

Python助力问卷调查数据处理和思考

■ 赵坤

在信息化时代,问卷调查作为社会科学、市场调研、公共政策制定等领域的核心信息收集手段,其应用场景日益广泛。但随着调研规模扩大,题型设计复杂,传统数据处理模式逐渐暴露短板,如数据量过载导致人工处理效率低下,流程繁琐易出现校验遗漏,复杂分析需求难以快速响应。笔者从工具开发视角,谈谈Python在问卷调查数据处理中的应用思路。

聚焦数据清洗智能化 筑牢数据质量基础

数据清洗是后续分析的前提,工具需实现“自动检测+规则化处理”的闭环,核心目标是降低人工校验成本,提升数据准确性。实现路径分为两步。

第一步模板导入与基准建立。先导入固定格式的问卷模板,模板需包含“题目、题型、权重、选项权重”四大核心字段,其中题型明确为单选、多选、填空、数据题四类,通过模板为后续数据

校验建立“标准参照系”,避免题目与数据错配。

第二步数据审核与匹配校验。导入原始调查数据后,系统自动触发两项核心校验。一是逻辑合规性审核,如多选题是否超出预设选项数量,单选题是否存在无效选项、数据题是否符合数值范围等;二是题目与数据的精准匹配,通过问卷模板的“题目”字段关联原始数据,确保每一条数据对应唯一题目;对检测出的无效值、缺失值,系统会按预设规则标记或提示人工干预。

推进处理流程自动化 打通全环节衔接壁垒

传统数据处理中“导入、清洗、分析、展示”的割裂式操作,易导致流程断档、数据损耗。工具需通过模块化设计实现全流程自动化衔接,具体架构分为四大独立且联动的模块。

一是数据导入模块。支持问卷模板与原始数据的批量导入,导入后自动完成格式识别与初步校验,校验通过后自动流转至数据清洗模块。

二是数据清洗模块。内置行业通用清洗规则(如剔除重复样本、修复格式错误数据),无需人工触发,可直接承接导入模块的数据并完成清洗,清洗后的数据自动同步至数据分析模块。

三是数据分析模块。待数据同步完成后自动开放,用户可按需选择基础汇总、分值计算等功能,分析过程无需手动干预,结果实时传递至结果展示模块。

四是结果展示模块。以结构化表格形式呈现分析结果(如选项占比、分值明细),支持实时查看与临时筛选,为后续导出提供可视化预览。

实现数据分析便捷化 降低专业操作门槛

针对工作人员的分析需求,工具需提供“低门槛+高功能”的分析能力,核心覆盖两大类功能,让复杂分析“一键可达”。

一是三大维度数据汇总功能。基础汇总:自动统计所有选择题的选项数量+占比情况,统计数据题的最大值、最小值和平均值。分类汇总:支持自定

义“筛选条件+分类维度”,实现交叉分析,如筛选“区域=A区”且“职业=机关单位”的样本情况,满足细分场景需求。单表汇总:将多分类维度的数据整合至同一表格,如在一张表中同步展示梁溪区、锡山区、滨湖区的数据情况,便于横向对比。

二是三大场景分值计算功能。基础分值计算:基于模板内置的“题目权重”与“选项权重”,自动计算单选题得分与样本总分。分类分值计算:按预设条件拆分分值,如分别计算苏南、苏中、苏北三地区的分值情况。单表分值计算:在同一表格中展示多分类维度的分值,如同步展示四辖区的总分、单题平均分,并新增“总分数据得分”(所有样本总分均值)“平均分”(各分类维度得分的均值)两类综合指标,满足宏观分析需求。

此外,工具内置“智能导出”功能,用户完成汇总或分值计算后,点击“智能导出”即可将结果导出为标准Excel表格,直接适配报告撰写场景,减少格式调整工作量。

(作者单位:国家统计局无锡调查队)

统计科普

正确解读粮食产量数据(一)

■ 王贵荣 常鹏

一、解读粮食产量的注意事项

在使用粮食产量数据时,需要注意以下三个问题:一是把握粮食产量的统计内涵。粮食产量从收获季节上看,包括夏粮、早稻和秋粮;从主要品种上看,包括谷物、薯类、豆类;其中谷物包括小麦、稻谷、玉米和其他谷物。在很多膳食营养书籍中,食物分类里所说的粮食其实大多是指谷物。所以人们习惯上说的粮食和统计上的粮食往往不是一回事,在使用粮食统计数据时要特别注意其统计涵义,这与使用其他统计数据的要求是一样的。

