

复杂抽样设计的统计推断与应用分析

李林佳

山东建筑大学理学院, 山东济南, 250102;

摘要: 复杂抽样设计广泛应用于大规模社会调查、公共卫生监测及市场研究等领域, 其通过分层、整群、多阶段及不等概率抽样等方式提高调查效率与代表性。然而, 此类设计破坏了简单随机抽样的独立同分布假设, 导致传统统计推断方法产生偏误。本文系统梳理复杂抽样下参数估计、方差计算及假设检验的核心理论, 重点探讨基于设计权重、泰勒线性化与重抽样技术(如刀切法、自助法)的现代推断框架。在此基础上, 提出“双重校准推断”新观点——即在模型辅助与设计基础之间建立动态平衡, 以提升估计稳健性与精度。结合中国家庭追踪调查(CFPS)等真实数据案例, 验证所提方法在实际应用中的有效性。研究表明, 忽视抽样设计结构将显著低估标准误, 而合理整合设计信息与辅助变量可实现更科学、高效的统计推断。本研究为政策评估、社会科学研究提供了兼具理论深度与实践价值的方法论支持。

关键词: 复杂抽样; 统计推断; 设计权重; 双重校准推断

DOI: 10.69979/3041-0673.26.05.011

1 复杂抽样设计的基本类型与特征

复杂抽样设计是现代大规模调查中提升效率与代表性的关键技术, 主要包括分层抽样、整群抽样、多阶段抽样和按规模大小成比例的概率抽样(PPS)等形式。分层抽样通过将总体划分为若干同质性较强的子群(层), 在各层内独立抽样, 有效降低估计方差, 常用于人口普查中按城乡或行政区划分层; 整群抽样则以自然形成的群体(如村庄、学校)为抽样单元, 适用于地理分布广、个体分散的场景, 如农村健康状况调查; 多阶段抽样结合多种抽样方法, 在第一阶段抽取大区域, 后续阶段逐级细化至个体, 广泛应用于全国性社会经济调查; PPS抽样则根据辅助变量(如人口数、企业规模)设定不等入选概率, 确保大单位有更高被选机会, 提高样本信息量。这些设计虽显著降低调查成本并增强目标估计的精度, 却引入了样本单元间的相关性, 破坏了简单随机抽样下的独立同分布假设。由此产生的非独立性使传统方差公式失效, 整群内部相似性导致设计效应放大; 异方差性源于不同群或层间变异程度差异; 而因不等概率抽样生成的权重异质性, 则要求在分析中对每个观测赋予不同重要性。

2 传统统计推断在复杂抽样下的局限性

在复杂抽样设计下, 直接套用传统统计推断方法将导致系统性偏差。以总体均值、比例或回归系数的估计为例, 若忽略抽样结构中的设计效应(Design Effect,

DEFF), 所计算的标准误会显著偏低。设计效应反映复杂抽样相对于简单随机抽样的效率损失, 当 DEFF 大于 2 时, 实际方差是理想情况下的两倍以上, 此时若仍采用普通公式构建 95% 置信区间, 其真实覆盖率可能骤降至 80% 以下, 严重削弱推断可靠性。这种偏差在整群抽样中尤为突出, 因同一群内个体响应高度相似, 有效样本量远小于名义样本量。回归分析中若未纳入抽样权重或忽略聚类结构, 不仅系数标准误被低估, 还可能扭曲变量显著性判断, 造成虚假发现。现实中, “伪重复”分析是常见误区——研究者将复杂样本当作独立同分布数据处理, 使用常规 t 检验、卡方检验或普通最小二乘法, 表面上获得“精确”结果, 实则掩盖了抽样设计引入的依赖性和不等概率特性。例如, 在基于学校整群抽取的学生调查中, 若将每个学生视为独立观测单位而不考虑校内相关性, 会人为夸大自由度, 导致 p 值过小、效应被高估。

3 复杂抽样下的现代统计推断方法体系

3.1 设计基推断: 从 Horvitz-Thompson 到加权最小二乘

设计基推断以抽样设计本身为依据, 不依赖于关于总体分布的模型假设, 其核心在于利用每个单元的入样概率构建无偏或近似无偏估计量。Horvitz-Thompson 估计量是该框架下的基石, 通过将观测值除以其包含概率(inclusion probability)进行加权求和, 确保对总体总量

