

DOI:10.13718/j.cnki.xsxb.2021.11.011

在线学习数据与学生学习成绩的相关性分析

——以大学概率论课程为例^①

熊思灿¹, 农莹²

1. 东华理工大学 理学院, 南昌 330013; 2. 华中师范大学 教育学院, 武汉 430079

摘要: 选取概率论课程 195 名学习者不同时段的在线学习数据和期末成绩数据进行研究, 并运用相关分析法、回归分析法以及因子分析法综合考察在线学习数据与期末成绩之间的相关关系。研究表明, 不同时段的在线学习数据均与期末考试成绩呈非常显著的正相关关系。在线学习数据可被分为反映“知识输入”“知识输出”和“互动交流”的 3 个类别。这 3 个类别的典型代表变量分别是在线学习访问数、在线考试成绩和讨论数。通过这 3 个典型变量建立的回归模型能较好地反映期末成绩的变化情况, 并进行学业预警。

关键词: 在线学习数据; 相关性分析; 回归分析; 因子分析; 学业预警

中图分类号: G40-057

文献标志码: A

文章编号: 1000-5471(2021)11-0084-06

On Correlation Analysis Between Online Learning Data and Students' Academic Performance

——With the University Probability Theory Course as an Example

XIONG Sican¹, NONG Ying²

1. School of Science, East China University of Technology, Nanchang 330013, China;

2. School of Education, Central China Normal University, Wuhan 430079, China

Abstract: Based on the online learning scores at different periods and final grades of 195 learners from the Probability Theory, the correlation between the online learning scores and final grades was comprehensively analyzed through correlation analysis, regression analysis, and factor analysis. The results show that the correlation between the online learning scores at different periods and the final grades are significantly positive. The online learning scores could be divided into three categories reflecting “knowledge input”, “knowledge output”, and “interactive communication”. The typical representative variables of these three categories are the visits, online test scores, and the number of discussions. The regression models established by these three typical variables can better reflect the changes in final grades and provide academic warning.

Key words: online learning scores; correlation analysis; regression analysis; factor analysis; academic warning

① 收稿日期: 2020-07-21

基金项目: 东华理工大学博士启动基金研究项目(DHBK2018052); 江西省基础教育研究课题(SZUDHZH2019-1133).

作者简介: 熊思灿, 博士, 副教授, 主要从事教育统计学方向研究.

为响应教育部 2018 年 4 月下发的《教育信息化 2.0 行动计划》的号召,我们从 2018 年秋季开始,对“概率论”课程进行了移动教学改革.将线下教学资源整合形成线上学习资源,借助于超星网络课程平台进行课程建设和管理,通过超星学习通 APP 开展线下为主、线上为辅的移动教学.通过两轮的教学实践,该课程平台积累了同学们大量的在线学习数据.如何对这些数据进行分析并挖掘其中蕴含的教育规律,成为了摆在我们面前的重要任务.

在线学习数据属于教育数据的一种,而教育数据又是大数据的一个子集,是一座可无限开采的“金矿”^[1].教育数据的思维和处理能力也是教师数据素养和核心能力之一^[2].国内已有不少学者基于教育数据开展了相关研究^[3].在在线学习行为对期末成绩的解释率方面,文献[4]从访问动机、资源类型、行为投入特征等三个方面进行了探讨,获得了 12 个反映在线学习的行为变量,这些变量对期末成绩的累计解释率为 65.7%.文献[5]则以文学批评课程为例,并从学习风格偏好的角度探讨了在线学习行为对学生期末成绩的间接影响情况.文献[6]也以“雨课堂”的在线学习数据为协变量,进行了期末成绩的单向协方差分析,结果表明“雨课堂”对教学和学习十分有帮助.文献[7]也以远程教育和网络教育应用课程为例,分析了学生在线学习数据,如学习满意度等,与期末成绩的相关关系.

