

基于分层广义线性模型的高校教学质量影响因素研究

周晓东¹, 王云娟^{2*}

¹上海对外经贸大学统计与数据科学学院, 上海

²上海立信会计金融学院统计与数学学院, 上海

收稿日期: 2025年10月5日; 录用日期: 2025年10月26日; 发布日期: 2025年11月10日

摘要

本文从微观数据的视角, 基于实际案例运用分层广义线性模型研究影响教学质量的影响因素。实证结果表明分层广义线性模型更加适合教学质量评价中数据分层和异质性的特点, 较传统模型分析结果更优。

关键词

教学质量, 分层广义线性模型, 异质性, 教学评价

Research on Factors Influencing Teaching Quality in Higher Education Based on Hierarchical Generalized Linear Models

Xiaodong Zhou¹, Yunjuan Wang^{2*}

¹School of Statistics and Data Science, Shanghai University of International Business and Economics, Shanghai

²School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai

Received: October 5, 2025; accepted: October 26, 2025; published: November 10, 2025

Abstract

This paper examines the factors influencing teaching quality from a micro-level data perspective, utilizing hierarchical generalized linear models based on real cases. The empirical results indicate

*通讯作者。

文章引用: 周晓东, 王云娟. 基于分层广义线性模型的高校教学质量影响因素研究[J]. 统计学与应用, 2025, 14(11): 45-53. DOI: 10.12677/sa.2025.1411309

that hierarchical generalized linear models are more suitable for analyzing data with inherent hierarchical structures and heterogeneity in the evaluation of teaching quality, yielding superior results compared to traditional models.

Keywords

Quality of Teaching, Hierarchical Generalized Linear Model, Heterogeneity, Teaching Evaluation

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

高校教学质量是人才培养的核心保障,其科学评价对于推动教学改进、提升教育质量具有重要意义[1]。现有研究已从多视角构建了教学质量评价的方法体系,如因子分析[2]、结构方程模型[2]、多面 Rasch 模型[3]、云-灰关联分析[4]、层次分析法[5]、BP 神经网络[6]、熵权法[8]以及混合偏态分布模型[9]等。近年来,亦有学者引入四层次分析框架[10]或结合偏最小二乘结构方程与人工神经网络[11],尝试提升评价的精确性与综合性。

然而,尽管方法日趋多元,多数研究在数据处理上仍存在一个局限性:未能充分考虑教学评价中普遍存在的嵌套数据结构。在实际教学环境中,学生个体嵌套于班级或课程中,课程又嵌套于教师或学院之下,形成典型的层次结构。这种数据结构导致观测值之间不具备独立性,若沿用传统回归或一般线性模型,不仅会低估标准误、增大I类错误风险,也无法有效分离不同层级(如学生层面与教师层面)的变异来源,从而导致估计偏误与推论失真。事实上,教育数据的嵌套特性及其对建模方法的内在要求,早在1990年代就已得到方法论学者的系统阐述。例如, Morris (1995) [12]明确指出教育数据天然适合分层建模,并深入探讨了早期模型在估计与推断中面临的挑战,强调了模型诊断与方差成分准确量化的重要性。

随着理论发展与计算工具的进步,分层模型在教育研究中的应用已逐步深入并证实其价值。例如,纪江明与葛羽屏(2015) [13]利用分层线性模型分析基础教育满意度,清晰地揭示出市民个体特征是解释城市间差异的主因,而城市层面财政指标的解釋力相对有限,这凸显了分层模型在辨析变量影响层级上的独特优势。王芳(2018) [14]基于分层线性模型研究了大学生教学满意度影响因素。郭建鹏等(2022) [15]基于全国大规模高校教师数据,采用两水平 HLM 模型,进一步揭示了教师教学投入同时受到个体因素(如性别、职称、自我效能感)和学校环境因素(如学校类型、教学环境)的跨层影响,并识别出学校环境通过自我效能感影响教学投入的中介路径。这些研究,连同 Sulis 等(2019) [16]运用广义线性混合效应模型对学生评教数据的分析,共同构成了将分层模型应用于教育评价问题的坚实实证基础,证明了其在处理异质性和进行跨层比较方面的优越性。