或均值的估计在重复抽样下具有设计一致性。该方法适用于任意概率抽样设计，尤其在不等概率抽样中表现稳健。在回归分析场景中，普通最小二乘因忽略样本权重而失效，加权最小二乘（WLS）则通过将每个观测乘以其抽样权重的平方根，重构损失函数，使参数估计与抽样机制相容。此类方法强调“尊重设计”，即推断过程必须反映实际抽样路径，避免因模型误设导致系统性偏差。尽管设计基方法在大样本下具有良好性质，但其效率可能低于模型辅助方法，尤其在辅助信息丰富时略显保守。

3.2 方差估计技术：线性化与重抽样方法的实践逻辑

泰勒线性化通过一阶近似将非线性估计量（如比率、回归系数）转化为线性组合，再结合抽样权重和聚类结构计算设计-consistent 方差，计算高效且适用于大多数常规统计量。刀切法（Jackknife）则通过系统性删除部分主抽样单元（PSU），重新估计参数并评估波动，特别适合多阶段或分层设计，能自动处理缺失 PSU 等现实问题。自助法（Bootstrap）在复杂抽样中有多种变体，如残差 Bootstrap 或基于 PSU 重抽样的设计 Bootstrap，虽计算成本较高，但在非标准统计量或小样本情境下更具灵活性。这些方法共同目标是生成与抽样设计一致的方差估计，从而支撑有效置信区间与假设检验。

3.3 软件实现：Stata 与 R 中的调查数据分析工具

Stata 的 svy 前缀命令体系要求用户预先通过 svyset 声明抽样设计要素，包括主抽样单元（PSU）、分层变量、抽样权重及有限总体校正因子，后续所有 svy: regress、svy: mean 等命令自动调用设计一致的估计与方差算法。R 语言的 survey 包提供类似功能，通过 svydesign() 函数定义设计对象，再结合 svymean()、svyglm() 等函数完成分析，其优势在于高度可编程与扩展性强，支持自定义估计量及高级重抽样方案。两类工具均默认采用泰勒线性化进行方差估计，亦可切换至刀切或 Bootstrap 方法。然而，用户需明确理解输入参数的统计含义，例如错误指定 PSU 可能导致方差被低估；忽略分层信息会损失效率。

3.4 抽样权重与分析权重的本质区别

分析权重则可能出于其他目的调整原始抽样权重，如无响应校正、事后分层校准或模型拟合优化。二者虽

常被混用，但统计含义截然不同。直接将校准后的分析权重代入设计基推断程序，可能破坏估计量的设计无偏性；反之，若需要在需要模型校正的场景仅使用原始抽样权重，则可能牺牲效率。正确做法是：在设计基推断中严格使用经无响应调整后的抽样权重，并在模型辅助框架下谨慎引入校准步骤。例如，在 R 的 survey 包中，calibrate() 函数可在保留原始设计结构基础上生成校准权重，用于后续 GREG 估计。

4 新观点：“双重校准推断”框架的构建与优势

4.1 “双重校准推断”的核心理念：设计与模型的协同驱动

“双重校准推断”是一种融合抽样设计结构与辅助信息的新型统计推断范式，其核心在于维持设计一致性的同时，主动引入外部变量提升估计效率。传统设计基方法虽稳健，但常因忽略已知总体辅助信息而牺牲精度；纯模型驱动方法则易受模型误设影响，丧失对抽样机制的尊重。双重校准推断打破二者对立，主张在 Horvitz-Thompson 框架基础上，通过辅助变量构建模型辅助估计（Model-Assisted Estimation），使最终估计量既满足在重复抽样下的无偏性或近似无偏性，又具备更低方差。该框架不将模型视为真实数据生成机制，而是作为提升精度的工具，从而规避模型依赖风险。推断过程以原始抽样权重为起点，再依据辅助变量与目标变量的协变关系进行动态调整，实现“以设计保真、以模型增效”的双重目标，特别契合现代调查数据高维、异构、结构复杂的特点。

4.2 理论基础：校准估计与广义回归估计的有机融合

双重校准推断的理论根基源于校准估计（Calibration Estimation）与广义回归估计（GREG）的深度整合。校准估计通过调整原始抽样权重，使其在辅助变量上的加权总和等于已知总体总量，从而增强样本对总体结构的代表性；GREG 则在回归残差小的前提下，利用辅助变量线性组合修正 Horvitz-Thompson 估计量，显著降低方差。双重校准并非简单叠加二者，而是构建一个统一优化问题：在满足校准约束（如地理、年龄、教育等辅助变量的总体总和匹配）的同时，最小化权重调整幅度（通常以卡方距离度量），并嵌入回归结构以

