

大数据时代统计数据质量提升路径的研究

尹小龙

湖南医药学院第一附属医院, 湖南 怀化 418000

摘要： 本文聚焦大数据时代统计数据质量提升路径。阐述了其研究背景与目的，强调数据质量对决策的关键影响及衡量指标。分析了面临的数据来源、观念方法及人员素质等挑战，针对性提出加深认识、加强制度保障、建设信息资源平台、完善体制机制及培养复合型人才等策略，旨在全面提升统计数据质量，适应大数据时代需求。

关键词： 大数据时代；统计数据质量；提升路径；数据来源

Research on the Path of Statistical Data Quality Improvement in the Era of Big Data

Yin Xiaolong

The First Affiliated Hospital of Hunan University of Medicine, Huaihua, Hunan 418000

Abstract： This paper focuses on the improvement path of statistical data quality in the era of big data. Its research background and purpose are expounded, emphasizing the key influence of data quality on decision and measuring indicators. This paper analyzes the challenges of data sources, concepts and methods and personnel quality, and puts forward strategies of strengthening understanding, strengthening institutional guarantee, building information resource platform, improving system and mechanism and cultivating compound talents, aiming at comprehensively improve the quality of statistical data and meet the needs of the era of big data.

Keywords： big data era; statistical data quality; improvement path; data source

引言

在当今大数据时代，数据量呈爆炸式增长，统计数据质量的重要性愈发凸显。然而，也面临诸多挑战，如何提升其质量成为关键问题。本研究旨在探索有效的提升路径，以适应时代发展需求，为决策等提供更可靠的数据支撑。

一、大数据时代统计数据质量的重要性

（一）对决策的影响

准确的统计数据在企业和政府决策中起着至关重要的作用，是决策的重要依据。在大数据时代，海量的数据为决策提供了更多的信息支持，但同时也对数据质量提出了更高的要求。

1. 企业决策依据

以阿里巴巴为例，通过大数据分析消费者行为、购买偏好和市场趋势，为商家提供个性化推荐、营销策略等支持，推动电商行业快速发展。准确的统计数据能够帮助企业了解市场需求、竞争对手情况以及自身经营状况，从而制定出更加科学合理的经营策略^[1]。例如，企业可以通过分析销售数据，了解产品的畅销程度和市场占有率，进而调整产品结构和生产计划；通过分析客户反馈数据，改进产品质量和服务水平，提高客户满意度和忠诚度。

2. 政府决策依据

政府依据统计数据进行规划和政策制定。例如，美国麻省理工学院的“10亿价格项目”，通过“网络抓取技术”，利用网上

购物交易数据计算日常通胀指数，为政府制定经济政策提供了重要依据。在国内，厦门市统计局依托厦门市电子政务内外网和政务云平台打造的“厦门市统计经济综合分析平台”，为政府决策提供可靠依据^[2]。政府可以通过统计数据了解经济运行状况、社会发展趋势以及民生需求等，从而制定出更加科学合理的政策措施，促进经济社会的协调发展。

（二）数据质量的衡量指标

数据质量的衡量指标包括准确性、完整性、一致性和及时性等，这些指标对于保证数据质量具有重要意义。

1. 准确性的关键作用

数据准确是避免错误决策的关键。在医疗保健领域，不准确的数据可能会导致医生做出错误的诊断和治疗方案，直接影响患者的健康和生命安全。在企业经营中，不准确的销售数据可能会导致企业制定错误的营销策略，影响企业的经济效益^[3]。因此，保证数据的准确性是提高统计数据质量的首要任务。