尽管如此,将分层广义线性模型(HGLM)系统应用于高校教学质量影响因素分析的研究仍相对匮乏。现有研究多集中于连续型结局变量(如满意度、投入度)的 HLM 分析,而教学评价实践中大量存在的分类、顺序等非连续型数据(如等级制评教结果)、非负等则需要更具普适性的 HGLM 框架。多数评价文献在构建模型时,未能从“数据结构驱动方法选择”的视角,系统回应嵌套数据与变量非正态性带来的双重挑战。值得注意的是,分层广义线性模型(HGLM)因其在处理层次结构、非独立数据方面的灵活性与鲁棒性,已在诸多社会科学领域成为成熟的分析工具。例如,在经济预测[17]、社会参与研究[18]、保险费率厘定

[19]以及公共服务与移民定居意愿[20]等研究中, HGLM 被成功用于分析复杂的多层作用机制, 充分验证了其方法论的普适性与有效性。

因此, 本研究旨在弥补上述研究空白, 通过引入分层广义线性模型, 系统分析高校教学中存在的多层数据结构, 并精准识别影响教学质量的关键学生层面与教师层面因素。本研究不仅从方法上回应了嵌套数据与非正态因变量带来的统计挑战, 拓展了 HGLM 在教育质量评价中的应用边界, 也在实证层面为后续基于多源、多层数据的教学改进策略提供了更可靠的分析基础。

2. 分层广义线性模型

令 y_{ij} 为第 i 个个体(组)的第 j 个观测值。分层广义线性模型由以下三部分构成:

(1) 分布形式:

$$y_{ij} | u_i \sim f_{Y_{ij}}(y_{ij} | \theta_{ij}, u_i), u_i \sim f_{U_i}(u_i),$$

其中

$$f_{Y_{ij}}(y_{ij} | \theta_{ij}, u_i) = \exp\left\{\phi^{-1}\left[y_{ij}\theta_{ij} - \psi(\theta_{ij})\right] + c(y_{ij}, \phi)\right\},$$

其中 θ_{ij} 称为典型参数, ϕ 称为散布参数。上述指数分布族包含一些常见的分布, 如指数分布、正态分布、二项分布、伽玛分布等。

(2) 系统部分:

$$\eta_{ij} = x'_{ij}\beta + z'_{ij}u_i,$$

其中 x_{ij} 和 z_{ij} 分别对应固定效应系数 β 和随机效应项 u_i 的解释变量所构成的向量。

(3) 连接函数:

$$g(\mu_{ij}) = \eta_{ij}, \mu_{ij} = E(Y_{ij} | u_i).$$

注意当上述模型中的区组效应 $u_i = 0$ 时, 上述三部分构成的模型称为广义线性模型。

对于上述分层广义线性模型, 其对数似然函数可以写为

$$L = \log \int \prod_{ij} f_{Y_{ij}}(y_{ij} | \theta_{ij}, u_i) f_{U_i}(u_i) du_i.$$

最大化上述对数边际似然函数, 即可求得模型中未知参数(包括 $Y_{ij} | u_i$ 以及 u_i 分布中的参数)的极大似然估计值。然而, 由于上述对数边际似然中包含高维积分, 因而增加了求解极大似然估计的难度。为此, 近年来相关学者提出了多种解决方案, 如 Gaussian-Hermite 近似或自适应 Gaussian-Hermite 近似、蒙特卡罗(MC)近似结合 Newton-Raphson 算法, MCEM 算法, Laplace 近似、MQL、PQL 方法、GEE 方法等。相关工作回顾可以参看文献[21][22]。除极大似然估计方法外, 在给定模型参数先验条件下, 亦可采用贝叶斯方法获取模型参数的贝叶斯估计。

3. 应用案例

本节我们通过两个实际例子介绍上述分层广义线性模型在教学评价中的应用。

3.1. 影响统计学通过率的影响因素分析

我们收集了某财经类高校非统计学专业统计学学习相关的数据。数据包含 549 名学生的期末考试成绩、任课老师、概率论课程成绩、学生性别、所在教学班级等信息, 其中涉及到 6 位任课老师, 6 个不同

