

基于t检验和随机森林算法的高等数学成绩主控因素分析

第五鹏祥¹, 姚熙媛¹, 赵兰苓¹, 许香敏¹, 江礼武²

¹中国石油大学(北京)理学院, 北京

²中国石油大学(北京)克拉玛依校区, 石油学院, 新疆 克拉玛依

收稿日期: 2024年12月18日; 录用日期: 2025年2月6日; 发布日期: 2025年2月17日

摘要

围绕不同专业学生学习高等数学成绩主控因素筛选问题, 构建了两个专业236名学生的高等数学学习过程指标和结课成绩数据集。运用描述性统计分析方法刻画成绩分布特征, 并借助t检验判断专业间学习数据和成绩差异的显著性。使用随机森林模型, 对影响学生高数成绩的进行分析, 发现课上习题作答率和课后习题得分率对结课成绩的影响更加显著, 到课率对结课成绩影响最不显著。基于数据分析结果, 针对性提出了面向所学专业进行差异化教学以及提升课堂互动、课后学习质量两大类教学改进措施。

关键词

高等数学, 主控因素, t检验, 随机森林, 教学建议

Analysis of Main Control Factors of Advanced Mathematics Performance Based on t-Test and Random Forest Algorithm

Pengxiang Diwu¹, Xiyuan Yao¹, Lanling Zhao¹, Xiangmin Xu¹, Liwu Jiang²

¹College of Science, China University of Petroleum, Beijing

²School of Petroleum, Karamay Campus, China University of Petroleum-Beijing at Karamay, Karamay Xinjiang

Received: Dec. 18th, 2024; accepted: Feb. 6th, 2025; published: Feb. 17th, 2025

Abstract

Regarding the screening of the dominant factors influencing the achievements of students from

文章引用: 第五鹏祥, 姚熙媛, 赵兰苓, 许香敏, 江礼武. 基于 t 检验和随机森林算法的高等数学成绩主控因素分析[J]. 创新教育研究, 2025, 13(2): 127-132. DOI: 10.12677/ces.2025.132094

different majors in advanced mathematics, a dataset of the learning process indicators and final course scores of 236 students from two majors in advanced mathematics was constructed. The descriptive statistical analysis method was used to depict the characteristics of the score distribution, and with the help of the t-test, the significance of the differences in learning data and scores among different majors was judged. By using the Random Forest model to analyze the factors affecting students' scores in advanced mathematics, it was found that the response rate of in-class exercises and the score rate of after-class exercises had a more significant impact on the final course scores, while the attendance rate had the least significant impact on the final course scores. Based on the results of data analysis, two major types of teaching improvement measures were put forward in a targeted manner, including differentiated teaching according to the majors studied and improving the quality of classroom interaction and after-class learning.

Keywords

Advanced Mathematics, Main Control Factors, t-Test, Random Forest, Teaching Suggestions

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

在高等教育体系中,高等数学(简称高数)作为理工科及部分文科专业的核心基础课程,不仅为学生后续专业课程的学习奠定坚实的数学基础,还培养了学生的逻辑思维、抽象思维及问题解决能力。然而,长期以来,高数课程因其抽象性、逻辑性强而成为学生普遍感到困难的一门课程,结课成绩往往呈现出较大的分布差异。探究学生成绩背后的主控因素,有利于针对性提出高数教学的改进措施。

利用大数据分析技术探索学生学业表现的影响因素已成为教育研究的新趋势。在此背景下,深入研究学生高数结课成绩的影响因素,并构建有效的预测模型,不仅有助于学生提升学业水平,还能的教学改进和个性化学习提供科学依据。赵红等[1]依据某重点高校大一新生数学公共课的成绩和问卷调查的多维数据,利用统计学方法进行了指标间的相关性分析、各个指标对数学成绩的差异性分析、数学成绩的影响因素和预测路径分析,得出相应的结论。孙茜等[2]采用多元线性回归分析法得出高数成绩与各类影响因素之间的关系。连高社等[3]建立层次分析法模型,探讨了各因素对大学高等数学成绩的影响路径和路径系数,提出了改进高等数学教学成效的建议和措施。

