

综合统计调查数据分析方法及应用研究

尧霞

山东省淄博市高青县田镇街道办事处，山东 淄博 256300

DOI:10.61369/SE.2026050041

摘 要： 本文综合探讨了统计调查数据分析的方法及其应用研究。通过对描述统计、推论统计、相关分析、回归分析等多种数据分析方法的系统梳理，结合实际应用案例，深入分析了这些方法在揭示社会经济现象、预测未来趋势、评估政策效果等方面的重要作用。研究强调了在数据分析过程中“以人为本”的理念，指出应充分关注人的需求和利益，以推动社会的全面、协调、可持续发展。最后，本文提出了优化数据收集、处理和分析流程的建议，以期提高数据分析的质量和效率，为相关领域的研究和决策提供科学依据。

关 键 词： 统计调查；数据分析方法；描述统计；推论统计；相关分析；回归分析

Research on Comprehensive Statistical Survey Data Analysis Methods and Their Applications

Yan Xia

Tianzhen Subdistrict Office, Gaoqing County, Zibo City, Shandong Province, Zibo, Shandong 256300

Abstract： This paper comprehensively explores methods for analyzing statistical survey data and their applied research. By systematically reviewing various data analysis methods, including descriptive statistics, inferential statistics, correlation analysis, and regression analysis, and integrating practical application cases, it delves into the significant roles these methods play in revealing socio-economic phenomena, predicting future trends, and evaluating policy effectiveness. The study emphasizes the "people-oriented" concept in the data analysis process, pointing out that full attention should be paid to human needs and interests to promote comprehensive, coordinated, and sustainable social development. Finally, the paper proposes suggestions for optimizing data collection, processing, and analysis procedures to enhance the quality and efficiency of data analysis, providing a scientific basis for research and decision-making in related fields.

Keywords： statistical surveys; data analysis methods; descriptive statistics; inferential statistics; correlation analysis; regression analysis

在信息化社会的今天，数据已经成为促进社会进步与经济发展的关键因素。统计调查作为一种重要的数据获取方式，它所采用的数据分析方法对揭示社会经济现象，预测未来发展趋势，评价政策效果都有着十分重要的作用。但是传统数据分析方法通常只注重技术的应用，忽略了在进行数据分析时“以人为本”的思想。为此，本研究旨在对统计调查数据分析方法及运用进行全面论述，并着重指出数据分析过程中，要充分重视人类的需要与兴趣，促进社会全面、协调，可持续发展。

一、综合统计调查数据分析的方法

（一）描述统计

描述统计作为综合统计调查资料分析的一种最基本手段，通过对资料进行整理，归纳与描述，使人们能迅速地认识资料的集中趋势，离散程度以及分布形态的特点。

集中趋势主要有平均数、中位数、众数等描述方法，平均数就是全部资料之和除以资料数目，可以反映资料的整体水平。比如在一个县居民收入调查中，计算出来的平均收入就能对这个县

的居民收入有一个大致了解。但是平均数易受极端值影响，此时中位数更重要。中位数指数据由小变大排好之后在中间的值。若数据个数为奇数则中位数为中间数字；若为偶数时，为中间两数之和。众数是数据中出现频率最高的一个数值，它能够揭示数据的集中区域^[1]。

描述离散程度常用极差、方差、标准差等指标。极差为资料中最大值和最小值之差，可以简便而直观地表示资料波动幅度，方差为各资料中数据与平均数之差的平方的平均值，标准差为方差平方根。方差与标准差越大，表明数据离散程度越大，反之越

小。比如在对两县域居民收入离散程度进行对比时，可通过方差、标准差等指标的计算判断哪一个县域居民收入差距较大。

分布形态描述以偏度为主，峰度为辅。偏度度量数据分布是否对称，当偏度等于0时，数据分布对称，0以上则右偏，0以下则左偏。峰度是用来描述数据分布尖峰程度的指标。当峰度为0时，其尖峰程度与正态分布相当，大于0意味着尖峰程度超过正态分布，而小于0则意味着更为平坦。对资料进行偏度、峰度分析可得知其分布是否服从正态分布、有无异常值。

在实践中，描述统计有助于迅速地理解数据的基本性质，并为后续分析提供依据。比如在市场调研中，运用描述统计的方法就能了解消费者年龄，收入，购买行为数据集中趋势与离散程度，为商家制定营销策略奠定基础。就社会科学研究而言，描述统计有助于研究人员对研究对象基本状况的理解和对问题进行深入剖析提供背景信息。

（二）推论统计

推论统计是在描述统计的基础上，通过对样本数据的分析来推断总体特征的一种方法，鉴于在多数场合下，不能对整体情况进行深入的调查，只能依赖于样本抽取来获取更全面的信息，因此可以推断统计在综合统计调查数据的分析过程中起到了不可或缺的角色^[2]。