专业, 9 个不同教学班级。我们关注的问题是哪些因素会影响学生该课程的通过率。图 1 给出了不同任课老师, 不同专业, 不同教学班级、不同性别学生、不同概率成绩分数段通过情况的条形图。卡方检验结果表明在 0.01 显著性水平下变量性别、概率论成绩对不及格率有显著影响, 而变量“所在教学班级”在 0.1 显著性水平下通过检验, 其余变量, 如任课老师, 专业却不显著。

为了考虑不同因素对学生通过率或不及格率的影响, 将学生是否通过统计学考试作为因变量, 而学生的性别(gender)、任课老师(teacher)、概率论成绩(score_prob)、所在教学班级(Class), 专业(major)作为解释变量建立广义线性回归模型。同时考虑到同一教学班级的学生成绩可能存在一定的相关性, 因此将学生所在教学班级作为区组因素, 最终建立分层广义线性回归模型:

$$y_{ij} | u_i \sim B(1, \theta_{ij}), u_i \sim N(0, \sigma^2),$$

$$\text{logit}(\theta_{ij}) = \log \frac{\theta_{ij}}{1 - \theta_{ij}} = \eta_{ij},$$

$$\eta_{ij} = \beta_0 + \beta_1 * \text{score_prob}_{ij} + \beta_2 * \text{gender}_{ij} + \beta_3 * \text{teacher}_i + \beta_4 * \text{major}_{ij} + u_i.$$

