

大学数学公共基础课学习效果的统计分析方法与实践

崔学慧, 王瑛子, 韩丽君, 郑淑贤

中国石油大学(北京)理学院, 北京

收稿日期: 2023年3月10日; 录用日期: 2023年4月14日; 发布日期: 2023年4月25日

摘要

高等学校对大学数学公共基础课学情分析有助于理清教学对象的特征、督促教师改进数学公共基础课的教学方法, 提升学生的学习效果。本文利用课程考核结果, 使用相关分析、假设检验、聚类分析等方法挖掘数学公共基础课程成绩分布的统计特征。通过得出的统计结果, 分析可能导致其分布特征的客观因素, 为提升教与学的质量提供恰当的建议。利用灰色预测方法预判学生成绩趋势, 提供教学过程的前置分析, 为教学设计完善化给予支撑, 从而进行差异化教学。该方法体系能够从宏观上进行学情分析, 分析过程直观简洁、科学有效, 可以滚动应用, 可移植性强。

关键词

数学公共基础课程, 相关分析, 假设检验, 聚类, 灰色预测

Statistical Analysis Methods and Practice of Learning Effects of Common Mathematics Courses during College

Xuehui Cui, Yingzi Wang, Lijun Han, Shuxian Zheng

College of Science, China University of Petroleum (Beijing), Beijing

Received: Mar. 10th, 2023; accepted: Apr. 14th, 2023; published: Apr. 25th, 2023

Abstract

The analysis of the academic situation of public basic mathematics courses in colleges and universities is helpful to clarify the characteristics of teaching objects, urge teachers to improve the

文章引用: 崔学慧, 王瑛子, 韩丽君, 郑淑贤. 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. DOI: 10.12677/ces.2023.114124

teaching methods of common mathematics courses, and improve students' learning effects. This paper uses the results of course assessment to explore the statistical characteristics of the score distribution of public basic mathematics courses by correlation analysis, hypothesis test, and cluster analysis. Through the statistical results obtained, the objective factors that may lead to their distribution characteristics are analyzed, and appropriate recommendations are provided for improving the quality of teaching and learning. The gray prediction method is used to predict the trend of student performance, provide pre-analysis of the teaching process, and provide support for the improvement of teaching design, so as to carry out differentiated teaching. The method systems which have strong portability can analyze the academic situation on the macroscopic level, and the analysis process is intuitive, concise, scientific and effective, and can be applied to rolling analysis.

Keywords

Common Mathematics Courses, Correlation Analysis, Hypothesis Test, Cluster, Gray Forecast

Copyright © 2023 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

随着信息化进程的深入,高校在学生成绩的信息化管理方面有着多维度的数据,拥有丰富的数据量,但仍处于数据输入、查询、简单统计和备份的早期阶段,不能及时有效地进行数据挖掘,不利于发现影响学生成绩的关键因素,如何通过数据挖掘数学公共基础课程成绩分布的统计特征,以及找出成绩与课程之间的关联性,来进行决策支持、指导教学,是目前教师必须要面对的问题[1][2]。

学生成绩是经常用来评价学生学习效果的基础指标,它可以反映学习情况,也能衡量高校办学水平。系统分析学生学习质量、分析教与学之间的相关性、统计不同院系之间成绩分布的显著性等,对于全面梳理和把握数学公共基础课程学习效果,提升教学水平,促进学风建设等有积极的作用。

本文通过统计分析方法处理大学数学公共基础课学生成绩,得到学生的学习情况,从学习效果反映教学质量,有效帮助教师指导学生学习,帮助教师做出明智的教学决策,使教学方法更加有效[3]。

2. 基于数据挖掘方法的大学数学公共基础课学习效果现状分析

2.1. 调查数据来源

本文所用数据来源于4门数学公共基础课程(高等数学、概率论统计基础、线性代数、数理统计)2017~2022年间学生的考试成绩。

该数据库包含以下信息:

- (1)基本信息:脱敏后的学生信息;
- (2)课程属性:学分、学时、课程名称、学期;
- (3)成绩信息:课堂期末成绩、课堂平时成绩、总评成绩。

2.2. 学生成绩特征描述

2.2.1. 描述性统计

利用SPSS软件对上述数据进行描述性统计分析[4]。

观察表 1 各科平均值均在 70~80 分之间,表明各门课程的试题难度适宜,其中数理统计平均值最高,为 78 分;线性代数平均值最低,为 71 分。各科标准差都大于 10,结合极差和标准差可知,各课程学生成绩分布均相对离散,其中高等数学标准差为 15.5 分,极差为 94 分。综合来看,各科考试均有一定难度。

