

# 人工智能时代《定性数据统计分析》课程 的教学改革探索

邓世容, 陈玉蓉, 许晓菲\*

武汉大学 数学与统计学院, 湖北 武汉 430072

DOI:10.61369/ASDS.2026020001

**摘要:** 人工智能 (AI) 与大数据技术的迅猛发展对传统统计学教育, 特别是《定性数据统计分析》教学提出了新的挑战与要求。本文探讨如何面向人工智能时代需求, 重构课程目标、内容与教学模式。改革的核心在于构建以“数智案例”为驱动的教学体系, 将机器学习、文本挖掘等 AI 前沿方法融入经典定性分析框架, 并着力建设跨学科、多场景的案例资源库。通过推动教学方法向项目驱动、合作探究式转型, 并建立多元化的实践与评价体系, 旨在培养具备扎实统计理论、熟练数据科学工具与跨界解决问题能力的复合型人才。

**关键词:** 人工智能; 定性数据分析; 教学改革; 案例教学; 跨学科; 实践能力

## Teaching Reform of “Statistical Analysis of Categorical Data” in the Age of AI

Deng Shirong, Chen Yurong, Xu Xiaofei\*

School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072

**Abstract:** The rapid development of artificial intelligence (AI) and big data technology poses new challenges and demands for traditional statistics education, particularly for the teaching of “Categorical Data Analysis”. This paper explores how to restructure course objectives, content, and teaching models to meet the needs of the AI era. The core of the reform lies in constructing a “data-intelligence case”-driven teaching system, which integrates cutting-edge AI methods such as machine learning and text mining into the classical categorical analysis framework, and focuses on building a cross-disciplinary, multi-scenario case resource library. By promoting a shift in teaching methods towards project-driven and collaborative inquiry-based approaches, and establishing a diversified practical and evaluation system, the aim is to cultivate versatile talents with solid statistical theory, proficiency in data science tools, and the ability to solve problems across disciplinary boundaries.

**Keywords:** artificial intelligence; categorical data analysis; teaching reform; case-based teaching; cross-disciplinary; practical ability

## 引言

在人工智能 (AI) 与大数据成为社会经济发展核心驱动力的时代, 数据形态与分析需求日益复杂。定性数据 (或称分类数据、属性数据) 作为反映事物属性、类别与关系的重要信息载体, 在社会科学、生物医学、市场分析、文本挖掘等领域呈现出海量、高维、多源的新特征。传统的《定性数据统计分析》课程, 其教学内容多侧重于列联表分析、Logistic 回归、对数线性模型等经典统计方法<sup>[1]</sup>, 在教学模式上往往存在重理论推导、轻实践应用, 案例陈旧且与前沿技术脱节等问题, 难以满足行业对具备 AI 素养和数据驱动决策能力的高层次应用统计人才的需求<sup>[2]</sup>。因此, 依托数智化技术, 对《定性数据统计分析》课程进行系统性、前瞻性的教学改革, 不仅是学科发展的内在要求, 更是服务国家战略需求、提升人才培养质量的关键举措。武汉大学以专业学位研究生教育综合改革项目为契机, 启动了《定性数据统计分析》数智案例特色课程建设, 本文旨在梳理并总结此项改革的设计理念、具体内容与实施路径。

基金项目: 2025年度武汉大学专业学位研究生教育综合改革项目-数智案例特色课程建设项目《定性数据统计分析》; 2024年度武汉大学本科教育质量建设综合改革项目-统计学一流本科专业建设探索项目子项目-《时间序列分析》课程体系建设。

作者简介:

邓世容, 武汉大学数学与统计学院, 副教授, 硕士生导师, 研究方向为纵向数据的建模与推断、生存分析、遗传统计等;

陈玉蓉, 武汉大学数学与统计学院, 副教授, 研究方向为生存分析。

通讯作者: 许晓菲, 武汉大学数学与统计学院, 特聘副研究员, 硕士生导师, 研究方向为复杂时间序列分析和预测、高维数据分析和网络数据分析等。

## 一、教学改革的核心驱动力与目标定位

本次课程改革的驱动力是多维度、立体化的，既源于外部环境的深刻变化，也基于对教育本质的重新思考。

### （一）核心驱动力

1. 技术驱动的知识迭代需求：机器学习、深度学习、自然语言处理等 AI 技术的突破，为解决高维分类、文本情感分析、图结构关系推断等复杂定性数据分析问题提供了强大工具<sup>[3, 4]</sup>。课程必须回应这种技术变革，将新工具、新方法有机整合进传统知识体系。

2. 产业升级的人才能力需求：用人单位不再满足于毕业生只会进行标准的假设检验或拟合预设模型，而是要求其能够从混乱的真实数据中定义问题、清洗整理、特征工程、选择合适的模型（包括经典统计模型和机器学习模型）、评估优化并商业洞察<sup>[5]</sup>。课程需要模拟这一完整的数据科学工作流程。