众多研究表明,高等数学成绩受多种因素共同影响。传统教学模式下,学生可能存在学习动力不足、学习方法不当等问题。吴国荣等[4]通过对农科类本科二批学生的调查发现,学生对高等数学的学习兴趣不高,课上主动参与度低,课后自主学习习惯差,这些因素显著影响成绩。于静等[5]阐述了传统高等数学课堂存在的问题,如学生学习动力不足、课前无预习、课堂互动少、课后不复习等。代伟等[6]指出在以教师讲授为主的传统教育模式下学生参与度低,往往依照教师的讲授接受式学习,缺少独立思考和自主探究的机会,解决实际问题的能力和创新能力薄弱。

随着现代信息技术的飞速发展,现代教学理念与工具不断革新,雨课堂等智慧教学平台逐渐融入高等数学教学过程[7]。现代教学与智慧教学平台的结合为改善这一状况带来了契机。张钧君等[8]在教学过程中利用雨课堂用于推送预习内容、进行课堂互动反馈和课后学习跟踪,改变了传统工程制图教学中学生接受式学习的局面,增强了学生的学习主动性。李彩丽等[9]探讨了智慧教学平台在生物化学教学中的应用,发现其在预习引导、课堂互动和课后反馈方面具有显著优势。教师可以根据智慧教学平台提供的

数据了解学生的学习情况，为学生提供针对性的学习建议，这种方式相比传统教学模式更能激发学生的学习兴趣 and 主动性。

借助目前雨课堂等智慧教学平台的发展和应用，在教学中收集更准确的学习数据成为可能，有望得到更加准确的高数学习成绩主控因素分析结果。本文利用雨课堂收集学生学习《高等数学》课程的过程数据，从课堂表现和课后学习两个角度分析学生情况，利用 t 检验、随机森林等方法，讨论了不同专业学生学习是否存在差异、学习成绩主控因素，针对性提出了《高等数学》教学建议。

2. 数据来源与研究方法

2.1. 数据来源

在高数课程教学过程中，依托雨课堂平台对学生课上表现和课后学习进行量化评价。本研究选取两个不同专业 236 名同学的高等数学过程和卷面成绩进行分析。过程指标是此次分析的自变量，包括课上指标：到课率(X_1)、课上习题作答率(X_2)、课上作答正确率(X_3)；课后指标：课后习题作答率(X_4)、课后作答正确率(X_5)。卷面成绩是此次分析的因变量，是学生结课成绩(Y)。其中，将有过程成绩但最后缺考的学生移除。表 1 是部分学生数据的展示。

Table 1. Students' written grades and process grades (partial data)

表 1. “学生卷面成绩和过程成绩(部分)”

Y	X_1	X_2	X_3	X_4	X_5
34	0.9767	0.48	0.6968	0.93548	0.7321
29	0.9535	0.57	0.6593	0.90323	0.8193
72	0.9302	0.55	0.6976	0.87097	0.8211
65	1	0.57	0.7778	0.87097	0.8704
60	0.8298	0.3778	0.9586	0.62857	0.6285
76	0.9787	0.8333	0.9955	0.65714	0.7887
70	0.6596	0.5667	0.8509	0.65714	0.7242
59	0.9362	0.7111	0.9022	0.65714	0.7658
79	0.9574	0.8556	0.9986	0.65714	0.8632
71	0.9787	0.8333	0.9635	0.65714	0.7857

2.2. 研究方法

(1) 描述性统计分析和 t 检验

对两个班级的数据进行描述性统计分析，了解均值、标准差的基本特征。均值反映了数据的集中趋势，标准差体现了数据的离散程度。通过这些统计量，可以初步了解两个班级高数成绩的整体水平、波动情况以及分布特征。

t 检验是一种假设检验方法，用于比较两个样本的均值是否存在显著差异。在本研究中，通过 t 检验来确定两个专业的成绩是否具有统计学上的显著差异。如果差异显著，则说明两个专业的成绩不是来自于同一个主体，即学生所学专业对高数教学产生了较大影响，应考虑基于专业的差异化教学。

(2) 随机森林特征重要性分析

随机森林是一种集成学习算法，它通过构建多个决策树并将它们的结果进行组合来进行预测。在特征重要性分析中，随机森林会衡量每个特征在构建决策树时的重要性。利用随机森林算法对表 1 中五个