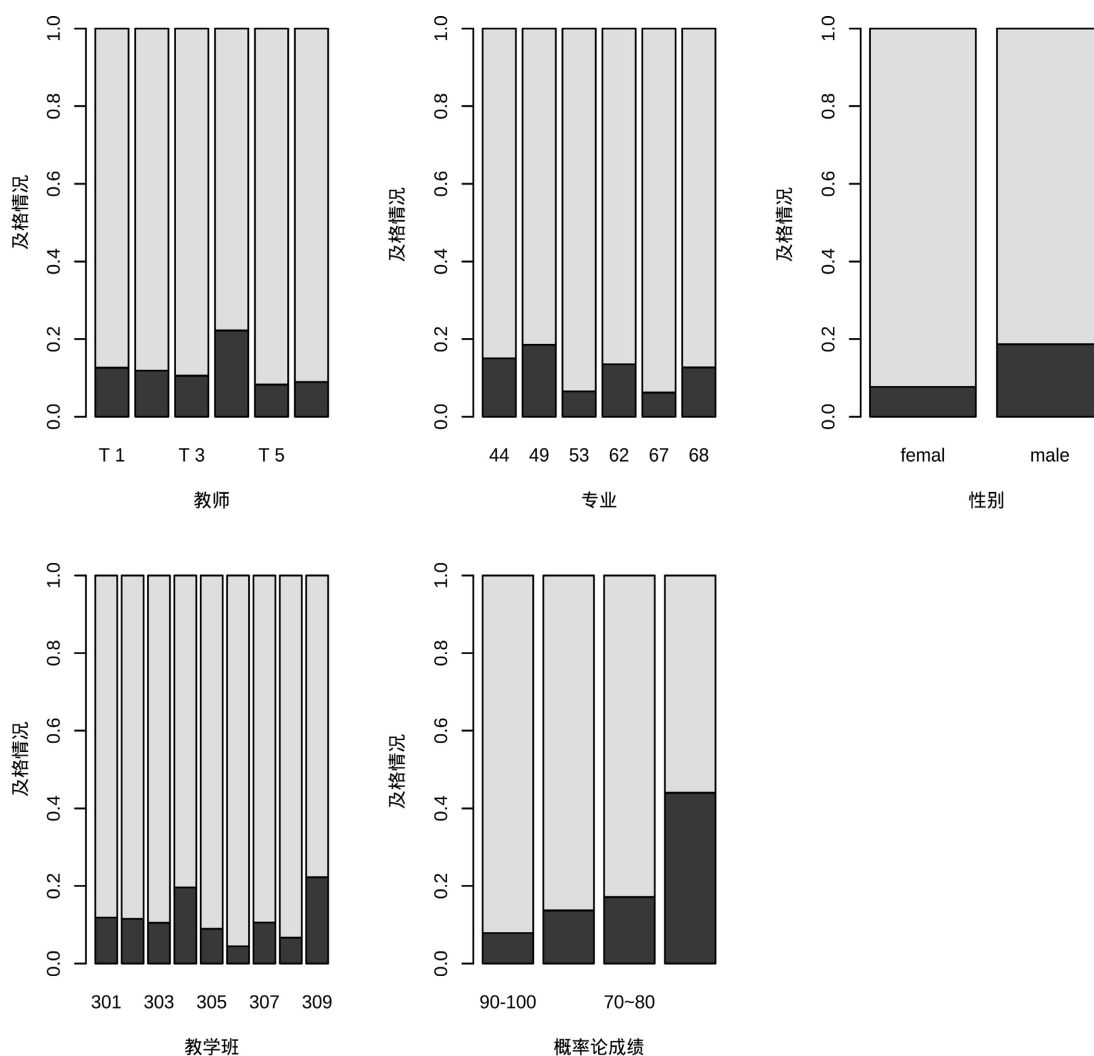


Figure 1. Bar chart of failure rates under different factors

图 1. 不同因素下不及格率条形图

应用 R 语言的 lme4 包中的 *glmer* 函数, 所得回归结果见表 1。

表 1 中模型 1 结果表明专业和任课老师均不显著, 而学生的性别与概率论成绩在 0.05 显著性水平下是显著的。此时随机效应的标准差接近于 0。去掉模型 1 中的不显著变量, 得到模型 2 的回归结果。同时, 为了比较不同模型的差异, 我们去掉模型 1 和 2 中的随机效应项, 建立一般的广义线性回归模型, 回归分析结果见模型 3 结果。模型 2 和模型 3 的回归系数的估计和标准差相差不大。另外由 AIC 和 BIC 的值可知模型 3 最优。上述结果表明, 学生的性别及概率论课程的学习对统计学通过率的影响比较显著, 而与任课老师、专业、所在教学班级等变量关系不大。分层广义线性模型与不考虑区组的一般广义线性模型分析结果差异不大。另外我们也采用贝叶斯方法(R 包 blme 中 *bglmer* 函数, 其中先验采用默认值)对上述模型进行了估计, 结果与表 1 中结果相差不大。

Table 1. Results of the regression analysis on the pass rate of the statistics course

表 1. 统计学课程通过率回归分析结果

	模型 1			模型 2			模型 3		
	估计量	标准差	P 值	估计量	标准差	P 值	估计量	标准差	P 值
(Intercept)	-1.55	1.26	0.22	-1.09	1.02	0.29	-1.10	0.98	0.26
Score_prob	0.04	0.01	0.00**	0.04	0.01	0.00***	0.04	0.01	0.00***
gender	-0.94	0.29	0.00**	-0.98	0.28	0.00***	-0.98	0.27	0.00***
teacher2	0.48	1.57	0.76	-	-	-	-	-	-
teacher3	0.92	0.66	0.17	-	-	-	-	-	-
teacher4	-0.92	0.95	0.33	-	-	-	-	-	-
teacher5	0.64	0.51	0.21	-	-	-	-	-	-
teacher6	-0.02	0.91	0.98	-	-	-	-	-	-
major49	1.55	0.81	0.05	-	-	-	-	-	-
major53	1.40	0.91	0.13	-	-	-	-	-	-
major62	0.29	0.65	0.65	-	-	-	-	-	-
major67	1.23	0.92	0.18	-	-	-	-	-	-
major68	0.33	1.59	0.83	-	-	-	-	-	-
Random effects		0.000			0.069		-	-	-
AIC		382.80			376.00			374.05	
BIC		433.10			393.30			386.97	