3. 学科交叉的研究范式需求：当代重大科学和社会问题的解决日益依赖于多学科数据的融合分析。定性数据分析作为通用方法论，需要与具体领域知识（如基因组学、社会学理论、消费者行为学）深度结合。课程应培养学生的跨界对话与合作能力。

4. 研究生教育的内涵发展需求：专业学位研究生教育尤其强调高层次、应用型、复合型人才的培养<sup>[6]</sup>。改革需强化实践创新环节，提升学生解决实际问题的职业胜任力。

### （二）目标定位

基于上述驱动力，本课程的总体定位为，建设一门深度融合人工智能思维、现代统计理论与多领域应用实践，以“前沿引领、案例驱动、交叉融合、能力为本”为鲜明特色的研究生层次核心课程<sup>[7, 8]</sup>。因此，本课程改革确立了清晰的三层目标定位：

1. 知识结构更新与重构：系统性地将高维数据处理技术（如正则化方法）、机器学习分类算法（如随机森林、支持向量机）<sup>[9, 10]</sup>、文本挖掘基础（如主题模型、词嵌入）、贝叶斯网络等前沿内容，与传统定性分析核心理论（列联表、Logistic 模型、对数线性模型）<sup>[1]</sup>进行逻辑衔接与对比融合，形成“经典-前沿”贯通的知识图谱。

2. 核心能力培养与提升：重点锻造四维能力：（1）技术实现能力：熟练使用 R/Python 进行数据操作、可视化及模型实现；（2）方法选择与评估能力：能根据问题性质、数据特征选择并合理评估经典统计模型与机器学习模型；（3）问题拆解与解决能力：通过完整案例和项目，训练从业务问题到统计问题转化、分析设计、结果解释与决策建议的全链条思维能力；（4）跨学科沟通与协作能力：通过跨学科案例分析和团队项目，学会与领域专家对话，理解问题背景。

3. 综合素养塑造与拓展：深化对数据伦理（隐私、公平性）、算法偏差的认知；培养严谨求实的科学精神与可重复研究的习惯；激发通过数据分析探索未知的创新意识。

## 二、教学改革的主要内容与创新实践

本课程改革将围绕“内容重构、资源创新、方法转变、评价

革新”四个支柱展开。

### （一）课程内容体系的重构与前沿深度融合

1. 新的课程内容体系设计遵循“固本强基、与时俱进、交叉拓展”原则。保留并强化核心基础模块：确保学生对属性数据分布、二维及多维列联表的独立性/关联性检验、Logistic 回归模型（二分类与多分类）、对数线性模型等经典理论有扎实理解<sup>[1]</sup>，这是理解更复杂模型的基石。参考教材如王静龙等编著的《属性数据分析》提供了坚实的理论基础。

2. 在经典模块中系统性注入 AI 视角与扩展。如列联表分析进阶：在讲授传统卡方检验后，引入大规模稀疏列联表的分析挑战；回归模型的扩展与对比：在深入讲解 Logistic 回归后，平行引入：（1）正则化 Logistic 回归（L1/Lasso, L2/Ridge），讲解其处理高维共线数据、进行特征选择原理；（2）机器学习分类模型簇，如决策树、随机森林、梯度提升机（如 XGBoost）、支持向量机<sup>[9, 10]</sup>，通过同一案例数据，比较这些“算法驱动”模型与“概率模型驱动”的 Logistic 回归在预测性能、解释性、计算效率等方面的优劣等。

### 3. 增设前沿与交叉应用专题模块<sup>[11]</sup>

通过构建不同的应用专题，以应对多学科交叉的复杂研究场景。例如，（1）聚焦文本数据的定性转化，涵盖文本预处理、主题模型及词向量（如 Word2Vec）等技术，将非结构化文本转化为结构化特征，支撑情感分析与主题挖掘。（2）高维与网络数据分析，例如在单细胞 RNA-seq 数据中通过高维特征进行细胞类型注释，或基于网络模型推断细胞通讯与社交社区结构，体现模式识别方法在生命科学与网络科学中的迁移与应用。（3）引入因果推断框架，通过倾向得分匹配等方法在观察性数据中控制混杂变量，推动分析从相关性识别迈向因果机制探索，从而系统提升研究的科学深度与实际解释力。