因素进行特征重要性分析，确定各因素对结课成绩的影响程度。

3. 结果与分析

3.1. 描述性统计分析

为了解两个班级学生在高等数学上的整体表现和成绩情况，将两份数据进行描述性统计性分析[9]，具体结果见表 2。

Table 2. Descriptive statistics on the scores of advanced mathematics
表 2. 高等数学成绩描述性统计

专业	人数	最高分	最低分	平均分	标准差
A	131	97	1	55.95	23.28
B	105	88	4	54.35	21.22

统计数据显示，专业 A 的最高分为 97，平均分为 55.95，分别高于专业 B 的最高分和平均分，两个专业的高数学习可能存在差异。进一步使用 t 检验方法，判断专业对高数学习影响的显著性。

3.2. t 检验

采用 t 检验对不同专业学生的高数学习过程指标和卷面成绩进行数据分析，判断不同专业学生的数据是否来自不同总体。t 检验结果见表 3。

Table 3. T-test for various indicators of advanced mathematics
表 3. 高等数学各项指标 t 检验

指标	t-statistic	P-value	结论
X ₁ (到课率)	11.80	1.56e-25	存在显著差异
X ₂ (课上习题作答率)	-9.23	1.62e-17	存在显著差异
X ₃ (课上习题得分率)	-11.36	4.06e-24	存在显著差异
X ₄ (课后习题作答率)	23.97	6.27e-65	存在显著差异
X ₅ (课后习题得分率)	4.24	3.25e-05	存在显著差异

t 检验结果表明，不同专业学生的高数学习过程指标和卷面成绩分别来自不同的总体，具有显著的差异性，即学生专业对高数学习有显著的影响

3.3. 特征重要性分析

Table 4. Analysis of the importance of features in the Random Forest for influencing factors.
表 4. 影响因素随机森林特征重要性分析

指标	重要性
X ₁ (到课率)	0.09580
X ₂ (课上习题作答率)	0.25100
X ₃ (课上习题得分率)	0.19146
X ₄ (课后习题作答率)	0.19786
X ₅ (课后习题得分率)	0.26389

为了进一步探究学生课上和课后表现对结课成绩的影响,采用随机森林算法筛选了课上(X_1, X_2, X_3)和课后(X_4, X_5)指标对结课成绩影响的重要性,结果见表4。

结果显示, X_2 (课上习题作答率)和 X_5 (课后习题得分率)对 Y (结课成绩)的影响更加显著, X_1 (到课率)对 Y (结课成绩)影响最不显著。这说明通过点名等手段使学生按时到课,很难实质性提高学生的高数学习效果。通过提升学生的课堂参与度和课后的作业练习质量,对提高高数学习成绩可能会有积极和有效的促进作用。

4. 结果讨论

4.1. 面向所学专业进行差异化教学的必要性

根据描述性统计和 t 检验结果可知,两个专业的学习过程数据源于不同总体且差异显著。学生分属于不同的专业,入学时学生生源水平存在差异;入学后专业课程内容、班级和宿舍的学习小环境也有区别,最终使得不同专业学生的成绩出现显著不同。

鉴于此,教师应依据学生的专业背景建立学生学习档案,记录学生的学习过程数据[10],深入了解不同专业学生的学习特点与需求,为每个学生制定个性化的学习计划和教学内容推荐,包括专属学习计划、精准教学内容推送。例如,对于理论性强的专业,设计层层递进、原理推导式的案例;对于应用型专业,引入真实项目拆解式案例,充分融合互联网资源开展沉浸式、个性化教学[11]。通过引导学生了解所学专业,激发学生的内在学习动力,并为不同水平的学生提供个性化的学习建议和指导,普遍地提高学生学习成绩,缩小群体间成绩差距。

4.2. 课堂互动和课后学习质量是影响成绩的重要因素

综合特征重要性分析的结果,可以得出影响《高等数学》结课成绩高低最主要的两个因素为课上作答率和课后习题作答率。因此教师一方面提高课堂参与度,增强教师与学生之间的互动[12],通过鼓励学生建立学习小组[13]和提供预习视频等措施降低学生参与课堂的难度,提高学习效率。