纵观已有研究,尽管存在着在线学习数据与期末成绩的相关性研究.但是在线学习数据往往只有一个学期的数据,且样本量相对较少,统计效力不足.此外,现有研究往往针对的是学期末的在线学习数据的总结性分析,较少有研究对不同时段的在线学习数据进行过程性分析.本研究将解答以下 3 个问题:①在线学习数据是如何影响期末成绩的?②在线学习数据可分为哪些大类?③利用过程性在线学习数据如何进行学业预警?

1 研究设计

1.1 研究对象和变量选取

本研究选取概率论课程的在线学习数据为例进行研究.概率论课程是我校数学与应用数学专业和统计学专业的专业基础课,在大学二年级上学期开设.课程总学时为 48 学时,每周安排 4 学时,共 12 周约 100 天左右.本研究数据采集自两个学期,其中:2018—2019 学年第 1 学期共 2 个教学班,分别由 2017 级数学与应用数学专业 42 人,以及 2017 级统计学专业 73 人组成;2019—2020 学年第 1 学期 1 个教学班,由 2018 级统计学专业 80 人组成.样本总容量为 195 人.

概率论课程依托超星网络课程平台运行,内容涵盖五章 27 小节的核心内容和拓展阅读以及章节测验等补充内容.教师通过超星学习通 APP 的各种控件,组织学生进行线下为主、线上为辅的移动教学.同时,课程平台记录下同学们的全程在线学习数据,包括课程积分、视频观看个数、视频观看时长、讨论得分、作业得分、在线考试得分、访问次数、签到次数、课程互动得分和任务点完成个数等共 10 项在线学习成绩.由于本研究横跨 2 个学期,涉及 3 个教学班,为保证数据的一致性和可比性,我们对前述 10 项在线学习成绩做归一化处理.处理方法是每位学习者的前述 10 项在线学习成绩除以该学习者所处学期和所在教学班的最高成绩,得到每位学习者的各项在线学习成绩相对于该学习者所处学期和所在教学班的最高成绩的满分率.这 10 项满分率被选作本研究的自变量.

1.2 数据采集和方法概述

数据采集分过程性数据和期末数据两部分.过程性数据采集仅在 2019—2020 学年第 1 学期教学过程中进行,采集时间一般是在距课程开课后,每隔 15 天左右进行.期末数据是在所有教学活动(含成绩报送等环节)结束后约 1 个星期内进行.过程性数据采集主要是前述 10 项在线学习成绩,而期末数据采集除了前述 10 项在线学习数据外,还采集了每位学习者的期末考试成绩.其中,期末考试成绩被选作本研究的因变量.

本研究中,在线学习成绩与期末成绩的相关性分析主要采用皮尔逊相关分析,不同等级(即期末考试成绩分段,见后文详述)同学的在线学习成绩之间的差异性比较采用方差分析法,在线学习成绩对期末成

绩的影响主要采用多元线性回归中的逐步回归法,而在线学习成绩分类则通过因子分析法得出.所有分析均在 SPSS 25.0 中完成.

2 数据分析

2.1 相关性分析及差异性检验

在线学习成绩与期末成绩的皮尔逊相关系数及检验 P 值如表 1 所示.从表 1 可以看出,各项在线学习成绩与期末成绩均存在非常显著的正相关关系.

表 1 在线学习成绩与期末成绩的皮尔逊相关系数及检验 P 值

在线成绩类别	相关系数	P 值
考试满分率	0.58	$<0.001^{***}$
讨论满分率	0.50	$<0.001^{***}$
访问数满分率	0.50	$<0.001^{***}$
课程互动满分率	0.50	$<0.001^{***}$
签到满分率	0.47	$<0.001^{***}$
作业满分率	0.47	$<0.001^{***}$
视频观看个数比	0.47	$<0.001^{***}$
任务点完成率	0.47	$<0.001^{***}$
课程积分比	0.45	$<0.001^{***}$
视频观看时长比	0.29	$<0.001^{***}$

注: $*** P \leq 0.001$.