Table 1. Descriptive statistics of course grades

表 1. 课程成绩描述性统计

		数理统计	概率统计基础	高等数学	线性代数
个案数	有效	3522	4007	7695	7287
	缺失	6418	5933	2245	2653
平均值		78.1542	77.2088	74.1746	71.3121
平均值标准误差		0.24971	0.20806	0.17691	0.17679
中位数		80.0000	79.0000	76.0000	72.0000
标准差		14.81943	13.17054	15.51849	15.09180
偏度		-0.693	-0.664	-0.555	-0.496
偏度标准误差		0.041	0.039	0.028	0.029
峰度		0.261	0.403	-0.070	0.029
峰度标准误差		0.082	0.077	0.056	0.057
极差		90.00	85.00	94.00	92.00
最小值		10.00	15.00	6.00	8.00
最大值		100.00	100.00	100.00	100.00
百分位数	25	68.0000	68.0000	64.0000	62.0000
	50	80.0000	79.0000	76.0000	72.0000
	75	90.0000	87.0000	86.0000	83.0000

2.2.2. 学生成绩分布主要特征

为了解学生成绩分布情况,利用 SPSS 软件对整理好的成绩进行正态性检验,判断学生成绩是否满足正态分布[5]。

Table 2. Normality test for each subject

表 2. 各科正态性检验

	Kolmogorov-Smirnov 检验 ^a			Shapiro-Wilk 检验		
	统计值	自由度值	显著性值	统计值	自由度值	显著性值
数理统计	0.070	831	0.000	0.961	831	0.000
概率统计基础	0.046	970	0.000	0.978	970	0.000
高等数学	0.076	1794	0.000	0.961	1794	0.000
线性代数	0.064	1741	0.000	0.971	1741	0.000

由表 2 中正态检验显著性值即 P 值可知 4 门公共数学基础课程学生成绩不严格服从正态分布,后续

分析应采用对正态分布要求不高的统计分析方法，提升可靠性。

根据上节描述性统计中的偏度和峰度数值可知，数理统计和概率统计基础课程学生成绩基础是左偏分布低峰态；高等数学、线性代数是左偏分布尖峰态。数理统计和概率统计是低峰态，说明这两门课程成绩离散。

3. 基于非参数统计方法的学生成绩统计分析

3.1. 课程成绩的 Spearman 相关分析

选取以上数据的期末成绩，因为不满足正态分布，故对其进行相关性分析采用 Spearman 方法[6]，得出相关系数，并做显著性检验，结果如下。

Table 3. Course correlation analysis for 2017~2019 class

表 3. 2017~2019 级课程相关性分析

年级		数理统计	概率统计	高等数学	线性代数	
2017级	数理统计	相关系数	1.000	.	0.626**	0.682**
		显著性(双尾)	.	.	0.000	0.000
		个案数	831	0	804	733
	概率统计	相关系数	.	1.000	0.652**	0.698**
		显著性(双尾)	.	.	0.000	0.000
		个案数	0	970	935	956
	高等数学	相关系数	0.626**	0.652**	1.000	0.645**
		显著性(双尾)	0.000	0.000	.	0.000
		个案数	804	935	1794	1685
	线性代数	相关系数	0.682**	0.698**	0.645**	1.000
		显著性(双尾)	0.000	0.000	0.000	.
		个案数	733	956	1685	1741
2018级	数理统计	相关系数	1.000	.	0.429**	0.461**
		显著性(双尾)	.	.	0.000	0.000
		个案数	830	0	782	702
	概率统计	相关系数	.	1.000	0.398**	0.361**
		显著性(双尾)	.	.	0.000	0.000
		个案数	0	971	936	956
	高等数学	相关系数	0.429**	0.398**	1.000	0.648**
		显著性(双尾)	0.000	0.000	.	0.000
		个案数	782	936	1765	1647
	线性代数	相关系数	0.461**	0.361**	0.648**	1.000
		显著性(双尾)	0.000	0.000	0.000	.
		个案数	702	956	1647	1732