### （二）以“数智案例库”为核心的教学资源立体化建设

案例库是连接理论与实践的桥梁。本课程致力于打造一个“来源真实、场景多元、配套完整、动态更新”的数智案例库。

#### 1. 案例来源的多元化

科研反哺教学案例：将项目团队在生物统计、计算社会科学等方向的在研成果，经过教学化设计，转化为前沿科研案例。学生能接触到真实的科研问题、数据和代码。

行业实践提炼案例：积极与企业、政府机构合作，将脱敏后的真实业务问题（如用户流失预测、政策效果评估）转化为教学案例。同时，充分利用 Kaggle (<https://www.kaggle.com>)、天池 (<https://tianchi.aliyun.com>) 等数据科学竞赛平台，以及 UCI (<https://archive.ics.uci.edu>)、GEO (<https://www.ncbi.nlm.nih.gov/geo/>) 等公共数据集，选取经典问题（如泰坦尼克号生存预测、鸮尾花分类）和前沿问题等。

跨学科综合设计案例：自主设计需要综合运用社会学问卷设计、经济指标、统计建模的案例，例如“基于多源数据（问卷+消费记录）的消费者可持续行为影响因素研究”<sup>[8]</sup>。

#### 2. 案例资源的数字化与结构化

每个人库案例均以标准模板构建，包含：（1）案例背景与业

务问题描述；（2）原始数据集（或获取链接）；（3）数据字典与清洗建议；（4）分步骤的R/Python代码脚本（含详细注释）；（5）多种分析方法（经典与机器学习）的实现与比较；（6）分析报告范文或要点提示；（7）教学使用指南（课时安排、讨论问题、知识点映射）。所有资源集成于课程在线平台（如校内学习管理系统），方便师生检索、下载与版本控制。这种基于项目的案例资源建设模式，符合建构主义学习理论，有助于学生在解决真实问题的过程中主动构建知识。

### （三）教学方法的协同创新与“学生中心”模式转变

推行“理论精讲、案例深析、项目实战、技术实训”四位一体的混合式教学模式。

1. 翻转课堂与理论精讲：将基础理论、公式推导等内容制作成短视频或精读材料，要求学生课前自学。课堂时间则用于梳理脉络、答疑解惑、深化理解，特别是讲清不同方法背后的统计思想与适用边界。

2. 案例驱动的沉浸式工作坊：针对核心知识点，设计配套的案例工作坊。课堂上，教师引导学生分组“啃”案例：从数据探索性分析开始，到尝试不同方法，比较结果，讨论“为什么这个方法在这里更有效？”“这个结果在业务上意味着什么？”，强化理论在具体情境中的应用。这种方法强调探究式学习，能有效提升学生的参与度和理解深度<sup>[7]</sup>。

3. 贯穿学期的项目式学习（PBL）：学生3-4人一组，在学期初选择一个感兴趣的课题（可从题库选或自拟，需教师审核）。在教师和助教的阶段性指导下，利用整个学期的时间，完成一个微型数据分析项目，最终提交项目报告、可复现代码库和答辩展示。PBL模拟真实工作场景，是对学生综合能力的全面淬炼，并能显著提高学生的学习动机和解决问题的能力<sup>[8]</sup>。

4. 强化的技术实训环节：安排专门的实验课或线上编程练习，聚焦数据清洗（如处理缺失值、类别不平衡）、可视化（ggplot2）、模型训练与调参、报告编写等实用技能，扫除技术障碍，让学生能把更多精力集中于方法论思考。

### （四）评价体系的多元化、过程化与能力导向改革

建立覆盖学习全过程、聚焦能力成长的评价体系，打破“一考定乾坤”的模式。

1. 过程性评价（占总评60%）：个人作业与编程练习（15%）：检验对基础理论和代码实现的基本掌握。案例分析报告（个人或小组，15%）深度评估对特定案例的分析逻辑、方法应用恰当性、结果解读深度和批判性思维。期末项目（小组，30%）：这是核心考核，从问题定义、数据质量、方法创新性与合理性、代码质量与可重复性、报告撰写与可视化水平、答辩表现等多个维度进行综合评分。引入企业专家或跨学科教师参与项目评审。

2. 课堂参与与贡献（10%）：鼓励提问、讨论和分享。

3. 终结性评价（占总评30%）：期末考试：题型侧重应用与理解，减少纯记忆性内容。可包括：给定场景的方法选择与理由阐述、对一段分析结果（或代码输出）的解读、小型案例的简要分析设计等，重点考察知识融会贯通的能力。

同时评价主体也可多元化，结合教师评价、同学互评（尤其

在小组项目中）、学生自评，并引入少量企业导师或领域专家评价（针对项目），使评价更全面、客观。这种多元评价体系有助于更全面地反映学生的学习成果。

## 三、改革实施的支撑体系、挑战与应对策略

### （一）关键支撑体系建设

“双师型”与“学习型”师资队伍：通过设立教研小组定期研讨、资助教师参加数据科学/AI短期培训与学术会议、鼓励教师与产业界合作或开展交叉学科研究、聘请企业资深数据分析师担任兼职讲师或项目导师等方式，持续提升教师团队的实践前沿敏感度和跨领域教学能力。