为了进一步考察期末成绩处于不同分数段的同学在线学习成绩之间的差异,我们将期末成绩按“85 分及其以上”,“75~84 分”,“60~74 分”和“低于 60 分”进行分组,并分别对应着“优秀”“良好”“及格”和“不及格”4 个等级.然后分别计算不同等级同学的在线学习成绩均值,并对均值差异采用方差分析法进行显著性检验.同时,计算不同等级同学各项在线学习成绩的 Eta 相关系数,结果如表 2 所示.

表 2 不同等级同学的在线学习成绩均值及均值比较的方差分析结果和 Eta 相关系数

在线成绩类别	不同等级同学的在线学习成绩均值				均值比较的方差分析		Eta 相关系数
	优秀	良好	及格	不及格	F 值	P 值	
考试满分率	0.716	0.657	0.450	0.262	31.906	$<0.001^{***}$	0.578
讨论满分率	0.890	0.876	0.816	0.615	21.422	$<0.001^{***}$	0.502
访问数满分率	0.601	0.503	0.364	0.286	21.265	$<0.001^{***}$	0.5
课程互动满分率	0.898	0.858	0.614	0.462	20.196	$<0.001^{***}$	0.491
签到满分率	0.772	0.802	0.439	0.316	17.067	$<0.001^{***}$	0.46
作业满分率	0.861	0.874	0.628	0.542	15.950	$<0.001^{***}$	0.448
视频观看个数比	0.942	0.906	0.787	0.689	15.904	$<0.001^{***}$	0.447
任务点完成率	0.9859	0.9875	0.9647	0.9456	15.389	$<0.001^{***}$	0.441
课程积分比	0.833	0.772	0.680	0.618	14.019	$<0.001^{***}$	0.425
视频观看时长比	0.522	0.508	0.311	0.290	5.425	0.001^{***}	0.28

注: $*** P \leq 0.001$.

从表 2 可以看出,不同等级同学的在线学习成绩之间存在非常显著的差异性.同时, Eta 相关系数表明,除了视频观看时长的 Eta 值 0.28 略低以外,其余各项在线学习成绩与期末成绩等级之间的 Eta 相关系数均比较高.其中,考试满分率与期末成绩等级之间的 Eta 值达到了 0.578.

2.2 回归分析及因子分析

我们以期末成绩为因变量,所有 10 项在线学习成绩为自变量,进行多元线性回归分析,并采用逐步回归法进行变量筛选.所得最终回归模型的调整 R^2 为 0.405,回归效果非常显著,参数估计结果如表 3 所示.

表 3 期末成绩对在线学习成绩的逐步回归最终模型及参数估计

自变量	未标准化回归系数	标准误	t 值	P 值	共线性容差	方差膨胀因子
(截距)	42.131	2.517	16.739	<0.001***		
考试满分率	19.026	4.727	4.025	<0.001***	0.513	1.949
访问数满分率	20.800	5.057	4.113	<0.001***	0.744	1.344
讨论满分率	10.887	3.814	2.854	0.005**	0.587	1.704

注:*** $P \leq 0.001$; ** $P \leq 0.01$.

从表 3 可以看出,最终被纳入回归模型的在线学习成绩有考试满分率、访问数满分率和讨论满分率,分别用 a, b, c 表示,则所得期末成绩 S 的回归方程如公式(1)所示:

$$S = 42.131 + a \times 19.026 + b \times 20.8 + c \times 10.887 \quad (1)$$

此外,从表 3 的共线性检验结果,即容差 (Tolerance) > 0.2 或者方差膨胀因子 (VIF) < 5 ,可以看出,被纳入模型的 3 个变量之间不存在多重共线性关系. 但 10 项在线学习成绩的 KMO 值为 $0.842 > 0.8$,且巴特利特球形检验的显著性概率 < 0.001 ,表明 10 项在线学习成绩之间存在相关性,支持做因子分析. 根据因子分析结果,提取两个主成分能解释总方差的 71.534%,其因子旋转后的空间成分图如图 1 所示.

