

DOI: 10.12361/2661-3263-06-06-141108

大数据背景下的企业统计学发展方向研究

普娅华

对外经济贸易大学统计学院, 中国·北京 100089

【摘要】在信息技术的飞速发展和数字化转型的浪潮下, 大数据正成为企业经营和决策的重要驱动力。企业通过收集、存储和分析大量的数据, 能够更深入地了解市场、客户和业务运作, 优化业务流程、提升效率、创新产品和服务。而在这个大数据的背景下, 统计学作为数据分析的核心工具, 具有着前所未有的重要意义。本文将探讨大数据背景下企业统计学的发展, 深入分析其意义、问题和挑战, 以期为企业在数据驱动的时代中发挥统计学的价值提供更清晰的指导和思路。从深化数据洞察、优化业务决策、应对挑战等多个维度, 我们将探讨企业如何在大数据时代充分发掘统计学的潜力, 为持续创新和可持续发展提供强有力的支持。

【关键词】大数据; 企业; 统计学

Research on the development direction of enterprise statistics under the background of big data

Yahua Pu

School of Statistics, University of International Business and Economics, China. Beijing 100089, China

[Abstract] With the rapid development of information technology and the wave of digital transformation, big data is becoming an important driving force for enterprise management and decision-making. By collecting, storing and analyzing vast amounts of data, companies can gain a deeper understanding of markets, customers and business operations, optimize business processes, improve efficiency, and innovate products and services. In this context of big data, statistics, as the core tool of data analysis, has unprecedented significance. This paper will explore the development of enterprise statistics in the context of big data, in-depth analysis of its significance, problems and challenges, in order to provide clearer guidance and ideas for enterprises to play the value of statistics in the data-driven era. From multiple dimensions such as deepening data insights, optimizing business decisions, and addressing challenges, we will explore how enterprises can fully tap the potential of statistics in the era of big data to provide strong support for continuous innovation and sustainable development.

[Keywords] Big data; Enterprise; Statistics

随着信息技术的飞速发展和互联网的普及, 大数据已经成为当今社会的一个重要特征。大数据的出现在各个领域引发了深刻的变革, 企业经营也不例外。企业在日常经营活动中产生大量的数据, 包括销售数据、用户行为数据、生产数据等, 这些数据蕴含着丰富的信息和价值。各种传感器技术的普及, 以及移动设备的广泛应用, 进一步增加了数据的产生速度和规模。

1 大数据背景下的企业统计学发展的重要意义

在当今大数据时代, 企业统计学的发展呈现出重要且广泛的意义。随着企业所面对的数据量急剧增加, 统计学的应用为企业转化海量数据为有价值的信息提供了有效手

段。统计学方法不仅可以揭示数据中隐藏的模式、趋势和关联, 更能够为决策制定提供准确的依据, 指导企业实现更加高效和明智的战略决策。企业统计学在市场竞争中的重要性不容忽视。通过深入分析市场数据, 企业能够更准确地洞察消费者需求、竞争态势等因素, 灵活调整产品策略和定价策略, 提升企业在市场中的竞争力。企业统计学还能帮助企业优化资源配置, 减少浪费, 提高效益。更为重要的是, 基于统计学的方法, 企业可以进行风险预测, 及早识别潜在的危机因素, 采取适当措施来降低风险带来的负面影响, 增强企业的抗风险能力。企业统计学也为个性化营销提供了支持。通过深入分析客户数据, 企业可以

精准洞察消费者的兴趣、购买习惯等，精心设计个性化的营销策略，提升客户满意度和忠诚度。企业统计学的发展对于创新至关重要。通过对数据的持续分析，企业可以及时发现市场机会和行业趋势，调整策略，实现创新的发展。政府和相关机构可以借助企业统计数据，制定更精准的政策和规划，推动社会经济的可持续发展。企业统计学的发展不仅对企业自身具有重要影响，还为国家和社会的整体发展提供了有力支持，为政府制定精准政策和推动可持续发展提供了数据依据。

2 大数据背景下的企业统计学发展存在的问题

2.1 数据隐私和安全问题

数据泄露导致用户隐私的暴露，进而影响用户的个人权益。如果企业未能有效保护用户数据，用户的个人信息被恶意利用，导致身份盗窃、金融诈骗等问题。这不仅给用户带来了财务损失，还影响用户的信任，对企业的声誉造成严重损害。数据泄露也导致企业的商誉受损。在信息时代，用户对于企业的信任是企业价值链中不可或缺的一环。一旦发生数据泄露事件，公众舆论和媒体的关注往往会影响企业的形象和声誉，导致客户流失、市场份额下降等问题，进而影响企业的长期发展。大数据的分析和挖掘也导致个人隐私的暴露。虽然大数据技术可以从海量数据中发现有价值的信息，但在这个过程中涉及到个人隐私的揭示。例如，通过分析用户的购物记录、搜索记录等，企业可以获得用户的个人喜好和兴趣，实现个性化推荐。然而，过度的数据分析使个人隐私暴露在公共领域，引发用户的担忧和不满。

