

数字化背景下城镇住户调查数据质量控制办法研究

孙 丽

辽宁省铁岭市清河区统计事务服务中心, 中国·辽宁 铁岭 112003

【摘 要】伴随数字化进程加速, 城镇住户调查模式发生根本性转变。数字化在提升调查效率与数据丰富度的同时, 也引发数据质量控制的新难题。本文系统剖析数字化对城镇住户调查数据质量的双重影响, 深入探讨当前数据质量控制面临的关键挑战, 从样本设计、数据管理、审核体系等多个维度提出优化策略, 旨在构建科学有效的数据质量控制体系, 为城镇住户调查工作提供有力支持。

【关键词】数字化; 城镇住户调查; 数据质量控制

1 引言

城镇住户调查是获取居民生活状况、经济活动信息的重要途径, 其数据质量直接关系到政府政策制定的科学性与有效性。在数字化时代, 信息技术深度融入调查工作, 电子记账、数据在线传输等技术的应用, 改变了传统调查模式, 显著提升了调查效率。然而, 数字化也带来了数据安全、审核难度加大等新问题, 数据质量控制面临新的挑战。深入研究数字化背景下城镇住户调查数据质量控制办法, 对提高统计数据质量、增强政府决策科学性具有重要意义。

2 数字化对城镇住户调查数据质量的影响

2.1 积极影响

数字化技术的应用极大地提高了数据采集效率。传统的纸质记账方式需人工收集、整理, 周期长且效率低。而电子记账借助专门的 APP 或软件, 调查户可随时随地记录收支信息, 数据能实时上传至调查系统, 大大缩短了数据采集时间, 使调查人员能够将更多精力投入到数据质量把控上。

电子记账系统具备自动编码、逻辑校验等功能, 能有效减少人为误差。在数据录入过程中, 系统可自动对收支项目进行分类编码, 并依据预设规则对数据进行初步审核, 及时发现和提示不符合逻辑关系的数据, 从而提高数据的准确性。

数字化调查手段拓宽了数据采集的维度。除传统收支数据外, 还能获取消费时间、地点、支付方式等信息, 为深入分析居民生活行为和消费模式提供了丰富的数据基础, 有助于挖掘数据背后的深层信息。

2.2 消极影响

数字化环境下, 数据安全面临诸多风险。数据以电子形式存储和传输, 容易成为网络攻击的目标。网络黑客可能窃取或篡改数据, 计算机病毒、恶意软件也可能导致数据丢失或损坏, 威胁数据的完整性和真实性。

部分调查对象对数字化操作存在抵触心理。一些年龄较大、文化程度较低的调查户, 因不熟悉电子记账设备和软件的操作, 影响记账的积极性和准确性。同时, 出于对个人隐私泄露的担忧, 部分调查对象对调查工作的配合度降低。

随着数据量的剧增, 数据审核工作面临巨大压力。传统的人工审核方式难以适应海量、复杂的数据审核需求。电子记账数据的格式和结构与纸质账本不同, 隐藏在其中的错误和异常情况更难被及时发现, 对审核算法和技术手段提出了更高要求。

3 当前数字化背景下城镇住户调查数据质量控制面临的挑战

3.1 调查样本代表性问题

在实际操作中, 受数字化系统更新难度、调查成本等因素制约, 样本轮换往往难以按时进行。样本老化使得调查数据无法准确反映城镇人口结构、经济活动等方面的变化, 影响调查结果的准确性和有效性。

数字化抽样过程中, 由于技术故障或抽样方法的局限性, 容易导致样本结构出现偏差。例如, 过度依赖网络抽样可能会遗漏网络接入不足或使用频率低的群体, 使得样本不能全面涵盖各类城镇住户, 降低了样本的代表性。

3.2 数据录入与传输问题

尽管电子记账系统具有一定的自动校验功能, 但调查户

在数据录入过程中仍可能出现错误,如金额输入错误、收支项目类别选择错误等。部分调查户对记账要求理解不透彻,也会导致数据录入不准确,影响数据质量。

在数据传输过程中,网络故障、系统不稳定等因素可能导致数据传输中断或丢失。特别是在网络信号不佳的地区,或调查系统进行维护升级时,数据传输风险更高,严重影响数据的及时性和完整性。

3.3 数据审核体系不完善

现有的数据审核规则多侧重于常见的逻辑错误和简单的数值范围判断,对复杂经济行为和数据间的深层次关联考虑不足。对于新兴消费模式,如共享经济消费、虚拟商品消费等,缺乏相应的审核规则,难以有效识别数据的合理性。