注: 记号 “***”、“**”、“*”、“.” 分别表示显著性水平 0.001、0.01、0.05、0.1 下显著。

针对上述发现, 今后我们在提升统计学课程学习效果时, 应增强学生在统计学前修课程概率论中的学习, 为统计学的学习奠定良好基础。同时在课程教学管理过程中, 注重性别差异对课程教学的影响, 针对不同性别特点选择不同的教学资料、教学策略和方法、考核方式以满足不同性别学生的学习需求, 提高课程教学效果。

3.2. 教学评价的影响因素分析

Hamermesh 和 Parker [23]基于一组大学教师的评教数据研究了教师颜值对教学评教的影响。具体数据见 R 语言 ARE 包 “TeachingRatings”。数据共 463 条记录, 包含受评教师的年龄(age), 是否为少数裔

(minority), 性别(gender), 学分(credits), 颜值(beauty), 教学评教分数(eval), 英语是否为母语(native), 是否为终身教职(tenure), 是否为高年级课程(division), 是否为单学分选修课(credits)等信息。

本案例中我们主要关注影响教师评教分数的影响因素。图 2 给出了不同因素不同水平下评教分数的箱线图及变量与评教分数的散点图。图 2 结果结合 t 检验或相关性检验(显著性水平为 0.05)表明, 性别(gender), 英语是否为母语(native), 是否为终身教职(tenure), 是否为单学分选修课(credits), 这四个分类变量对评价结果影响显著。另外颜值(beauty)与评价分数(eval)分数之间有较强的相关性。