2. 完整性与一致性

数据的完整和一致对反映真实情况起着重要作用。完整性要

作者简介：尹小龙（1973.07-），男，汉族，湖南邵东市，大学本科，统计师，研究方向为：卫生统计。

求统计数据应覆盖全部的总体或能体现其代表性的样本，不应有遗漏或缺失的情况。缺乏完整性的数据会导致分析和结论的误差。一致性要求统计数据应在不同时间点、不同数据源或不同统计单位之间保持一致。一致性表明数据的稳定性和可靠性。例如，在企业财务决策中，数据的完整性和一致性对于准确评估企业的财务状况和经营成果至关重要。如果财务数据存在遗漏或不一致的情况，可能会导致企业做出错误的财务决策，影响企业的生存和发展。

二、大数据时代统计数据质量面临的挑战

（一）数据来源问题

在大数据时代，数据来源的广泛性和复杂性给统计数据质量带来了巨大挑战。企业部分统计指标的来源往往难以准确把握^[4]。以工业企业为例，工业总产值的统计就存在诸多难题。部分企业在填报工业总产值时，可能由于对统计口径理解不准确、会计核算方法不一致等原因，导致填报数据错误。一些企业将半成品价值重复计算，或者未将自制设备等纳入总产值统计范围，使得工业总产值数据不能真实反映企业的实际生产情况。而且，除了工业总产值，其他统计指标如原材料消耗、能源使用等的来源也不好掌握。企业的生产经营环节众多，数据分散在各个部门和系统中，缺乏统一的标准和规范进行收集和整理，这就增加了数据来源的不确定性和复杂性。

（二）观念与方法落后

统计观念和方法的落后严重影响了统计数据质量。一些统计人员仍然停留在传统的统计思维模式中，对大数据的认识不足，导致统计工作的准确性和及时性下降。他们习惯于采用旧的调查方法和统计手段，不能充分利用大数据技术获取更全面、准确的数据^[5]。例如，在进行市场调研时，仍然依赖传统的问卷调查方式，而忽视了网络数据、社交媒体数据等丰富的数据源。这种观念的落后使得统计数据难以跟上市场变化的节奏，无法及时为决策提供有效的支持。同时，统计方法的落后也导致数据质量低，统计职能弱化。一些统计分析方法过于简单，不能深入挖掘数据背后的规律和关系，无法为企业和政府提供有深度的决策建议。而且，传统的统计方法在处理海量、多样的数据时显得力不从心，容易出现数据处理不完整、分析结果不准确等问题。

（三）人员素质问题

基层统计人员的素质问题也是影响统计数据质量的重要因素。部分基层统计人员业务素质偏低，缺乏系统的统计知识和专业技能培训。他们在数据收集、整理和分析过程中，可能会出现操作失误、计算错误等问题^[6]。例如，在进行数据录入时，由于粗心大意导致数据录入错误；在进行数据计算时，使用错误的公式或方法，使得计算结果偏差较大。此外，一些基层统计人员不会合理运用数据，仅仅是简单地收集和上报数据，而不能对数据进行深入分析和解读，无法为企业和政府提供有价值的信息。他们对数据分析软件和工具的掌握程度也不够，不能充分利用先进的技术手段提高数据处理效率和质量。这就使得统计数据在基层

环节就存在质量隐患，进而影响到整体数据的准确性和可靠性。

三、提升统计数据质量的策略

（一）加深认识

强调各级统计机构应重视大数据应用，营造良好环境。

1. 克服恐惧顺应转变

在大数据时代，各级统计机构和管理部門不应畏惧大数据带来的变革，而应积极克服对大数据的“恐惧症”，顺应“信息资源”向“信息资产”的转变。据统计，目前已有超过70%的统计机构开始尝试将大数据应用于统计工作中，但仍有部分机构因对大数据的不了解而持观望态度。各级统计机构应认识到大数据的发展是不可阻挡的趋势，只有积极主动地适应新形势，才能在统计工作中抢占先机。

2. 重视大数据统计工作

重视大数据统计工作对于提升统计数据质量具有重要意义。大数据的应用可以拓宽数据来源，提高数据的准确性和及时性。通过对海量数据的分析，可以发现传统统计方法难以察觉的规律和趋势，为决策提供更有价值的参考^[7]。例如，在市场调研中，利用大数据分析消费者的行为和偏好，可以更准确地了解市场需求，为企业制定营销策略提供依据。同时，重视大数据统计工作还可以营造宽容的统计研究应用环境，鼓励创新，推动统计工作不断发展。