2.2 数据质量不稳定

大数据往往涵盖多个来源，其中部分数据因为技术问题、人为操作或其他原因而缺失或不完整。数据缺失会导致统计分析中的样本不完整，进而影响分析的准确性和代表性。如果数据缺失不随机，导致样本偏差，影响结果的可信度。大数据源头众多，数据的录入、收集和转换环节容易引入错误。这些错误来自于人为输入错误、传感器故障、数据传输错误等。数据的不准确性会影响统计分析的准确性和可信度，甚至导致错误的决策。由于数据来源多样，不同来源的数据存在定义、格式、单位等方面的一致性。在进行统计分析时，如果没有进行适当的数据转换和标准化，不一致的数据导致分析结果的混乱和误解。在大数据环境中，数据会出现重复或冗余的情况，即相同或相似的数据在不同数据集中出现。这会导致统计分析中的样本重复，影响分析结果的精确性和准确性。大数据往往是从用户行为、社交媒体、传感器等多个渠道收集而来，因此受到采样偏倚的影响。特定群体的数据更容易被收集，而其他群体的数据被忽略，引入样本偏倚，影响分析结果的普遍性和适用性。大数据环境中数据更新迅速，但某些数据会过时，而分析过时数据导致不准确的结论。数据时效性也影响长期趋势的分析和预测。大数据中存在隐含的偏见，比如用户在社交媒体上的表达不够真实，数据

中存在主观性的因素。这些隐含偏见导致分析结果不够客观和准确。

2.3 多源数据整合困难

不同数据源使用不同的数据格式、结构和编码，导致数据难以直接整合。例如，一个数据源使用CSV格式，另一个数据源使用Excel格式，这就需要在整合过程中进行格式转换，增加了复杂性。不同数据源采用不同的数据标准和命名规范，导致数据元素名称不一致，相同概念的数据在不同数据源中有不同的名称，增加了数据整合的难度。不同数据源的数据质量存在差异，一些数据源更加准确和可靠，而其他数据源存在错误或不完整的数据，这影响到整合后数据的准确性。不同数据源采集数据的粒度不同，例如一个数据源以天为单位采集数据，另一个数据源以小时为单位采集数据，这就需要进行数据粒度的调整和对齐。不同数据源采集的数据涵盖不同的领域和维度，这使得在整合过程中需要确定哪些数据对于分析具有价值，哪些数据可以被舍弃。在整合多源数据时，涉及到数据安全和隐私的问题。不同数据源包含敏感信息，需要确保在整合过程中数据的安全性和隐私保护。多源数据是异步获取的，不同数据源的更新频率不同，导致数据整合后存在时间上的差异，需要解决数据同步的问题。在整合多源数据时，存在重复的数据和记录，需要进行数据去重和匹配，以确保数据的唯一性和一致性。数据整合过程需要使用专门的技术和工具来处理不同数据源的数据，不同数据源之间的集成和整合需要投入更多的技术和资源。

3 大数据背景下的企业统计学优化发展策略

3.1 提高数据隐私和安全保障水平

企业应该制定明确的隐私政策和数据使用准则，明确数据收集和使用的范围、目的和方式。隐私政策应该简明扼要、易于理解，并确保用户在数据使用中有明确的权利和选择。要确保企业的数据处理行为符合相关法律法规，遵守数据保护和隐私合规标准。采用数据加密技术，对敏感数据进行加密存储和传输，确保即使数据被泄露，也难以解读。例如，在政府统计学实施过程中，已知M/M/1/K队列的数据流到达强度 λ 与服务强度 μ ，则排队系统队长概率表示为：

$$P'_k = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{K+1}} * \left(\frac{\lambda}{\mu}\right)^k \quad 0 \leq k \leq K$$

可以得到队列空闲概率 $P'_0(n)$ 和阻塞概率 $P'_{\text{block}}(n)$ 。

$$P'_0(n) = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{K+1}}$$

$$P'_{\text{block}}(n) = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{K+1}} * \left(\frac{\lambda}{\mu}\right)^k$$

得到的队列空闲概率 $P'_0(n)$ 与传输成功概率 P_s 可以得到以活动节点数为状态变量的 Markov 链转移概率矩阵 P 。由转移概率矩阵可以得到 Markov 链的稳态分布 $\pi(n) = \pi(n) \cdot P$ 。由于转移矩阵是有限状态的和齐次不可约的所

以其稳定分布一定存在。由此可见,企业可以在大数据技术的帮助之下有效提升数据因素保护水平。建立严格的访问控制机制,限制员工和系统管理员对数据的访问权限,确保只有授权人员可以访问敏感数据。采用多因素身份认证等技术,确保用户在访问企业平台时进行有效的身份验证,防止非法访问。限制员工的系统权限,将最小化原则应用于数据处理,减少内部人员的风险。建立健全的数据安全管理体系,包括数据安全策略、风险评估、漏洞修复、紧急响应计划等。定期进行安全风险评估,识别潜在的风险和威胁,及时采取措施进行修复和预防。为员工提供有关数据隐私和安全的培训,增强其对数据安全的意识和理解。