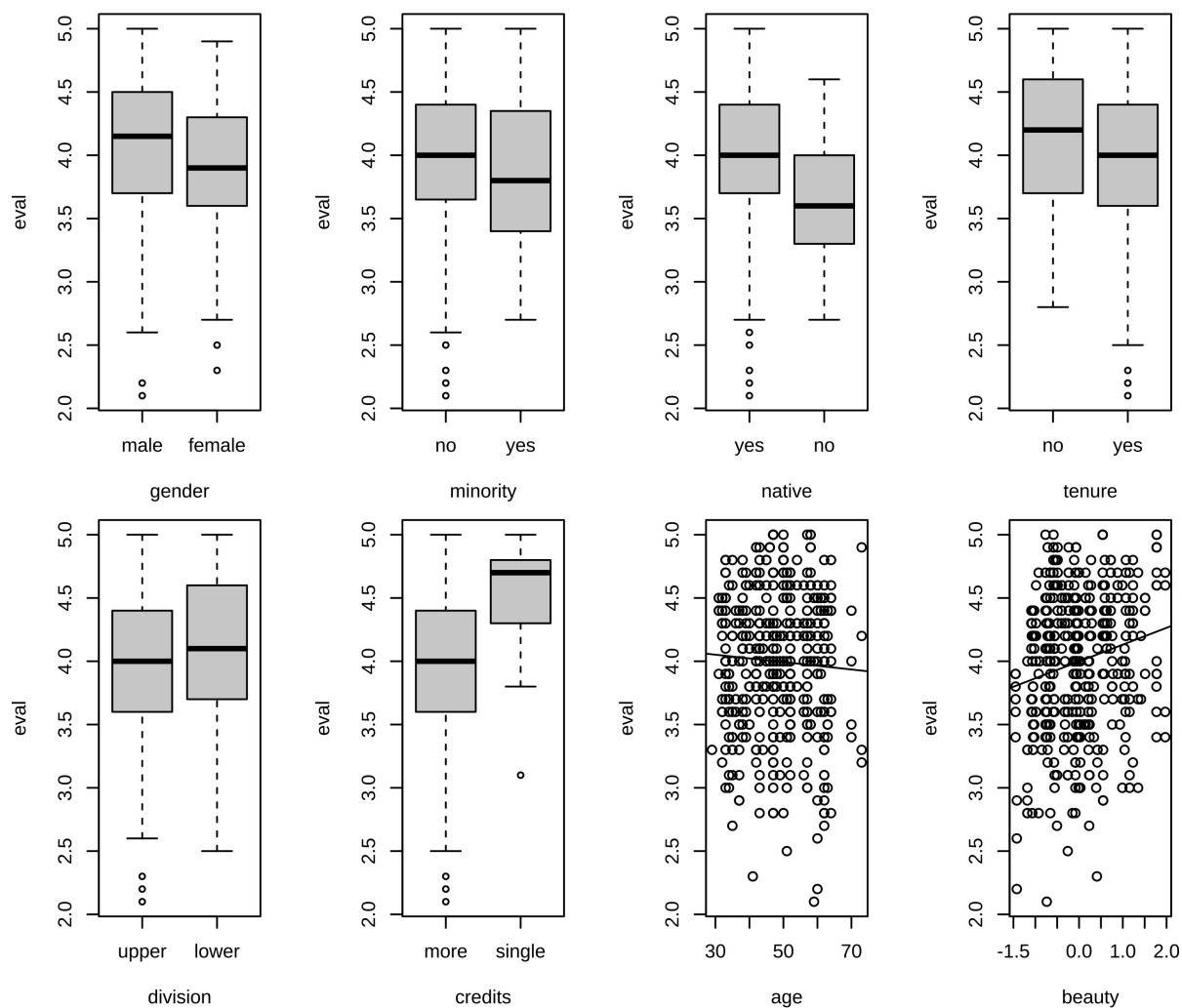


Figure 2. Box plot or scatter plot of teaching evaluation scores under different influencing factors

图 2. 不同影响因素下教学评教分数箱线图或散点图

注意到评教分数为非负值, 因此我们首先采用伽玛回归模型对数据进行建模分析。伽玛回归模型如下:

$$\begin{aligned}
 y_i &\sim \Gamma(\theta_i, \phi), \\
 g(\theta_i) &= \log \theta_i = \eta_i, \\
 \eta_i &= \beta_0 + \beta_1 * beauty_i + \beta_2 * age_i + \beta_3 * gender_i + \beta_4 * minority_i \\
 &\quad + \beta_5 * native_i + \beta_6 * tenure_i + \beta_7 * division_i + \beta_8 * credits_i
 \end{aligned}$$

采用 R 语言基础包中的 `glm` 函数进行回归分析, 相关回归结果见表 2 中的模型 1。去除模型 1 中的

不显著变量, 得到约简模型 2。结果表明颜值、性别、课程学分都对评教有显著影响, 而年龄、是否为终生教职、是否为高年级课程对评教影响不显著。

另外由于数据中一个老师参与评教的课程可能不止一门, 同一个老师的评教分数可能存在相关性, 因此考虑将评价教师作为区组因素, 建立分层伽玛回归模型:

$$\begin{aligned} y_{ij} | u_i &\sim \Gamma(\theta_{ij}, \phi), u_i \sim N(0, \sigma^2), \\ g(\theta_{ij}) &= \log \theta_{ij} = \eta_{ij}, \\ \eta_{ij} &= \beta_0 + \beta_1 * \text{beauty}_i + \beta_3 * \text{gender}_i + \beta_4 * \text{minority}_i \\ &\quad + \beta_5 * \text{native}_i + \beta_6 * \text{tenure}_i + \beta_8 * \text{credits}_{ij} + u_i \end{aligned}$$

利用 R 语言 lme4 包中的 *glmer* 函数进行回归分析。结果见表 2 中的模型 3, 结果表明变量是否为少数裔、英语是否为母语在模型中都不显著, 去掉不显著变量后得到约简分层回归模型 4。根据 AIC 和 BIC 准则, 模型 4 为最优模型。模型 4 表明: 颜值对评教有正的影响, 男教师比女教师评分倾向于高, 学分少的课程比学分多的课程评分倾向于高。类似于案例 1, 我们也采用贝叶斯估计方法对上述模型 3 和模型 4 进行了估计, 结果与表 2 中的结果相差不大。