二是注意粮食产量的公制和习惯计量单位。在正式的统计年鉴和统计公报中的粮食播种面积和产量都是用公顷、吨、公斤这些公制单位。而有些统计报告和政府工作报告,以及很多有关经济形势分析的文章里出于习惯性原因,往往采用亩、斤等计量单位。要注意这些计量单位的换算关系,1公顷=15亩,1吨=1000公斤,1公斤=2市斤。在遇到万亩、亿斤等计量单位时,要注意换算关系。

三是注意国际对比时的统计口径。在阅读国外有关粮食论文和资料进行国际对比时,除要对中国的粮食统计内涵有所了解外,还要了解国际上粮食统计口径,特别是搞清楚对比国家或国际组织的粮食统计内涵。国际组织和很多国家的粮食就是指谷物,有些国家如印度尼西亚、巴基斯坦等亚洲、非洲的发展中国家的粮食口径是包括谷物和薯类的。

二、粮食产量历史数据解读

党的十八大以来,以习近平同志为核心的党中央高度重视粮食生产,把粮食安全作为治国理政的头等大事,提出确保谷物基本自给、口粮绝对安全的新粮食安全观,牢牢守住粮食安全主动权,带领亿万人民走出了一条中国特色粮食安全之路。我国粮食综合生产能力稳步提升,粮食产量稳定增长,连续8年稳定在1.3万亿斤以上,种植结构持续优化,区域布局更加合理,质量效益不断提高,以占世界9%的耕地、6%的淡水资源,养育了世界近1/5的人口,有力回答了“谁来养活中国”的问题,中国人的饭碗越端越稳。粮食生产取得举世瞩目的历史性成就,为世界粮食安全作出了积极贡献。

(一)粮食产量稳定增长

党的十八大以来,我国粮食连年丰收,粮食需求刚性增加,但稻谷、小麦保持多年产大于需,粮食库存充裕,市场供应充足。从粮食价格看,近几年受成本上涨推动,主要粮食品种价格有所上涨,但总体保持平稳运行。2015年开始,我国粮食产量迈上并站稳1.3万亿斤新台阶。2022年,粮食产量为13731亿斤,创历史新高,比2012年增加1486亿斤,增长12.1%。稻谷、小麦、玉米、大豆等主要品种产量均实现不同程度的增加。2022年我国人均粮食产量486.1公斤,连续多年远超世界平均水平,也高于国际公认的400公斤粮食安全线。

一是粮食单产水平持续提高。党的十八大以来,我国粮食单产除2016年因灾和种植结构调整略有下降外,其他年份均实现增长,是我国粮食增产的主要贡献因素。2013年、2015年和2019年我国粮食单产水平先后站上360公斤/亩、370公斤/亩和380公斤/亩三个台阶,2022年我国粮食单产386.8公斤/亩,比2012年增加29.9公斤/亩,增长8.4%。单产提高主要得益于我国农业综合生产能力不断提升,粮食作物基本实现良种全覆盖,农机农艺结合,良种良法配套,农田基础设施明显改善,防灾减灾能力进一步增强。全国已建成早涝保收、高产稳产的高标准农田2012年仅1亿亩,2022年增加到10亿亩;农业科技贡献率2012年为53.5%,2022年达到62.4%;农作物耕种收综合机械化率2012年为57%,2022年达到73%。

二是粮食播种面积稳中有增。党的十八大以来,我国粮食播种面积稳定在17亿亩以上,整体呈波动上升趋势。2012-2016年,粮食播种面积连续增长,2016年达到17.88亿亩,比2012年增加7293万亩,增长4.3%。2017年农业种植结构调整,“镰刀弯”地区调减玉米播种面积,实行“粮改饲”“粮改豆”,同时增加了花生、中草药材等非粮作物,粮食播种面积连续三年下降,三年间减少4750万亩。2020年,各地加大了对粮食生产支持力度,层层压实粮食生产责任,落实各项补贴政策,提高农民种粮积极性,我国粮食播种面积恢复至17.52亿亩,比2019年增加1057万亩,增长0.6%,扭转了下滑态势。2022年粮食播种面积177498万亩,比2020年增加2346万亩,增长1.3%;比2012年增加5946万亩,增长3.5%。