Continued

2019级	数理统计	相关系数	1.000	.	0.662**	0.697**
		显著性(双尾)	.	.	0.000	0.000
		个案数	1037	0	975	894
	概率统计	相关系数	.	1.000	0.658**	0.707**
		显著性(双尾)	.	.	0.000	0.000
		个案数	0	1087	1057	1073
	高等数学	相关系数	0.662**	0.658**	1.000	0.643**
		显著性(双尾)	0.000	0.000	.	0.000
		个案数	975	1057	2094	1961
		相关系数	0.697**	0.707**	0.643**	1.000
		显著性(双尾)	0.000	0.000	0.000	.
		个案数	894	1073	1961	2062
线性代数	相关系数	0.662**	0.658**	1.000	0.643**	
	显著性(双尾)	0.000	0.000	0.000	.	
	个案数	894	1073	1961	2062	

各门课程的考试成绩有显著的正相关关系, 由此可知, 4 门课程之间关联性较强, 学生若想打好数学基础, 就应该注重每一科目的学习, 不能忽视任一数学基础课程。同时表明, 大一数学基础课的学习质量对后续学习有较大影响, 要引起学生自身及管理部門的重视。

对比表 3, 2017 级和 2018 级的课程关联性, 发现 2018 级高等数学与其它 3 门课程的关联性较低, 可重点关注 18 级高等数学的教学情况和课程设置情况, 并在之后的教学中及时改进; 高等数学和线性代数的相关性三年中非常稳定, 相关系数保持在 0.65 左右, 这两门课程在数学基础课阶段为互补课程。

3.2. Kruskal-Wallis 检验

Kruskal-Wallis 检验是用于推断计量资料或等级资料的多个独立样本所来自的多个总体分布是否有差别的方法。

选择 2019 年 4 门课程的成绩数据, 使用假设检验来检验不同专业之间对同一课程的接受程度是否存在差异。原假设被定义为各专业以相同的方式学习相同的课程, 根据数据特征选择 Kruskal-Wallis 检验[7], 分析结果如下。

Table 4. Hypothesis test of student score in different majors

表 4. 不同专业学生成绩假设检验

课程	原假设	检验	显著性	决策
数理统计			0.000	拒绝原假设
概率统计基础	在专业类别中, 总成绩分布相同	Kruskal-Wallis 检验	0.000	拒绝原假设
高等数学			0.000	拒绝原假设
线性代数			0.000	拒绝原假设

由表 4 得知不同专业对各课程的掌握情况都不相同。故针对假设检验的结果, 接下来选取聚类分析的方法, 对各专业对课程的掌握情况进行分类并排序, 对学生的差别情况以及专业的排名进行更细致的分析。

选取 2016~2020 级 4 门数学公共基础课成绩数据, 检验不同年级学生对同一课程的学习是否存在差

异。原假设被定义为同一课程的学习在不同年级是相同的，进行分析。

Table 5. Hypothesis test of student score in different grades

表 5. 不同年级学生成绩假设检验

课程	原假设	检验	显著性	决策
数理统计			0.000	拒绝原假设
概率统计基础	在年级类别中， 总成绩分布相同	Kruskal-Wallis 检验	0.000	拒绝原假设
高等数学			0.000	拒绝原假设
线性代数			0.000	拒绝原假设

由表 5 得出 2017 和 2020 级，2018 和 2020 级的 4 门数学公共基础课程的成绩分布基本相同，不存在差异。2017 与 2019 级的课程成绩和 2018 与 2019 级的课程成绩差异最大。综上所述，可以认为不同年级对同一数学公共基础课程的掌握情况不同。

3.3. K-means 聚类分析

因为高等数学与线性代数课程的典型相关关系，选取 2017~2019 级全体学生高等数学和线性代数成绩数据，用 K-means 聚类将不同专业对课程的掌握情况进行分类，每个班级的分数为聚类平均数，这些平均数是确定学生成绩和分配高低分的标准和依据，利用 SPSS 对数据进行聚类分析[8] [9]。

Table 6. ANOVA table

表 6. ANOVA 表

	聚类		误差		F	显著性
	均方	自由度	均方	自由度		
高等数学	405.144	3	9.281	32	43.652	0.000
线性代数	226.426	3	7.788	32	29.075	0.000

观察 ANOVA 方差分析表 6，F 值是各科对聚类的贡献，重要程度为高等数学 > 线性代数，学校可适当增加重要课程的课时，并选取有经验的教师承担教学，提高教学效果。