高校课堂教学质量评价的科学化, 强烈依赖于能够精准衡量并排除多种干扰因素的定量模型。传统的评价指标体系虽涵盖了教学态度、内容、方法等维度[24], 但越来越多的证据表明, 学生评教数据深受嵌套结构与系统性偏见的双重困扰。例如, 最新研究发现, 学生基于“公平性感知”的评价具有高度不稳定性, 且这种偏见无法通过简单的统计控制来消除 (Buchanan *et al.*, 2025 [25])。这揭示出, 在评价指标之外, 数据本身的结构与效度问题已成为影响结果科学性的关键。若忽略评教数据中嵌套关系, 以及性别、评分公平性等潜在偏见, 任何基于传统线性模型的分析都可能得出误导性的结论。因此, 我们必须借助如分层广义线性模型这类能够同时定义随机效应以处理嵌套数据、并纳入协变量以控制偏见的现代定量手段, 才能有效剥离无关因素的干扰, 获得对教学质量更真实、更科学的评价结果。

Table 2. Regression results on teaching evaluation scores
表 2. 评教分数的回归分析结果

	模型 1			模型 2			模型 3			模型 4		
系数	估计量	标准差	P 值	估计量	标准差	P 值	估计量	标准差	P 值	估计量	标准差	P 值
Intercept	1.42	0.02	0.00***	1.40	0.01	0.00***	1.37	0.02	0.00***	1.37	0.02	0.00***
beauty	0.04	0.01	0.00***	0.04	0.01	0.00***	0.04	0.02	0.05·	0.04	0.02	0.05*
scale (age)	-0.01	0.01	0.35	-	-	-	-	-	-	-	-	-
genderfemale	-0.05	0.01	0.00***	-0.04	0.01	0.00***	-0.06	0.03	0.07·	-0.06	0.03	0.05·
minorityyes	-0.04	0.02	0.03*	-0.05	0.02	0.02*	-0.02	0.05	0.65	-	-	-
nativeyes	-0.06	0.03	0.03*	-0.06	0.03	0.02*	-0.08	0.06	0.20	-	-	-
tenureyes	-0.02	0.02	0.34	-	-	-	-	-	-	-	-	-
divisionlower	0.00	0.01	0.94	-	-	-	-	-	-	-	-	-
creditsingle	0.15	0.03	0.00***	0.16	0.03	0.00***	0.11	0.03	0.00***	0.11	0.03	0.00**
Random effects	-			-			0.07			0.08		
AIC	742.00			737.47			596.11			594.82		
BIC	783.38			766.44			629.21			619.65		

注: 记号 “***”、“**”、“*”、“·” 分别表示显著性水平 0.001、0.01、0.05、0.1 下显著。

4. 结论

本研究针对教学质量评价数据中存在的分层结构与异方差性特点, 提出了分层广义线性模型的分析方法。通过对两项实际数据的分析, 验证了该方法在此类复杂数据情境下的有效性与可行性。结果表明, 该模型能够更精准地识别影响教学质量的关键因素, 为教育管理者与教师采取针对性改进措施提供了实证支持。需要指出的是, 教学质量评价是一个复杂的系统工程, 未来研究应结合质性分析、课堂观察等多种方法, 构建更为综合的评价体系, 从而更全面地推动课堂教学质量的持续提升[25]。