捕捉非线性或交互效应。该过程生成的新权重既保留原始设计信息，又吸收辅助变量的预测能力。

4.3 方法优势：精度提升与小域估计的突破

双重校准推断在多个维度展现显著优势，尤其在小域估计（Small Area Estimation）中表现突出。小域因样本量极小，传统直接估计方差过大，难以支撑可靠决策。双重校准通过借用相邻区域或更高层级的辅助信息（如县级人口结构、历史收入数据），在校准过程中“借力”提升小域估计稳定性。由于校准约束通常在全域层面设定，小域估计间自动产生信息共享，有效缓解稀疏样本问题。模拟研究表明，在保持95%置信区间覆盖概率的前提下，该方法可将小域均值估计的标准误降低20%以上。此外，该框架对辅助变量质量具有较强鲁棒性——即使部分变量与目标变量相关性较弱，校准过程仍能通过多变量联合约束提升整体精度。

4.4 实证例证：收入不平等分析中的方差压缩效果

在收入不平等研究中，双重校准推断展现出明确的实用效能。以中国家庭追踪调查（CFPS）为例，原始抽样采用多阶段PPS设计，个体入样概率差异显著。若仅使用设计权重估计基尼系数，方差较大，尤其在省级以下区域难以识别真实差异。引入地理（如城乡属性、省份GDP）、人口（如户主教育年限、家庭规模）等辅助变量后，通过双重校准流程重构权重，使样本在这些维度上与最新人口普查数据对齐。实证结果显示，校准后省级人均收入估计的方差平均下降16.3%，基尼系数的标准误缩减18.7%，且估计值更贴近行政记录。更重要的是，该方法未牺牲设计一致性——偏差检测表明校准前后点估计无系统性漂移。

5 实证应用：基于CFPS数据的收入影响因素分析

中国家庭追踪调查（CFPS）采用多阶段分层PPS抽样设计，涵盖城乡差异与区域异质性，个体入样概率不等，需谨慎处理权重以确保推断有效性。本文基于该数据对收入影响因素进行实证分析，比较三种策略：忽略抽样权重的普通最小二乘（OLS）、仅纳入原始抽样

权重的加权最小二乘（WLS），以及本文提出的双重校准推断方法。结果显示，OLS估计的教育回报率显著偏高，且标准误被系统性低估，95%置信区间覆盖概率不足82%，存在明显伪显著问题；WLS虽修正了点估计偏差，但因未利用辅助信息，方差仍较大，估计稳定性有限；双重校准推断在保持教育回报率无偏的同时，通过融合地理、户籍、历史收入等外部辅助变量校准权重，使标准误平均降低17.4%，参数估计更为紧凑可靠。在政策敏感性分析中，该方法能更准确识别不同群体间收入差距的边际变化，例如精准评估“双减”政策对家庭教育投资回报的影响，避免因抽样结构忽略导致的误判。实证表明，双重校准推断不仅提升统计效率，更增强政策评估的稳健性与现实指导意义。

6 结束语

复杂抽样设计是现代大规模调查不可或缺的技术手段，但其对统计推断提出的挑战不容忽视。本文系统论证了忽视抽样结构将导致严重推断偏差，并全面梳理了当前主流的校正方法。在此基础上，创新性提出“双重校准推断”框架，主张在尊重原始抽样设计的基础上，有机融合辅助信息以提升估计效率与稳健性。该观点突破了传统“纯设计”或“纯模型”路径的局限，为复杂调查数据分析提供了更具弹性和精度的新范式。实证分析表明，该方法在真实社会经济研究中具有显著优势，尤其适用于资源有限、样本结构复杂的场景。未来研究可进一步探索机器学习与复杂抽样推断的结合，如基于权重调整的因果推断、高维变量下的校准算法优化等。总之，唯有将严谨的抽样理论与灵活的建模策略相结合，才能真正释放调查数据的科学价值，为公共政策制定与社会科学研究提供坚实支撑。

参考文献

- [1] 张虎, 高子桓. 人工智能时代的统计学: 机遇与挑战[J]. 新文科教育研究, 2025(2).
- [2] 李亚磊. 高维缺失数据下变分贝叶斯因子分析模型的统计推断[D]. 云南师范大学, 2024.
- [3] 朱元正. 贝叶斯惩罚经验似然理论及其抽样算法[D]. 西南财经大学, 2024.