4. 基于灰色预测的学生成绩预测

Table 7. Predicted grades for next year's grades for two courses

表 7. 两门课程来年成绩预测成绩

数理统计	2016 级	2017 级	2018 级	2019 级	2020 级
实际值	69.10	79.70	84.82	77.97	
预测值		81.675	80.827	79.987	79.156
绝对误差			2.3819%		
概率统计基础	2016 级	2017 级	2018 级	2019 级	2020 级
实际值	73.49	76.89	82.93	75.29	
预测值		79.148	78.367	77.594	76.828
绝对误差			2.7837%		

灰色预测是一种对含有不确定因素的系统进行预测的方法。灰色预测通过关联分析和灰色生成来寻找系统变动的规律,然后建立微分方程模型来预测未来发展趋势。灰色预测模型适用于样本少、数据不完整或不可靠的情况,运算方便、精度高,适用于对学生成绩进行预测。

因为对于绝大多数高校而言概率类的课程最后开设,所以利用 Matlab 语言对数理统计及概率统计基础课程成绩数据进行预测,结果如下。

由表 7 可知,两门课程下一年的成绩均有下降的趋势,考虑到受疫情影响,教师在之后教学中需加强疫情期间线上学习的效率。

5. 结论

本文提出的评价过程复杂度不高,既有描述性统计,也有基于假设检验的科学分析过程,能够系统综合评价学生的数学公共基础教学的学习效果。提供教学过程的前置分析。结合分析结果,采取丰富多样的教学手段,重点监控,提升学习能力与学习效果。

(1) 根据课程的 Spearman 相关性分析结果发现,各课程学生成绩相关度高,要使学生打好数学基础,就应该注重每一科目的学习,不能厚此薄彼。

(2) 根据对不同专业学生成绩的 K-W 检验结果可以得到,部分课程及部分专业的学生成绩差别较大,教师需要重点关注这些课程及专业的学习情况。根据对不同年级学生成绩 K-W 检验结果可以得到,不同年级对同一课程的掌握情况不同,教师应针对不同年级学生的实际情况进行差异化教学设计,以学生为中心,综合把握学情,多措并举提升教学质量,夯实本科生的数学基本功,提升学生的数学基础理论水平。

(3) 由 K-Means 聚类分析结果得出重要课程贡献度较大,学校可适当增加重要课程的课时以利于提高教学效果。教师应鼓励学生,使他们不仅把学习当成任务,更要把学习当成兴趣,注重巩固数学基础。教师应该帮助中等偏上的同学提高线性代数基础,从而全面发展,课后注重培养空间想象力。授课过程中重点关注学习成绩相对落后的学生,激发学生求知欲,结合平时测验,发现不足并督促学生及时改正。

参考文献

- [1] 吴文玲. 基于数据挖掘技术的课程相关性分析及其应用研究[D]: [硕士学位论文]. 成都: 四川师范大学, 2018.
- [2] 陈喜华. 数据挖掘在学生成绩分析中的应用研究[D]: [硕士学位论文]. 广州: 华南农业大学, 2017.
- [3] 张甜, 尹长川, 潘林, 等. 基于改进的聚类和关联规则挖掘的学生成绩分析[J]. 北京邮电大学学报(社会科学版), 2018, 20(2): 91-96.
- [4] 张丁丁. 描述性统计分析及 SPSS 实现[J]. 协和医学杂志, 2018, 9(5): 447.
- [5] 李廉, 张万红. 高校课程考试成绩正态分布辨析[J]. 中国考试, 2021, 348(4): 86-93.
- [6] 钱晨坚. Spearman 秩相关系数比较的假设检验及样本量估计新方法研究[D]: [硕士学位论文]. 广州: 南方医科大学, 2021.
- [7] 蒲虎. Kruskal-Wallis 检验原理介绍及其应用[J]. 兴义民族师范学院学报, 2019, 122(4): 108-111.
- [8] 王世纯, 许新华, 黄嘉成, 等. K-means 聚类算法在高校学生成绩分析中的应用研究[J]. 湖北师范大学学报(自然科学版), 2019, 39(3): 113-118.
- [9] 钟文精, 焦中明, 蔡乐. 基于 K-Means 算法的学生成绩聚类分析[J]. 教育信息技术, 2021(5): 56-58.