(二)种植结构持续优化

一是稻谷、小麦产量基本稳定。稻谷和小麦是我国主要口粮作物。党的十八大以来,我国继续实施稻谷和小麦最低收购价,特别是近两年针对农资价格上涨导致的种粮成本增加,党中央多次向实际种粮农民发放一次性补贴,农民种粮积极性得到有效保障。随着单产水平的提升,产量稳中有增,2022年稻谷产量4170亿斤,比2012年增加39.3亿斤,增长1.0%。2022年小麦产量2754亿斤,比2012年增加305.0亿斤,增长12.5%。稻谷、小麦作为重要口粮作物,目前库存充裕,供应充足,已连续多年产大于需,有力确保了“谷物基本自给,口粮绝对安全”。

二是玉米、大豆产量实现较快增长。党的十八大以来,调整玉米产业结构和区域布局,加大高产稳产、耐密抗倒、抗病抗虫、宜机械化作业的紧凑型玉米品种推广力度,稳步推进玉米生物育种技术应用。2022年玉米产量5544亿斤,比2012年增加952.9亿斤,增长20.8%,为粮食连年丰收作出了突出贡献。实施大豆振兴计划以来,我国大豆产量快速增长,2022年大豆产量达406亿斤,比2012年增加137.0亿斤,增长51.0%。

三是杂粮、薯类产量稳步提高。党的十八大以来,随着生活水平提高,人民群众不仅要“吃得饱”,还要“吃得好”,居民膳食结构持续优化,杂粮需求量不断增加。各地立足自身资源禀赋,大力发展特色杂粮,通过建设标准化基地,发展社会化服务,着力打造“一乡一业,一村一品”新格局,巩固脱贫攻坚成果,推进乡村产业振兴。2022年谷子、高粱等杂粮产量196亿斤,比2012年增加36.0亿斤,增长22.4%。各地区在稳定现有马铃薯生产基础上,因地制宜扩种补种,2022年薯类产量595亿斤,比2012年增加18.9亿斤,增长3.3%。

(摘自《领导干部应知应会主要统计指标诠释》,中共中央党校出版社 中国统计出版社)

怎样提升工业价格调查企业的调研质效

■ 陈瑜

工业企业作为经济活动主体,通过调研其产品价格变动、财务状况、经营策略、面临的问题、意见建议等生产经营情况,可以直观展现复杂多变的经济环境,帮助统计调查人员、政策制定主体准确理解、把握生产形势。笔者结合工作实践,就如何做好工业价格调查样本企业的调研工作谈几点心得。

挖掘数据价值

注重建立数据资源整合机制。如,单个企业价格数据分别与上下游企业、同行业样本企业的趋势作对照,筛选价格趋势不一致等情况,及时与企业核实,确认无误后,分析指标变动原因,探索和发现经济社会发展过程中出现的新情况、新问题和新变化。

确保调查样本的代表性。选择本地区产值占比大、有发展前景的行业企业,持续关注企业生产经营状况,捕捉价格变动的深层次因素,有利于挖掘出相关行业价格传导的经济热点。

多元收集信息

在实地走访或数据核查时,与企业统计部门、财务部门开展座谈交流,详细了解企业发展情况,探讨在当前经济形势下,企业面临困难、政策期盼和应对措施。在赴企业实地调研前,可以提前查阅该企业所在行业的一些新闻资讯,了解行业最新发展状况和潜在风险;查阅该企业门户网站,了解企业成长历程、业务开展、科研成果,形成对该企业的初步认识和了解;结合当前政策环境和经济形势热点内容,提前明确调研主题,做到知己知彼。

强化“数库”“智库”融合,聚焦党委政府中心工作,及时了解本地区惠企政策措施,加强涉企政策执行情况和落实效果监测,通过评估判断政策出台对调查指标的影响程度,丰富案例分析内容,增强案例权威性和参考价值。

通过关注龙头、示范企业等典型报道,进而了解不同产业发展和结构演变过程,加深对行业市场状况的理解和前景趋势预判,同时结合增加值、固定资产投资



原永红 绘图

等资讯,确保案例分析符合经济运行实际。

建立合作机制

有时无法在案例分析中通过现有样本事例,准确反映经济发展中的一些特殊情况,因此,需要建立多方合作机制。

建立部门协作关系。通过每一次调研契机,与发改、经信、商务等与工价调查息息相关的部门建立长期友好合作关系,共享相关数据、资源,互通有无,积极主动共同开展有关调研分析。