数字化教学与实践平台：建设或利用现有的高性能计算集群或云服务器，为学生提供统一的、预装必要软件和库的云端编程环境（如RStudio Server），解决学生本地安装配置困难、计算资源不足的问题。

制度化激励机制：将案例开发、教学改革成果纳入教师的绩效考核、职称评定和教学评优体系，激发教师投入改革的持续动力<sup>[6, 8]</sup>。

### （二）面临的主要挑战与应对策略

挑战一：有限课时与海量内容的矛盾。实施“核心必修+前沿选修”模块化设计。核心经典理论和方法保证所有学生掌握；前沿专题（如文本挖掘、高维数据）以工作坊、专题讲座或线上资源的形式提供，供学有余力的学生选修。充分利用线上资源进行课前预习和课后拓展。

挑战二：案例库的持续维护、更新与知识产权。建立案例库的协同建设与动态更新机制，鼓励历届优秀学生项目经打磨后转化为新案例。明确案例使用规范，对涉及敏感数据的商业案例进行深度脱敏和教学化改编，优先选用开源数据或合作开发。

挑战三：学生基础差异与个性化教学。提供分层次的编程入门指导材料和“技能加油站”微视频。在PBL项目中，允许学生根据自身兴趣和能力选择不同难度的课题。实施助教辅导制和小组内的互帮互助，这符合差异化教学的原则<sup>[7]</sup>。

挑战四：新型能力评价的效度与信度。开发详细的项目评价量规，提前向学生公布，使评价标准透明化。对“批判性思维”“创新性”等较主观的维度，通过具体的行为描述（如“能识别并讨论所用模型的至少一个局限性”）来使其可观测、可评估。使用评价量规能有效提高评价的一致性和公平性。

## 四、结论与展望

面向人工智能时代的《定性数据统计分析》课程教学改革，绝非简单的技术工具叠加或案例数量的增加，而是一场涉及教育理念、知识体系、教学方法与评价范式的系统性、深层次变革<sup>[2-5]</sup>。武汉大学此次以“数智案例”为特色和抓手的改革实践，将探索一条将经典统计理论根基与前沿人工智能方法有机融合，并通过项目式、探究式学习将其转化为学生核心能力的可行路径<sup>[7, 8][11]</sup>。

改革强调“从实践中学习，为实践而学习”，旨在培养出能驾驭智能时代复杂数据分析挑战的、既懂统计原理又擅技术实现、既有专业深度又有跨界视野的新一代应用统计人才。

展望未来，课程改革仍需在以下几个方面持续深化：一是进一步拓展产学合作生态，引入更多真实的产业问题与数据资源；二是探索人工智能赋能教学本身，如利用大语言模型作为编程辅

助答疑工具、自动生成模拟数据用于教学、智能分析学生代码常见错误模式等<sup>[9]</sup>；三是加强教学研究的实证跟踪，通过前后测、访谈、毕业生追踪等方式，科学评估改革对学生学习效果和长期职业发展的影响。唯有坚持持续迭代与创新，才能确保课程建设始终与技术进步和时代需求同频共振，为我国高层次应用型统计人才的培养贡献坚实力量。

## 参考文献

- [1] 王静龙, 梁小筠, 王黎明. 属性数据分析 [M]. 北京: 高等教育出版社, 2013.
- [2] 邓国强, 陈佳蕙. 新质生产力引领下的大数据时代应用统计硕士人才培养研究 [J]. 江苏科技信息, 2025, 42(08): 9-13.
- [3] 徐宗本. 人工智能的10个重大数理基础问题 [J]. 中国科学: 信息科学, 2021, 51(12): 1967-1978.
- [4] 游士兵, 徐小婷. 统计学方法的发展及其在大数据中的应用 [J]. 统计与决策, 2020, 36(04): 31-35.
- [5] 张娟娟, 朱芳芳. 人工智能背景下统计学课程教学改革探讨 [J]. 对外经贸, 2022, (10): 150-153.
- [6] 孙毅. 统计学专业研究生课程教学的研究与思考 [J]. 科教导刊, 2021, (16): 32-34+53.
- [7] 马鹏程. 基于 OBE 理念的应用统计学专业课程教学模式改革与实践——以属性数据分析课程为例 [J]. 大众科技, 2025, 27(05): 69-73.
- [8] 马鹏程, 吴孙勇, 刘红卫, 等. 基于“一核两翼三化四融”模式的教学探究与实践——以应用统计学专业课“属性数据分析”为例 [J]. 科技风, 2025, (11): 99-101.
- [9] 李航. 统计学习方法 [M]. 第2版. 北京: 清华大学出版社, 2019.
- [10] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.
- [11] 徐锋. “属性数据分析”的课程案例和课程考核改革 [J]. 大众科技, 2023, 25(11): 108-110+114.