在推论统计中，参数估计是其中最重要的一个部分。参数估计有点估计与区间估计两种。点估计是用样本统计量来估计总体参数，例如用样本平均数估计总体平均数。区间估计则是在点估计的基础上，给出一个区间范围，以一定的概率保证总体参数落在这个区间内。作为一个例子，提供了一个置信度区间，确保了95%的置信度使得总体的平均值落在这一区间之内。区间估计能对总体参数范围有更精确的认识，并降低估计误差。在推论统计中，假设检验又是其中一种重要手段，假设检验先提出原假设与备择假设，再依据样本数据对检验统计量进行统计并判定其显著性程度，最终依据检验统计量概率值与显著性程度进行决策。若检验统计量概率值低于显著性水平时，拒绝原假设而采纳备择假设；否则就不会排斥原来的假设。假设检验有助于判定样本数据能否支持自己的假说，进而推断总体特征。

以某县域居民健康状况调查为例，可选取部分居民为研究样本，并通过样本数据分析推测县域居民健康状况。如果了解该县居民的平均身高是否高于全国平均水平，可以提出原假设“全县居民平均身高与全国平均水平相当”，备择假设“全县居民平均身高比全国平均水平高”。再提取样品，统计样品平均身高并对统计量进行检验，判定其显著性水平达到0.05。若检验统计量概率值在0.05以下，拒绝原假设而接受备择假设——县域居民平均身高大于全国平均水平；其实不然，并没有拒斥原来的假定，也就是无法认为全县居民平均身高比全国平均水平高。

（三）相关分析

相关分析是综合统计调查数据分析中用于研究两个或多个变量之间关系的一种方法，利用相关分析可以知道变量间有无联系及联系的力度与方向^[3]。

以相关系数作为变量间相关程度的量度，常见的相关系数包

括皮尔逊相关系数，斯皮尔曼相关系数等，皮尔逊相关系数是用来度量两变量间线性关系的指标，其范围为-1~1。当相关系数达到1时，这意味着两个变量之间存在完全的正相关性；当相关系数达到-1时，这意味着两个变量之间存在完全的负相关性；当相关系数为0时，表示两个变量之间不存在线性关系。斯皮尔曼相关系数被用来量化两个变量间的秩相关性，特别是在变量不遵循正态分布或存在异常值的场合。相关分析有助于找出变量间可能存在的联系，从而为后续研究提供了线索，比如，当考察一个县教育水平与经济发展情况时，就能从有关分析中知道教育水平与经济的发展有无联系。若找出教育水平和经济发展呈正相关关系的证据时，就能深入研究教育在经济发展中的作用机理，从而为制订促进经济发展政策奠定基础。

在实践中相关分析也可应用于变量筛选与预测。在多元回归等统计建模过程中，可利用相关分析甄别与因变量有关的自变量以提高其精度。同时通过历史数据相关分析可找出变量间关系模式以预测未来走势。例如，在对某县的房地产市场进行分析时，可以通过相关分析来了解房价与多个因素（比如地理位置，房屋面积，周围的配套设施等等）之间的关系。若发现房价与其所处地理位置及周边配套设施有很强的正相关关系则房地产投资决策中可着重关注上述因素。同时对历史数据进行相关分析可构建房价预测模型供房地产市场参与主体决策参考。

（四）回归分析

回归分析作为综合统计调查数据分析的一种重要手段，主要应用于对变量间因果关系的研究，通过构建数学模型对因变量进行数值预测。

回归分析以确定自变量与因变量间函数关系为中心。在实践中，一般都是根据所研究的问题性质及资料的特性来选用适当的回归模型。常用回归模型包括线性回归，多项式回归和逻辑回归。线性回归被认为是最基础且最常被采用的回归模型之一，其假定因变量和自变量呈线性关系，利用最小二乘法对模型进行参数估计。比如在对某县域居民收入和教育水平关系的研究中，以居民收入为因变量、教育水平为自变量构建线性回归模型。从样本数据中分析可得出教育水平与居民收入之间的影响系数，来判断教育水平与居民收入之间的关系。

多项式回归方法特别适合于因变量和自变量间存在非线性关系的场景。它引入自变量高次项对数据进行拟合，较好地体现了变量间的相互联系。比如，当对一个县农作物产量和施肥量的关系进行研究时就可能发现，产量和施肥并不存在单纯的线性关系，而表现为先增加后减少。此时可利用多项式回归模型引入施肥量二次项或者三次项进行数据拟合。