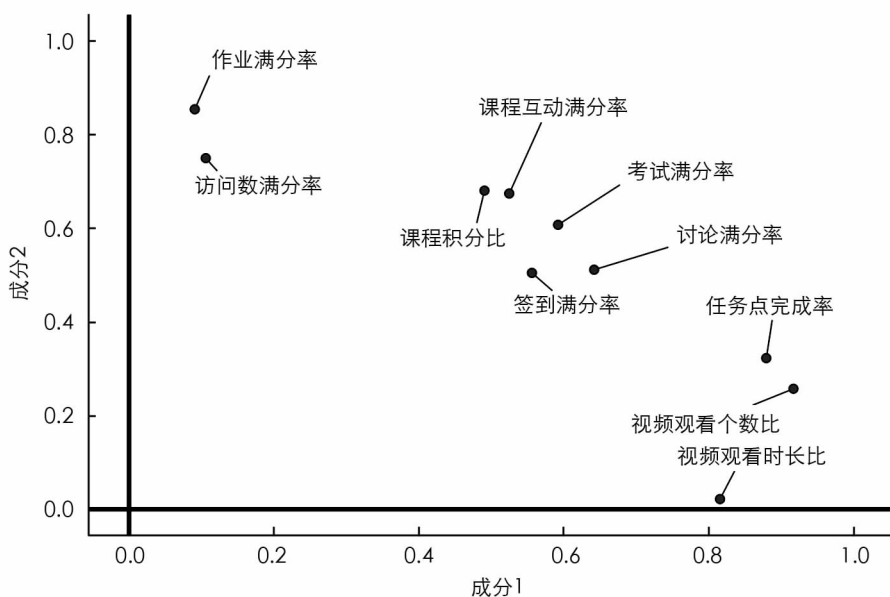


图 1 旋转后空间中的成分图

从图 1 可知,主成分 1 在视频观看时长比和视频观看个数比以及任务点完成率上面取得较大的值,可以理解为“知识输入”,重点反映知识进脑的学习程度. 主成分 2 在作业满分率和访问满分率上面取得较大的值,可以理解为“知识输出”,重点反映知识转化能力的强弱. 课程互动满分率、课程积分比、签到满分率、考试满分率和讨论满分率与所提取的两个主成分都有关系. 这些变量可以理解为既有输入又有输出的“互动交流”部分. 因此,全部 10 项在线学习成绩被归为 3 大类别,分别是“知识输入”“知识输出”和“互动交流”. 再结合回归分析结果可知,这 3 大类别在线学习数据的显著代表变量分别是访问数满分率、在线考试满分率和讨论满分率.

2.3 过程性在线学习成绩分析与学业预警

2019—2020 学年第 1 学期学习过程中,我们采集了距首次开课时间(2019 年 9 月 11 日)不同时长(15 d, 30 d, 48 d, 56 d, 65 d, 76 d, 88 d 和 99 d)共 8 期的在线学习成绩,并将这些在线学习成绩按照课程视频得分占 15%,讨论得分占 10%,作业得分占 20%,考试得分占 15%,访问数得分占 10%,签到得分占 15%以及课程互动得分占 15%,计算每位学习者的过程性综合成绩. 然后,将这 8 期的在线学习成绩代入回归方程(1)中,计算各期的期末成绩预测值.

经检验,除首期(即距首次开课 15 d 时,首期因还未进行在线考试,从而数据有缺项)外,其余各期的

综合成绩和预测成绩均与期末考试成绩呈显著正相关关系, 将每期的综合成绩和预测成绩分别与期末考试成绩求皮尔逊相关系数, 然后绘制相关系数与距首次开课天数的折线图(图 2).