基金项目

上海高校市级重点课程建设项目“贝叶斯统计”(2024)。

参考文献

- [1] 陈磊, 朱金秀, 丁小庆. 新时代高校教师教学质量评价的改革与实践[J]. 中国大学教学, 2024(4): 68-74.
- [2] 张丽. 基于结构方程模型的高校教师教学质量评价[J]. 统计与决策, 2014(16): 61-63.
- [3] 盛艳燕, 赵映川. 基于多面 Rasch 模型的大学教师课堂教学能力评价方法研究[J]. 高教探索, 2015(2): 70-74.
- [4] 冉燕辉, 唐万梅. 基于云-灰关联分析的教学评价研究[J]. 重庆师范大学学报(自然科学版), 2015, 32(1): 142-146.
- [5] 董国玉, 祁迎春. 层次分析法在高校课堂教学质量评估中的应用研究[J]. 中国成人教育, 2017(9): 53-56.
- [6] 范岩, 马立平. 优化 BP 神经网络的高校教学质量评价模型[J]. 统计与决策, 2018, 34(2): 80-82.
- [7] 刘忠铁, 于洪鉴. 高校学生评教的实证研究[J]. 中国大学教学, 2021(8):51-56.
- [8] 任吉, 于佳琪. 熵权法在硕士研究生方法论课程质量评价体系构建中的应用[J]. 黑龙江高教研究, 2021, 39(9): 105-110.
- [9] 周晓东, 陈沐沂. 基于尺度混合偏正态分布的考试成绩分析[J]. 统计学与应用, 2024, 13(5): 1677-1689.
- [10] Cook, S., Watson, D. and Webb, R. (2024) Performance Evaluation in Teaching: Dissecting Student Evaluations in Higher Education. *Studies in Educational Evaluation*, **81**, Article 101342. <https://doi.org/10.1016/j.stueduc.2024.101342>
- [11] Liu, R., Zhang, Y. and Fang, X. (2025) Research on the Influencing Factors of Business Teachers' Teaching Quality: PLS-SEM and ANN Approach. *The International Journal of Management Education*, **23**, Article 101134. <https://doi.org/10.1016/j.ijme.2025.101134>
- [12] Morris, C.N. (1995) Hierarchical Models for Educational Data: An Overview. *Journal of Educational and Behavioral Statistics*, **20**, 190-200. <https://doi.org/10.3102/10769986020002190>
- [13] 纪江明, 葛羽屏. 分层模型视角下中心城市基础教育满意度影响因素研究——基于“2012 新加坡连氏中国城市公共服务质量调查”的实证分析[J]. 教师教育研究, 2015, 27(2): 1-7.
- [14] 王芳. 基于分层线性模型的大学生教学满意度影响因素分析[J]. 复旦教育论坛, 2018, 16(1): 48-55+97.
- [15] 郭建鹏, 唐祯, 吕帅. 什么因素影响了中国高校教师的教学投入——基于个体与环境两水平变量的分层线性模型分析[J]. 教育发展研究, 2022, 42(9): 57-70.
- [16] Sulis, I., Porcu, M. and Capursi, V. (2019) On the Use of Student Evaluation of Teaching: A Longitudinal Analysis Combining Measurement Issues and Implications of the Exercise. *Social Indicators Research*, **142**, 1305-1331. <https://doi.org/10.1007/s11205-018-1946-8>
- [17] 石磊, 向其凤, 陈飞. 多水平模型及其在经济领域中的应用[M]. 第2版. 北京: 科学出版社, 2021.
- [18] 彭大松, 苗国. 家庭化流动背景下非户籍人口的社区参与研究——基于广义分层线性模型的分析[J]. 人口与发展, 2020, 26(5): 62-72.
- [19] 孙维伟, 张连增, 胡祥. 基于分层广义线性模型的非寿险费率厘定精算模型研究[J]. 统计与信息论坛, 2017, 32(6): 48-54.
- [20] 甘行琼, 李玉姣. 城市基本公共服务对新生代农民工定居意愿的影响——基于广义分层线性模型的实证研究[J]. 贵州社会科学, 2019(4): 126-133.
- [21] Tuerlinckx, F., Rijmen, F., Verbeke, G. and De Boeck, P. (2006) Statistical Inference in Generalized Linear Mixed Models: A Review. *British Journal of Mathematical and Statistical Psychology*, **59**, 225-255.

<https://doi.org/10.1348/000711005x79857>

- [22] Dean, C.B. and Nielsen, J.D. (2007) Generalized Linear Mixed Models: A Review and Some Extensions. *Lifetime Data Analysis*, **13**, 497-512. <https://doi.org/10.1007/s10985-007-9065-x>
- [23] Hamermesh, D.S. and Parker, A. (2005) Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity. *Economics of Education Review*, **24**, 369-376. <https://doi.org/10.1016/j.econedurev.2004.07.013>
- [24] 李毅, 向浩. 基于“三维一体”的高等教育课堂教学质量评价研究[J]. 高教学刊, 2019(17): 81-83.
- [25] Buchanan, E.M., Miranda, J.F. and Stephens, C. (2025) The Reliability of Student Evaluations of Teaching. *Assessment & Evaluation in Higher Education*, **50**, 997-1011. <https://doi.org/10.1080/02602938.2025.2504618>