鉴于课后学习对高数成绩的重要性,教师可利用互联网在线课程、学习平台等丰富的学习资源,善于运用大数据工具,为学生提供多样化[14]的课后学习支持。比如,在在线教育平台[15]精准投放个性化课后作业,依据学生课堂表现、阶段测试结果智能匹配难度适宜、知识点关联紧密的习题,并配备多维度答案解析,涵盖文字详解、动画演示、名师微课讲解。同时,鼓励学生建立在线学习小组,合理利用即时通讯工具进行实时讨论和问题解答,找到适合自己的学习方法,提高学习效率。教师也可在线参与学生的讨论,强化在线教育平台的互动功能,及时给予指导和反馈。

5. 结论

(1) 收集了两个专业236名同学的高等数学学习过程指标和结课成绩。 t 检验结果显示,不同专业学生的高数学习过程指标和结课成绩存在显著差异,推荐面向所学专业进行差异化教学,包括设计和专业吻合的案例,进行案例式教学等方式。

(2) 通过随机森林特征重要性分析,发现课上习题作答率和课后习题得分率对结课成绩的影响更加显著、到课率对结课成绩影响最不显著。教师在高数教学中,要注重提高课堂互动和课后学习质量。

(3) 本文并未考虑学生在学习高数成绩前的数学基础,未考虑学生课后努力程度和课余生活的影响,教师的主观因素也没有进行控制变量,且样本容量有限。通过长期和大量学习过程数据的收集分析,有望得到更加准确的结果。

参考文献

- [1] 赵红,张若军,赵菲.数学成绩的影响因素分析与数学素养提升的探讨——H大学大一新生数学公共课的多维

- 数据分析[J]. 高等理科教育, 2024(5): 25-32.
- [2] 孙茜, 蔡择林. 地方本科院校高等数学成绩影响因素的回归分析[J]. 湖北师范大学学报(自然科学版), 2020, 40(3): 80-83.
 - [3] 连高社, 陈小彪. 基于层次分析法的大学生高等数学成绩影响因素分析[J]. 大学数学, 2021, 37(4): 70-78.
 - [4] 吴国荣, 刘宇菲, 杨彩琴, 等. 高等数学学习成绩影响因素的调查分析——以内蒙古农业大学农科类本科二批录取学生为例[J]. 内蒙古农业大学学报(社会科学版), 2019, 21(1): 19-23.
 - [5] 于静, 杨立敏, 王晶晶, 等. 雨课堂在高等数学混合式教学中的应用[J]. 教育教学论坛, 2020(32): 265-267.
 - [6] 代伟, 杨洋, 鄂成国, 等. “以学生为中心”高等数学教学改革与实践[J]. 河北环工程学院学报, 2023, 33(4): 90-94.
 - [7] 崔学慧, 王瑛子, 韩丽君, 等. 大学数学公共基础课学习效果的统计分析方法与实践[J]. 创新教育研究, 2023, 11(4): 809-815.
 - [8] 张钧君, 薛燕, 吕建国. 基于“雨课堂”的线上线下混合式教学模式探索与实践——以工程制图与 CAD 课程为例[J]. 河北北方学院学报(社会科学版), 2024, 40(1): 104-106.
 - [9] 李彩丽, 李永泉. 基于雨课堂的形成性评价体系在《生物化学》教学中的应用[J]. 中国继续医学教育, 2022, 14(7): 1-4.
 - [10] 江志冬. 统计分析理论在大学生学习成绩分析中的应用[J]. 辽宁师专学报(自然科学版), 2022, 24(1): 10-14.
 - [11] 潘兴侠, 郭琦茹, 林楠. 本科生高等数学成绩影响因素调查——基于 logistic 回归模型的分析[J]. 大学数学, 2021, 37(4): 60-69.
 - [12] 曹明, 李宸锋. 互联网环境对高校师生教学互动水平的影响研究[J/OL]. 煤炭高等教育, 2024: 1-10. <http://kns.cnki.net/kcms/detail/32.1365.g4.20241209.1109.001.html>, 2024-12-11.
 - [13] 郑金山. SPSS 数据处理软件在高等数学成绩分析中的应用[J]. 科技经济市场, 2023(1): 134-136.
 - [14] 孙茜, 蔡择林. 地方本科院校高等数学成绩影响因素的回归分析[J]. 湖北师范大学学报(自然科学版), 2020, 40(3): 80-83.
 - [15] 史新竹, 裴秀丛, 李舒, 等. 教育信息化背景下医学院校教师网络教学平台的应用[J]. 卫生职业教育, 2024, 42(14): 106-110.