状态和形式,从而推导出总体的发展趋势和方向,为不同领域的决策者提供有效依据,帮助他们做出更合理的决策。传统统计学主要利用具体的数据信息来总结和推断总体的特征,虽然也可以了解数据的动态发展,但个体获得的知识信息有限。因此,在方法整合方面,需要保留传统统计学中演绎推理的优势,同时结合大数据的优势,将单一的方法整合在一起,解决一个问题。这样可以提高分析的效率和质量,同时完全掌握数据信息的发展规律和变化趋势。总的来说,大数据的发展需要将传统统计学与大数据的优势相结合,以提高分析的效率和质量。通过综合利用不同的方法和技术,可以更好地理解数据的发展规律和趋势,并为决策提供准确和可靠的依据。

四、与大数据相关的统计学未来发展动态

(一) 在面对大数据质量问题时可采取的措施:

1. 数据源管理: 加强对数据源的管理,确保数据的来源可靠性和准确性。对于存在虚假数据或假数据的情况,可以通过数据验证和验证机制来识别和过滤掉不合格的数据。
2. 数据清洗和预处理: 在数据进入分析环节之前,进行数据清洗和预处理,包括去除重复数据、纠正错误数据、填补缺失数据等,以提高数据的一致性和准确性。
3. 数据质量评估: 建立数据质量评估指标体系,对数据进行定期评估和监控,发现和解决数据质量问题。可以使用数据质量度量指标,如准确性、完整性、一致性、时效性等来评估数据质量。
4. 数据安全和隐私保护: 在处理大数据时,要注重数据安全和隐私保护,采取合适的安全措施,如数据加密、访问控制、身份认证等,以保护数据的完整性和隐私。
5. 数据治理: 建立完善的数据治理机制,包括数据质量管理策略、数据管理流程和规范,明确数据的责任人和权限,确保数据的合规性和可追溯性。
6. 技术创新和人才培养: 积极探索和应用新技术,如人工智能、机器学习等,提高数据质量管理的效率和准确性。同时,加强对数据质量管理人员的培训和提升,提升他们的专业知识和技能。通过以上措施,可以改进和提高大数据质量。

(二) 数据要素的开发与应用研究

随着数据经济的崛起,对数据经济产业进行科学核算和培育大数据市场是非常重要的。如何将统计学中的数据进行进一步开发与研究,在此有以下几点分析:

1. 数据经济产业核算: 建立统一的数据经济产业核算体系,包括数据产业链、数据要素投入与产出、数据产业增加值等指标,

以科学评估和核算数据经济的规模和贡献。

2. 培育大数据市场: 建立健全的数据市场机制,包括政策支持、法律法规、市场准入和监管机制等,鼓励和促进数据的交易和共享,提高数据的可交易性和流通性。

3. 加强数据资源的价值实现: 通过数据分析和挖掘,开发数据新产品和服务,提供有价值的应用服务,实现数据资源的价值和变现,促进数据经济的发展。

4. 推动数据开放共享: 建立数据开放共享的机制和平台,鼓励各方主体主动分享数据资源,促进数据的交流与合作,推动数据的开放和共享,促进数据要素新动能的发挥。

5. 法律规范与产权交易: 加强数据法律法规的制定和完善,明确数据的产权和交易规则,为数据交易提供法律保障,进一步推动数据的整合和流通。

6. 人才培养和创新能力提升: 加强对数据经济相关人才的培养和引进,提高数据分析和应用技术的创新能力,为数据经济产业的发展提供人才和技术支持。

(三) 社会经济大数据计算研究

随着智能化技术的发展,新型数据的使用给社会科学研究者提供了许多新的机会和挑战。这些新型数据可以通过先进的机器学习和数据挖掘技术进行分析和理解,从而揭示出社会经济的新态势、探索新理论、发现异常和预测未来发展趋势等。这种研究方法被称为计算社会经济学,它将社会科学、经济学和计算机科学的理念和方法融合在一起。

计算社会经济学的应用领域非常广泛,可以扩展到司法领域中的法庭第三方数据证据的分析和认定,使用物联网数据进行决策科学化的应用,以及使用通信数据进行金融风险防控等。通过对大规模真实数据的研究,特别是与经济发展和社会问题相关的数据,可以开辟应用统计学研究的新领域。

这种新型数据的使用不仅可以提高数据的稠密性和实时性,还可以为多源数据的整合提供新的研究领域。它使社会科学研究者能够更好地理解社会经济的演变和变化,为社会发展提供更准确的预测和决策支持。同时,它也需要社会科学研究者不断学习和应用先进的技术,以更好地利用这些新型数据进行研究和分析。

五、结语

总之,计算社会经济学的发展是社会科学、经济学和计算机科学的深度融合,它为我们科技高速发展的大数据时代,统计学将面对着更多的机会和风险。在这样的一个大历史背景下,统计人员应建立高度自信,紧随社会发展,把科学统计方法和大数据有效融合,在科学的引导下创新和发展,以提高统计质量,从而发挥统计的更大价值。

参考文献:

- [1] 未良莉, 黄虹. 大数据时代经管类专业统计学课程教学改革思考 [J]. 合肥学院学报: 综合版, 2019 (1): 132-135.
- [2] 赵建霞. 大数据时代医院统计工作模式的改进思考 [J]. 办公室业务, 2019 (5): 29.
- [3] 直敏, 高天哲, 孙杨. 基于大数据分析的计算机信息处理技术应用 [J]. 计算机产品与流通, 2021 (2): 7-8.
- [4] 官汝锋, 刘姣, 丁文静, 等. 基于大数据背景探究应用统计学的创新 [J]. 数字化用户, 2019, 25 (10): 199+211.

作者简介: 云昊 (1975-), 女, 中级统计师, 中级经济师, 学历: 本科主要研究方向: 公共管理。