部分地区在数据审核中仍以人工审核为主,缺乏先进的数据分析技术和软件工具支持。人工审核效率低且易出现疏漏,与发达国家广泛应用大数据分析、人工智能等先进技术进行数据审核相比,存在明显差距,难以满足数字化背景下的数据审核需求。

4 数字化背景下城镇住户调查数据质量控制办法

4.1 优化调查样本设计

科学合理的样本设计是保障调查数据代表性的基础。利用数字化技术构建动态监测系统,能够实时追踪城镇住户在人口结构、经济活动等方面的变化特征。该系统可整合多源数据,如人口普查数据、就业统计数据等,通过数据分析模型,及时发现住户特征的变动趋势。基于此,依据人口流动、产业结构调整等关键指标,灵活调整样本轮换计划,确保样本始终契合城镇住户的实际构成。同时,定期对样本的代表性进行量化评估,运用统计学方法分析样本与总体在关键变量上的差异,一旦发现样本出现偏差,及时替换不符合当前城镇住户结构特征的样本单元,保证调查数据能够精准反映总体情况。

分层抽样是提高样本均匀性和代表性的有效方法。按照城镇住户的收入水平、年龄结构、职业类型等多个维度进行细致分层,确保各层内住户具有相似的特征,层与层之间存在显著差异。在各层中按照既定比例抽取样本,使样本能够覆盖不同收入层次、不同年龄段、不同职业类型的住户群体。结合数字化手段,将调查户的基本信息与政府相关部门数据库、社会经济数据平台进行比对,实现样本的精准匹配。通过这种方式,有效避免样本偏差,确保调查数据能够全面、均衡地反映各类城镇住户的生活状况和

经济行为。

4.2 加强数据录入与传输管理

数据录入环节的准确性直接影响数据质量。电子记账软件的数据录入界面设计应遵循简洁直观、易于操作的原则。对于容易混淆的收支项目,提供详细且通俗易懂的提示信息 and 操作示例说明,帮助调查户准确理解各项收支分类标准。同时,增强软件的实时校验功能,在调查户录入数据时,系统自动对数据格式、数值范围、逻辑关系等进行即时检查。一旦发现数据异常,如金额为负数、支出明显超出合理范围等情况,系统立即弹出错误提示,并提供修改建议,引导调查户及时修正错误,从源头上降低数据录入错误率。

数据传输的稳定性和可靠性对数据的完整性和及时性至关重要。在数据传输过程中,采用多重备份技术,在调查户的记账设备本地进行数据存储备份,同时在数据传输的中间节点设置缓存和备份机制。建立完善的数据传输备份与恢复机制,明确数据传输故障时的应急处理流程和责任分工。一旦出现网络故障、系统崩溃等情况,能够迅速启动备份数据恢复程序,确保数据不丢失。加大对网络基础设施建设的投入,优化网络布局,提升网络信号覆盖范围和强度,同时加强调查系统的稳定性建设,定期进行系统维护和升级,降低数据传输风险,保障数据能够及时、完整地上传至调查系统。

4.3 完善数据审核体系

全面、科学的审核规则是数据质量审核的核心依据。随着城镇住户经济行为日益多样化和复杂化,数据审核规则需要不断拓展和完善。除了传统的逻辑关系审核,如收入与支出的平衡关系、各类收支项目的占比合理性等,还需针对新兴经济现象和消费模式,如共享经济消费、虚拟商品消费、数字金融交易等,制定专门的审核规则。明确这些新兴领域数据的合理范围、计算方法和逻辑关系,确保审核规则能够覆盖所有可能出现的数据类型和经济行为。建立审核规则动态更新机制,根据经济社会发展变化和审核实践经验,定期对审核规则进行修订和完善,保证审核规则始终具有时效性和有效性。

引入先进的数据分析技术是提升数据审核效率和准确性的关键。大数据分析技术能够对海量的历史住户调查数据和相关经济数据进行深度挖掘和分析,通过建立数据模型,发现数据中的潜在规律和异常模式。例如,利用时间

序列分析模型预测住户收支的正常波动范围,通过聚类分析方法识别具有相似消费行为的住户群体,从而快速定位异常数据。人工智能的机器学习算法可以自动学习数据特征和审核经验,不断优化和调整审核规则。通过对大量历史数据的学习,机器学习模型能够识别出人工难以发现的复杂数据模式和异常情况,提高审核的精准度和效率。将大数据分析和人工智能技术与传统审核方法相结合,形成多层次、智能化的数据审核体系。