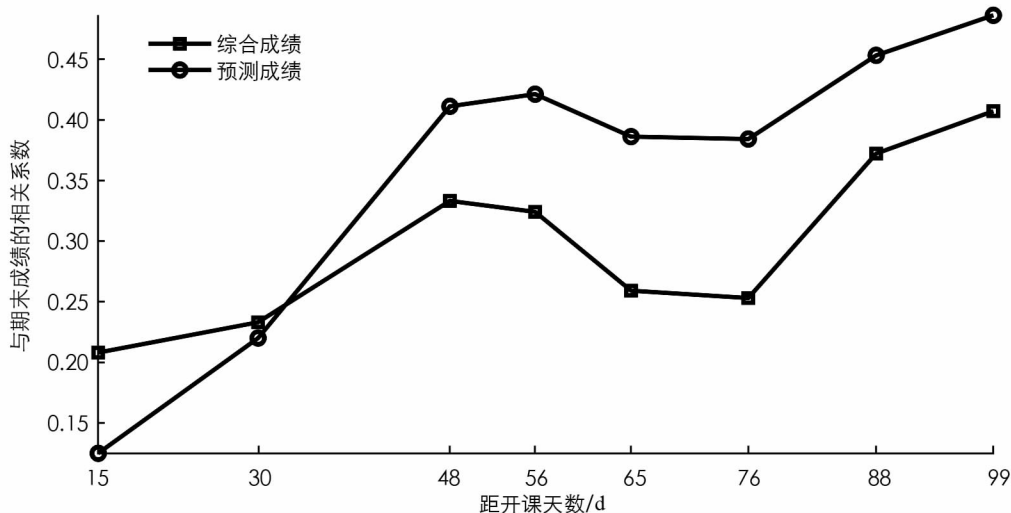


图 2 各期综合成绩和预测成绩分别与期末成绩的皮尔逊相关系数折线图

从图 2 可知, 随着学习过程的逐步推进, 综合成绩和预测成绩与期末成绩的相关系数呈逐渐上升趋势. 在中期, 出现了小幅下滑现象, 且持续约 28 d 左右之后再次回升, 并保持高相关性至学期末. 此种下滑现象揭示了在线学习的某种懈怠行为, 其具体原因还有待进一步研究. 一般地, 预测成绩与期末成绩的相关关系显著高于同期综合成绩与期末成绩的相关关系, 反映出本文所得回归方程 (1) 在日常教学中的稳健性, 以及利用综合成绩作为同学们平时学习质量的评价指标的合理性.

鉴于预测成绩与期末成绩的强相关关系, 本文以最后一期的预测成绩建立其与期末成绩的回归模型. 模型拟合 R^2 为 0.237, 显著性检验 P 值 < 0.001 , 拟合效果十分显著. 参数估计及回归系数检验结果如表 4 所示.

表 4 期末成绩对最后一期预测成绩的回归模型及参数估计

自变量	未标准化回归系数	标准误	t 值	P 值
(截距)	16.483	12.855	1.282	0.204
最后一期预测成绩	0.857	0.174	4.917	$< 0.001^{***}$

注: $*** P \leq 0.001$.

由表 4 可得, 期末成绩 S 对最后一期预测成绩 d 的回归方程为

$$S = 16.483 + 0.857 \times d \quad (2)$$

根据公式 (2), 我们可以进行学业预警. 比如, 为使期末成绩不低于 60 分, 则最后一期预测成绩至少不低于 50.778 分. 对于那些最后一期预测成绩低于 50.778 分的同学, 我们可以提前干预, 督促这部分同学加紧学习, 以防考试挂科, 实现在线学习成绩的学业预警功能.

3 结论与建议

我们以大学概率论课程为例, 对在线学习数据如何影响学生学习成绩进行了深入研究, 得到如下结论, 并据此提出了相关建议.

1) 在线学习成绩与期末成绩呈显著正相关.