纵深拓展调研范围。针对企业统计人员行业情况了解面较窄的问题,积极与企业领导沟通,协调有关专业技术人员或行业一线人员答疑解惑,挖掘经济现象背后的本质问题。

掌握调研技巧

确定调研主题后,根据具体情境运用多样化调研手段。问卷调查方法适用于大规模收集数据,通常采取网络二维码问卷或纸质问卷进行收集;访谈法适用于寻找个体案例,深入了解特定群体对象的观点和需求,可以通过面对面座谈、电话访谈等方式进行;观察调研法适用于客观观察社会、经济现象,通过直接观察调研对象的行为或所处的环境来获取信息。例如走访企业时直接观察企业生产经营情况,判断数据、信息来源可靠性;文献调研方法适用于获取信息背景,也可通过与其他部门沟通了解相关调研背景、政策环境。

(作者单位:国家统计局嘉兴调查队)

优化寄回带回收入记账的两种方法

■ 张志如

外出从业人员寄回带回收入是居民收入的重要组成部分。然而,实际工作中发现,因调查对象记账能力存在差异,面临对指标理解易产生偏差、收入编码变动频繁、收入记账不易把握等问题。笔者从记账方法层面探索两种优化方案,以期避免寄回带回收入记账问题的发生。

分步记录

核心思路在于将外出从业人员收入的记录过程分解为两个清晰的步骤,从而降低调查员的认知负担,提高记账准确性。

具体而言,在记账APP或账本上,将“外出从业总收入”与“实际寄回带回金额”两个科目紧密排列在一起,并配有清晰的图标和文字说明,引导调查员按步骤填写。第一步,要求调查员先填

写外出成员的全部收入,包括工资、奖金、津贴等(不论是否实际寄回);第二步,调查员单独记录当月实际通过汇款、现金带回等方式转入家庭用于生活消费的金。

第一步明确“他挣了多少钱”,强调收入的完整性,避免因误解“寄回带回”概念而导致收入漏记。第二步明确“他给了家里多少钱”,强调收入部分性,避免收入多记。将复杂的“寄回带回”概念拆解成两个简单明了的概念:“他挣了多少钱”和“他给了家里多少钱”,调查员无需理解统计学的概念边界,只需要进行事实记录。

设置“寄回带回收入”≤“外出从业总收入”的逻辑关系,系统自动计算实际寄回带回金额占全部收入的比例,并在界面上直观展示。设置合理的比例波动阈值,对异常波动数据进行预警,帮助调查员和业务人员核对数据是否真实准确。

按比例折算

核心思路是将“哪些收入属于寄回带回”的判断环节从调查员手中剥离,交由系统或统计方法完成。

具体而言,首先记录外出从业人员全部收入,引导调查员先完整记录外出从业人员的全部工资收入,避免因“只需记寄回带回部分”而漏记或错记。其次依据科学合理的比例折算为寄回带回收入。

比例设定的科学性是本方法的关键,需综合考虑多方面因素。一是全国或地区物价指数和平均工资水平,不同区域的经济差异;二是家庭结构,如常住成员数量、是否有赡养老人等;三是家庭消费支出占收入的比例,通常寄回带回的以保障基本生活为主。因此,比例设定的依据,可参考全国或地区物价指数、平均工资水平、家庭常住成员占比以及家庭消费支出占收入的比例等因素综合确定。为保持推算结果的科学性、准确性,应建立折算

比例的动态调整机制,定期根据经济与社会发展状况对比例进行调整。

该方法调查员无需判断“哪些钱寄回了”,只需记录总收入,其余工作由系统完成。这样做,一方面减轻了调查员的记账压力,只需记录全部工资,简化了记账过程;另一方面,通过科学、差异化的比例设定,使得估算出的“寄回带回收入”更接近真实,可能比调查员主观判断的记录更加真实准确。

“分步记录法”注重过程引导,概念清晰,易于理解,记录的是实际发生的、确切的汇款和带回金额,避免了估算带来的偏差。但需要调查员有一定的配合度和记账能力,另外记账人员还必须与外出务工人员保持频繁沟通,才能准确知道其“总收入”;“按比例折算法”依托系统智能处理,极大的简化记账流程,通过统一的模型进行折算,可以有效减少因调查员个体理解差异而带来的数据波动。

(作者单位:国家统计局灵璧调查队)