4.4 强化调查人员与调查对象培训

调查人员是数据质量控制的关键执行者,其专业素养和业务能力直接影响调查工作的质量。定期组织调查人员参加数字化调查技术和数据质量控制培训,培训内容涵盖电子记账系统的操作技巧、数据审核的方法和标准、调查样本的管理和维护等方面。邀请行业专家学者和具有丰富实践经验的一线工作人员进行授课,通过理论讲解、案例分析、实际操作演练等多种方式,提高调查人员对数字化调查技术的掌握程度和数据质量控制能力。建立调查人员交流平台,鼓励调查人员分享工作经验、交流遇到的问题和解决方案,促进调查人员之间的相互学习和共同进步。同时,制定调查人员考核制度,定期对调查人员的工作表现和业务能力进行评估,将考核结果与绩效挂钩,激励调查人员不断提升自身素质。

调查对象是数据的直接提供者,其配合度和记账能力对数据质量有着重要影响。针对不同类型的调查对象,采用差异化的培训方式。对于熟悉数字化操作的年轻调查对象,通过线上视频教程、在线学习平台、社交媒体群组等方式,提供便捷的培训资源和及时的在线答疑服务。对于年龄较大、操作困难的老年调查对象,组织调查人员进行上门一对一指导,帮助他们熟悉电子记账设备和软件的操作流程。培训内容不仅包括电子记账软件的使用方法,还着重讲解住户调查的重要意义、记账要求和注意事项,提高调查对象对调查工作的认识和重视程度。通过发放通俗易懂的培训资料,如图文并茂的记账指南、操作手册等,辅助调查对象学习和掌握记账知识,增强调查对象的记账能力和配合意愿。

4.5 建立数据质量监督与评估机制

内部监督是保障调查工作各环节数据质量的重要防线。在调查机构内部建立严格的数据质量监督制度,成立专门的数据质量监督小组,明确监督小组的职责和权限。

监督小组定期对调查数据的采集、录入、审核等环节进行全面检查,采用随机抽样、重点抽查等方式,对调查数据进行质量评估。制定详细的数据质量检查标准和评估指标体系,对数据的真实性、完整性、准确性进行量化评分。对于发现的数据质量问题,及时进行问责,要求相关责任人限期整改,并跟踪整改情况。建立数据质量问题反馈机制,将检查过程中发现的问题和改进建议及时反馈给相关部门和人员,促进调查工作流程的优化和数据质量的持续提升。

引入第三方机构对城镇住户调查数据质量进行评估。第三方机构凭借其独立性和专业性,从客观角度对样本代表性、数据准确性、调查流程规范性等方面进行全面评价,并出具详细的评估报告,提出改进建议。调查机构根据评估报告,针对性地调整数据质量控制措施。

5 结论

数字化为城镇住户调查带来机遇的同时,也使数据质量控制面临诸多挑战。通过分析数字化对数据质量的影响和当前面临的问题,本文提出了一系列涵盖样本设计、数据管理、审核体系、人员培训以及监督评估等方面的数据质量控制办法。这些办法相互关联、相互支撑,构成了一个完整的数据质量控制体系。在实际工作中,各地区应结合自身实际情况,灵活运用这些办法,不断提升城镇住户调查数据质量,为政府决策和社会发展提供可靠的数据保障。随着数字化技术的不断发展,城镇住户调查数据质量控制工作也需要持续创新和完善,以适应新的形势和要求。

参考文献:

- [1] 刘萍. 提高住户调查数据质量的研究[J/OL]. 中文科技期刊数据库(全文版) 社会科学, 2022(10) [2022-10-01].
- [2] 夏雪. 关于住户统计调查数据质量控制方法和应用研究中出现的问题及建议[J]. 市场调查信息(综合版), 2021, (8): 1-1.
- [3] 赵彦云. 大数据赋能政府统计现代化[J]. 中国社会科学评价, 2021, (03): 108-115+159.
- [4] 王文娜, 胡杨莹, 吴啸滇, 等. 浙江数字化改革背景下公共数据在统计调查中的创新应用研究[J]. 统计科学与实践, 2021, (12): 23-26.

作者简介:

孙丽(1976.1—), 性别: 女, 民族: 汉, 籍贯: 辽宁铁岭, 职称: 高级统计师, 学历: 大专。研究方向: 统计。