逻辑回归在因变量为二分类变量时应用较多。它是通过构造逻辑函数，对因变量取某数值的可能性进行预测。例如，在研究某县居民是否购买商业保险时，可以将是否购买商业保险作为因变量（取值为0或1），居民的收入、年龄、教育水平等作为自变量，建立逻辑回归模型。对样本数据进行分析可得出各自变量在居民商业保险购买过程中的作用大小，为保险公司营销策略的制定奠定了基础。

回归分析广泛应用于综合统计调查的数据分析。一方面有助于变量间因果关系的深刻理解，从而为科学研究与政策制定提供理论支撑。另一方面又能通过预测模型的构建为企业及政府部门的决策提供参考。以经济领域为例，回归分析可应用于市场需求，股票价格的预测等方面；从社会科学的角度来看，回归分析可应用于教育，医疗和就业的研究。

二、综合统计调查数据分析方法的应用

（一）完善数据收集机制：确保数据的全面性和准确性

数据收集作为数据分析工作的基石，健全的数据收集机制可以保证数据的全面性与准确性，从而为之后数据处理与分析工作打下坚实基础。就健全数据收集机制而言，首先必须明确采集对象与范围，保证采集到的资料能综合反映研究对象真实状况。同时也要建立科学的数据采集方法与过程，其中包括明确数据来源，设计问卷或者数据抓取工具以及明确数据采集时间与周期。为了更有效地提升数据采集的速度和品质，可以利用如大数据和云计算这样的现代信息技术，来达到数据的自动收集和实时更新的目的。另外，还要建立并完善数据审核与校验机制，严格把关与校验所采集数据，保证数据准确可靠。数据收集机制不断完善的同时，也要重视数据隐私保护与安全，在对个人数据或者敏感信息进行采集与处理的过程中，必须严格按照有关法律法规以及道德规范进行，保证数据合法合规。

（二）增强数据处理能力：提升数据处理效率与准确性

数据处理是数据分析的关键环节之一，一个强大的数据处理能力能够显著提高数据分析的效率和精度。就增强数据处理能力而言，首先必须提高数据清洗与预处理能力。数据清洗就是将采集的原始数据经过去噪，去重及填充缺失值处理，从而达到去除数据噪声及异常值并改善其品质的目的。数据预处理主要涉及对数据格式的转化和标准化操作，以确保后续的数据分析任务能够无障碍地进行^[1]。

为进一步提升数据处理效率与准确性，可引进数据挖掘，机

器学习等先进数据处理技术与工具，这些技术与工具可以自动化完成数据清洗与预处理烦琐任务，极大的降低了人工负担。同时，它们还能够从数据中挖掘出潜在的规律和特征，为数据分析提供更有价值的信息。在增强数据处理能力时，还要注意其实时性与动态性。在数据量越来越大、更新速度越来越快的情况下，必须建立实时数据处理与动态更新机制来保证数据处理的及时性与准确性。

（三）创新数据分析方法：发掘数据背后所蕴含价值

在综合统计调查数据分析方法中，数据分析是终极目标，也是核心环节。一种创新性的数据分析方法可以深度挖掘其背后所蕴含的价值并对决策进行强有力的支撑。创新数据分析方法时，首先应注重数据可视化表达，通过可视化表达可以把繁杂的数据信息变成直观的图形与图像，从而使得对数据的判读变得更有效、更精确。比如可以通过散点图，热力图这些可视化工具显示变量间的关联关系及变化趋势。另外，可根据具体研究问题与需要探索新型数据分析方法与模式，比如在预测分析方面，可以采用时间序列分析和回归分析的统计方法来构建预测模型，在文本分析方面，可以使用自然语言处理技术实现文本数据的情感分析和主题识别。这些新型数据分析方法与模型，可以给人们带来更全面，更深刻的的数据洞察与决策支持。在对数据分析方法进行革新的同时，也要重视学科之间的协作和沟通。数据分析是一项涉及众多学科领域，并且需要各学科间进行协作和沟通才能够达到较好成效的综合工作。

三、结语

文章通过系统地论述统计调查数据分析方法，并对这些方法在实践中进行深入研究，全面论证数据分析对于揭示社会经济现象的作用、预测未来走势，评价政策效果的重要性。在今后的研究中要不断对数据分析方法应用及优化策略进行深入探究，希望能够提升数据分析质量与效率，对相关领域研究与决策提供更加强大的支撑。

参考文献

- [1] 王少卿. 我国社会经济统计制度方法完善对策研究 [N]. 河南经济报, 2024-07-20(009).
- [2] 赵明. 加快修订《金融统计管理规定》制度的思考 [J]. 当代金融家, 2024(06): 146-147.
- [3] 程豪, 陈毅, 申金升. 统计与调查双轮驱动平台建设路径 [J]. 科技智囊, 2024(05): 39-46.
- [4] 田毅清, 陈银珠. 数字赋能自然资源综合统计的浙江实践 [J]. 浙江国土资源, 2024(03): 40-41.