（二）加强制度保障

论述建立健全统计工作制度，规范数据应用。

1. 完善规范与制度

制定相关管理办法和要求对于保障大数据统计工作的标准规范协调性至关重要。例如，建立大数据采集标准体系，可以规范数据采集的方式、流程和技术路线，确保数据的可靠性和可比性。据不完全统计，目前已有部分地区制定了大数据统计相关的规范与制度，取得了良好的效果。另外完善统计业务的分类与指标设置，也可以提高数据的准确性和完整性。

2. 加强指标设置与培训

优化统计指标和加强业务培训对提高数据质量起着重要作用。合理的统计指标可以更准确地反映经济社会现象，为决策提供有力支持。同时，加强对统计人员的业务指导与培训，可以提高他们对统计指标的理解和运用能力，确保数据的准确采集和处理^[8]。例如，通过开展定期的培训课程和研讨会，可以让统计人员及时了解最新的统计方法和技术，提高他们的业务水平。

（三）建设信息资源平台

1. 平台架构与功能设计思路

以医疗系统为例，要建设好统计大数据信息资源平台，总体架构应当采用模块化设计，以云计算、大数据、人工智能等先进技术为支撑，构建起一个涵盖医疗卫生机构各领域的数据整合与分析体系。平台应具备全面且强大的功能，其核心是医疗数据集成与共享平台，它能无缝对接医院信息系统等各类医疗信息系统，实现数据的高效采集和整合。例如，在医院数据监测方面，

平台可以实时获取各级医院的门诊和住院病例信息，包括患者的基本信息、症状表现、诊断结果、检查检验数据等多维度信息，通过标准化的数据格式转换和校验，形成完整的疾病监测数据集。同时，平台还可以设有医疗数据分析与预测模块，运用机器学习算法对大量的医疗数据进行深度挖掘和分析，为卫生健康管理和决策提供科学依据。

2. 数据质量管控与标准化

关于信息平台的数据审核机制，应建立严格的数据质量管控体系，从数据采集源头抓起，制定详细的数据采集规范和标准^[9]。例如，对于病历数据的录入，规定统一的格式和必填项，确保数据的完整性和准确性。同时，平台可以利用人工智能技术对采集到的数据进行实时校验和质量评估，自动识别和纠正数据中的错误和异常值。在数据标准化方面，平台应当依据国际和国内的医疗卫生行业标准，建立统一的数据字典和编码体系，对疾病诊断、药品名称、医疗服务项目等进行标准化编码，确保不同医疗机构的数据具有可比性和一致性。此外，平台还应定期对数据进行质量审核和反馈，督促医疗机构改进数据质量，从而为卫生统计工作提供高质量的数据保障。

（四）完善信息化体制机制

讨论结合大数据要求，完善统计信息化体制机制。

1. 适应技术发展趋势

随着数字政府的推进和 5G、人工智能、物联网等技术的应用，数据采集、处理和展示呈现智能化、便携式和移动化的趋势。统计机构应建立相应的体制机制，适应这些技术发展趋势。例如，加快统计大数据信息资源平台的建设、完善运维和管理体制，确保平台的高效运行。同时，利用“云端集中”的优势，实现数据的集中管理和共享，提高数据的利用效率。

2. 创新统计工作生态圈

大数据为统计工作带来了新的机遇，如开源互联网思维以及众创、众包、众扶、众筹平台的建设等，为大数据统计工作和大数据的统计应用发展提供了良好的“生态圈”。统计机构可以充分利用这些新机遇，创新统计工作方式和方法。例如，通过众包平台，邀请社会公众参与数据采集和分析，提高数据的准确性和及时性。同时，大数据的共享、开发应用以及大数据统计产品等也为统计数据质量管控打开了新的空间。