在线学习成绩与期末成绩之间普遍存在显著的正相关关系. 其中, 相关系数大于 0.5 的有在线考试满分率、讨论满分率、访问数满分率和课程互动满分率. 这几个指标均能较好地反映学生的在线学习积极性和在线学习水平. 此外, 随着学生期末成绩等级的下滑, 几乎所有在线学习成绩项均呈下滑趋势, 仅有视频观看个数比、任务点完成率和签到满分率几个指标在期末成绩为“良好”等级时达最高值点. 这表明, 期末考试处于“良好”等级的同学, 平时会十分关注动动手指就能完成的, 比如签到这样的任务. 但是, 处于

“良好”等级的同学在其他项目,比如作业满分率、在线考试满分率等方面,却明显低于期末成绩处于“优秀”等级同学的相应成绩。即“优秀”的同学之所以优秀,除了做好诸如“签到”等易于完成的任务外,平时还花了很多精力在衡量知识转化的“作业”“考试”等方面。此外,我们还可以得到,期末考试“不及格”的同学在平时各项成绩排名中都靠后。如何让这部分同学也积极提高学习成绩,是值得我们深思的又一个课题。

2) 在线学习成绩可归并为反映“知识输入”“知识输出”和“互动交流”的 3 个类别。

因子分析表明,同学们的在线学习成绩可归并为反映“知识输入”“知识输出”和“互动交流”的 3 个类别。其中,反映“知识输入”的有视频观看个数比、视频观看时长比、任务点完成率等;反映“知识输出”的有作业满分率、访问满分率等;而反映“互动交流”的则有课程互动满分率、课程积分比、讨论满分率等。这 3 个类别的显著代表变量分别是访问满分率、在线考试满分率和讨论满分率。因此,为使学习者获得较好的学习成绩,教师应重点关注课程访问热度、课程讨论频率和在线考试成绩这 3 项指标,学习者也应尽力提升自己在这 3 项指标上的成绩。

3) 阶段性在线学习数据与期末成绩呈显著正相关关系且具有学业预警功能。

课程学习进行过程中的阶段性在线学习数据与期末成绩呈显著正相关关系,且相关系数随着学习时长的推进呈上升趋势。选取最后一期在线学习数据,代入回归方程(1)得期末成绩的预测值。对于预测值低于 50.778 分的学习者,教师可以提前进行干预,以防学生出现不及格现象。从而实现阶段性在线学习数据的学业预警功能。

教师作为信息化教学的组织者、主导者和具体实施者,应主动关注、收集、整理和分析相关教育数据,让经验以数据为支撑,让教学通过规律来指导,实现国家提出的发展和提高本科教育质量的宏伟目标。

参考文献:

- [1] 杨现民,唐斯斯,李冀红. 发展教育大数据:内涵、价值和挑战 [J]. 现代远程教育研究,2016(1): 50-61.
- [2] 李 新,杨现民. 教育数据思维的内涵、构成与培养路径 [J]. 现代远程教育研究,2019,31(6): 61-67.
- [3] 尹忠根,杨宗友,李采丰. 基于“互联网+”的大学体育课内外健康数据一体化实践研究 [J]. 西南师范大学学报(自然科学版),2020,45(4): 108-114.
- [4] 张媛媛,李 爽. MOOC 课程资源访问模式与学习绩效的关系研究 [J]. 中国远程教育,2019(6): 22-32, 93.
- [5] 李艳红,徐 敏. 学习风格偏好影响移动学习态度和行为的实证研究——以“文学批评”课程为例 [J]. 电化教育研究,2019,40(2): 45-53.
- [6] 姚 洁,王伟力. 微信雨课堂混合学习模式应用于高校教学的实证研究 [J]. 高教探索,2017(9): 50-54.
- [7] 胡 勇. 在线学习过程中的社会临场感与不同网络学习效果之间的关系初探 [J]. 电化教育研究,2013,34(2): 47-51.

责任编辑 张 构