3.2提高数据质量的稳定性

建立严格的数据采集流程,确保数据来源的可信度和准确性。对采集的数据进行清洗,识别和处理数据中的错误、缺失和重复项。制定数据采集和清洗的标准操作规程,确保数据的一致性和准确性。统一数据格式、单位和命名规范,避免不同数据源的数据不一致问题。将不同来源的数据进行标准化和转换,确保数据在进行统计分析前具有一致性。例如,在公式:

$$P_{fd}(n) = P_{block} + (1 - P_{block}) * P_{drop}$$

其中队列拥塞概率 P_{block} 由式(3-15)求得。超出最大允许重传次数导致数据包丢弃的概率 P_{drop} 表示为:

$$P_{drop} = (1 - P_s) * (1 - P_r * P_s)^{R_{max}} \\ = \sum_{n=1}^M P_{fd}(n) * \pi(n)$$

由此可见,大数据技术在统计学领域的应用有助于数据质量和稳定性的提升。采用数据验证算法和方法,自动检测数据中的错误和异常值。引入数据校验环节,对数据进行逻辑性和合理性的验证,保证数据的准确性。建立数据质量监控系统,实时追踪数据的质量指标,发现和解决数据问题。设置数据质量阈值,当数据质量指标超过阈值时,及时进行预警和处理。应用数据质量评估工具,对数据进行定量和定性的质量评估,识别数据质量问题。利用统计分析方法,分析数据质量问题的影响程度和原因。对数据来源进行验证和审查,确保数据提供者的可信度和合法性。针对重要数据,可以进行独立的第三方验证,增加数据的可信度。制定数据治理策略,明确数据质量的责任和流程,确保数据质量问题得到有效解决。建立数据质量管理团队,负责监督数据质量,及时响应和处理问题。培训员工关于数据质量的重要性和处理方法,提高数据采集和处理的意识。

3.3提高多源数据整合水平

定义统一的数据标准和命名规范,确保不同数据源采集的数据在命名、单位、格式等方面保持一致。制定数据质量标准,要求每个数据源在采集、存储和传输数据时遵

循一定的质量标准。在数据整合前,对每个数据源进行数据清洗和预处理,包括去除重复数据、填充缺失值、纠正错误数据等,以提高数据质量和一致性。使用专门的数据集成工具和平台,可以帮助企业更有效地整合多源数据。这些工具可以自动化数据转换、匹配和整合过程,减少手动操作,提高效率。开发数据映射和转换规则,将不同数据源中的数据映射到统一的数据模型中,确保数据元素之间的对应关系,实现数据的整合。建立数据集市或数据仓库,将不同数据源的数据集中存储,提供统一的数据访问接口和查询工具,方便用户进行数据分析和查询。制定数据同步和更新策略,确保不同数据源的数据能够及时同步和更新,保持数据的一致性和最新性。在整合过程中要注重数据安全和隐私保护,对敏感数据进行加密和脱敏处理,确保数据在传输和存储过程中不被泄露。建立数据质量监控机制,定期对整合后的数据进行质量评估,发现并解决数据质量问题,保障统计分析的准确性。在数据整合过程中,需要跨不同部门和团队之间进行协作和沟通,确保数据整合的顺利进行,消除信息孤岛。为涉及数据整合的人员提供培训和技术支持,使他们掌握数据整合工具和技术,提高数据整合的能力。

结论

统计学作为一门研究数据的科学,正发挥着日益重要的作用。统计学通过收集、整理、分析和解释数据,帮助企业从数据中发现有价值的信息,指导企业的战略和决策。在大数据时代,统计学不仅可以帮助企业更好地理解数据,还可以从中发现隐藏在数据背后的规律和趋势,为企业提供更准确的指导。然而,虽然大数据为企业统计学的发展带来了新的机遇,但也存在一些问题和挑战。大数据的规模和复杂性使得数据的整理和清洗变得更加困难,需要企业投入更多的人力和资源。大数据中存在着噪声、误差和不完整性,如何在这些干扰因素下进行有效的统计分析也是一个难题。随着隐私和安全问题的凸显,如何在保护用户隐私的前提下进行数据分析也是一个值得思考的问题。因此,有必要深入研究企业在大数据时代中如何充分发挥统计学的优势,应对各种问题和挑战,推动企业在数据驱动下取得更大的成功和发展。

参考文献:

- [1] 姬丽. 大数据背景下的企业统计学发展方向分析[J]. 现代商业, 2021, (29): 157-159.
- [2] 尹良喜. 煤炭火工企业统计学与信息学融合的技术探讨[J]. 淮北职业技术学院学报, 2011, 10(03): 67-69.
- [3] 南英子. 企业统计学课程教学改革方向探讨[J]. 教育教学论坛, 2011, (05): 201-202.

作者简介:

普娅华(1989.1—),女,汉族,云南省昆明市,学历本科,中级经济师,对外经济贸易大学统计学院在职人员高级课程研修班学员,研究方向为统计学大数据科学与应用方向。