（五）培养复合型人才

强调培养既懂统计业务又掌握信息技术的人才。

1. 内部培养与外部引进

内部培养和外部引进相结合是培养复合型人才的有效方式。通过内部培训，可以提高现有统计人员的信息技术水平，使其更好地适应大数据时代的统计工作要求^[10]。同时，引进大数据、云计算等信息化领域专业人才，可以为统计机构带来新的思想和技术。例如，一些高校和科研机构已经开设了大数据统计相关的专业课程，为统计机构培养了一批高素质的专业人才。

2. 合作培养创新人才

与互联网企业、科研院所、高校、社会团体等合作培养创新人才具有重要意义。通过学术研究、技术合作和应用交流，可以

不断培养创新型人才，为统计工作的发展提供人才支持。例如，湖南大学金融与统计学院提出“以大数据思维为导向，能力培养为核心”的“三维同步”培养模式，构建复合型创新人才培养体系，取得了良好的效果。

四、结论与展望

在大数据时代，统计数据质量面临着诸多挑战，但通过采取一系列有效的策略，可以显著提升统计数据质量。我们强调了各级统计机构应加深对大数据的认识，克服恐惧顺应转变，重视大数据统计工作，营造良好的统计研究应用环境。加强制度保障，完善规范与制度，加强指标设置与培训，确保大数据统计工作的标准规范协调性和数据的准确性、完整性。建设信息资源平台，采用先进技术架构，实现数据质量管控，提高数据采集效率。完善信息化体制机制，适应技术发展趋势，创新统计工作生态圈。培养复合型人才，通过内部培养与外部引进相结合、合作培养创新人才等方式，为统计工作提供有力的人才支持。

这些策略对于提升大数据时代统计数据质量具有重要意义。准确可靠的统计数据是企业 and 政府决策的重要依据，能够帮助企业了解市场需求、制定经营策略，为政府规划和政策制定提供支持，促进经济社会的协调发展。

未来可进一步深入研究大数据技术在统计数据质量提升中的具体应用和创新模式。探索如何更好地整合多源数据，提高数据的融合度和可用性。加强对统计数据质量评估体系的研究，建立更加科学、全面的评估指标和方法。关注人工智能等新兴技术与统计工作的深度融合，以及如何利用这些技术进一步优化统计流程和提高数据质量。同时，持续关注人员素质提升的长效机制，为大数据时代统计工作的发展提供坚实的人才支撑。

参考文献

- [1] 廖云辉, 陈华. 大数据背景下统计数据质量影响因素分析[J]. 大众标准化, 2022,(02):137-139.
- [2] 姬亚敏. 大数据背景下提升政府统计数据质量路径研究[D]. 郑州大学, 2021.
- [3] 秦宏卫, 王小宏. 大数据提升统计数据质量[J]. 中国农村金融, 2015,(18):53-54.
- [4] 张海亮. 大数据背景下统计数据质量控制方法研究[J]. 生产力研究, 2022,(05):156-160.
- [5] 宋书现. 大数据时代公民数据素养的提升路径探赜[J]. 产业与科技论坛, 2020(10):2.
- [6] 赵东升, 姜蒙, 彭婷婷. 大数据时代计量管理质量提升路径研究[J]. 中文科技期刊数据库(全文版)工程技术, 2021(12):3.
- [7] 陈方, 孙文超, 王松. 企业物料数据质量提升路径研究[J]. 科技视界, 2021(22):2.
- [8] 黄小敏. 大数据时代促进统计调查高质量发展研究[J]. 中小企业管理与科技, 2020(20):2.
- [9] 周林兴, 林凯. 大数据时代档案数据质量控制: 现状、机制与优化路径[J]. 档案与建设, 2022(2):4-8.
- [10] 王凯玲, 邵竞莹. 大数据时代高校统计工作优化路径探析[J]. 活力, 2022(12):178-180.