

Analysis of Influencing Factors on College Students' Mathematics Scores Based on Multiple Ordered Logistic Regression

Xuejiao Liu, Xiusen Li*, Yue Sun

School of Mathematics and Statistics, Shandong University of Technology, Zibo Shandong
Email: *leexiusen@163.com

Received: Jun. 29th, 2019; accepted: Jul. 11th, 2019; published: Jul. 18th, 2019

Abstract

By using the method of multiple ordered logistic regression analysis for modeling, through the usual learning situation, college students' performance rating has been predicted and analyzed. The results show that plan for postgraduate entrance examination, careful listening in class and mathematics result of college entrance examination will have a greater impact on college students' math scores tending to be excellent. Consciously doing homework and taking notes have no significant effect on the final math scores of college students, while properly previewing before class helps in the scores. At the same time, the model prediction accuracy can reach 51.69%.

Keywords

Undergraduate Mathematics, Usual Learning Situation, Multiple Ordered Logistic Regression

基于多项有序逻辑回归的大学生数学成绩影响因素分析

刘雪娇, 李秀森*, 孙悦

山东理工大学数学与统计学院, 山东 淄博
Email: *leexiusen@163.com

收稿日期: 2019年6月29日; 录用日期: 2019年7月11日; 发布日期: 2019年7月18日

摘要

运用多项有序逻辑回归分析的方法, 建立模型, 通过平时学习情况对大学生成绩等级进行分析预测, 结论*通讯作者。

果表明,有考研计划、听课认真情况以及高考数学成绩会对大学生数学成绩趋于优秀有较大的影响,自觉做作业和整理笔记对大学生期末数学成绩没有较大的促进作用,课前适当程度的预习对大学生数学成绩有一定的帮助,同时模型预测准确度达51.69%。

关键词

大学生数学,平时学习情况,多项有序逻辑回归

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

肖婕[1]等应用方差法分析了生源地、任课教师等因素对大学数学、英语成绩的影响,宁静[4]等应用聚类法分析了高考总分对大学数学、英语成绩的影响,王亮红、彭慧芹[2]应用聚类分析法探讨了聚类分析技术在教学管理中的应用,白春玲[3]等应用广义线性回归模型研究了学生学年间成绩的相关性。本文研究大学生数学成绩与高考成绩、考研意向、课堂表现等诸多影响因素之间的相关性,与以往研究相比,范围较为集中,主要考虑学业方面的因素,以山东理工大学财管、市销和金融三个专业的学生为调查对象,设计调查问卷,回收有效问卷 239 份。

2. 逻辑回归分析的假设检验

使用有序 Logistic 进行回归分析时,需考虑四个假设:1) 因变量唯一,且为有序多分类变量;2) 存在一个或多个自变量,可为连续、有序多分类或无序分类变量;3) 自变量之间无多重共线性;4) 模型满足“比例优势”假设。

直接将分类变量纳入 Logistic 回归方程,则软件会将分类变量按连续变量处理。例如,按“文科”——1,“理科”——2 进行编码,纳入方程,方程会认为“理科”是“文科”的两倍,所以,为了解决这个问题,在对假设进行验证之前,需要用一系列的二分类变量“是”或“否”来表示原始的分类变量,这些新的二分类变量称为“哑变量”。下面我们对假设进行判断,此处重点分析假设 3~4。

见表 1,容忍度均远大于 0.1,方差膨胀因子均小于 10,所以不存在多重共线性。

Table 1. Coefficients

表 1. 系数^a

模型	共线性统计量		
	容差	VIF	
1			
	高考数学	0.701	1.426
	课前预习情况	0.674	1.484
	听课认真程度	0.306	3.267
	课堂内容当时理解程度	0.314	3.185
	课堂纪律遵守情况	0.635	1.576
	回答问题跟上思路	0.308	3.245
	玩手机	0.509	1.965

Continued

1	复习课堂内容情况	0.437	2.288
	作业认真情况	0.300	3.335
	作业独立情况	0.261	3.835
	作业自我批改情况	0.260	3.842
	wenli	0.765	1.307
	k	0.720	1.389
	zijue	0.670	1.492
	baocun	0.578	1.729
	cuoti	0.724	1.381
	biji	0.768	1.303
	moni	0.689	1.452
wangjie	0.695	1.439	

a. 因变量: 等级。

见表 2, 平行线检验中卡方 = 46.678, $P = 1$, 大于 0.05, 说明平行性假设成立, 即各回归方程相互平行, 可以使用有序 Logistic 过程进行分析。

Table 2. Test of parallel lines

表 2. 平行线检验^c

模型	-2 对数似然值	卡方	df	显著性
零假设	165.986			
广义	119.309 ^a	46.678 ^b	118	1.000

零假设规定位置参数(斜率系数)在各响应类别中都是相同的。

a. 在达到最大步骤对分次数后, 无法进一步增加对数似然值。

b. 卡方统计量的计算基于广义模型最后一次迭代得到的对数似然值。检验的有效性是不确定的。

c. 联接函数: Logit。

3. 多项有序逻辑回归分析

对模型中所有自变量的偏回归系数是否全为 0 进行似然比检验, 判断模型拟合程度。

原假设为所有纳入自变量的系数为 0, 我们所得出的模型拟合信息表(表 3)中显著性小于 0.05, 说明至少有一个变量系数不为 0, 且具有统计学显著性即模型整体有意义。换言之, 拟合自变量的模型拟合优度好于仅包含常数项的模型。

Table 3. Model fitting information

表 3. 模型拟合信息

模型	-2 对数似然值	卡方	df	显著性
仅截距	281.571			
最终	165.986	115.584	59	0.000

联接函数: Logit。

根据 SPSS 使用的模型以及参数估计值表格, 可以列出方程:

$$\ln\left(\frac{\gamma_1}{1-\gamma_1}\right) = 5.126 - 0.079 \text{高考成绩} - (0.54 \text{文理科} - 2.589 \text{考研} + \dots - 0.34 \text{作业批改情况})$$

$$\ln\left(\frac{\gamma_2}{1-\gamma_2}\right) = 8.536 - 0.079 \text{高考成绩} - (0.54 \text{文理科} - 2.589 \text{考研} + \dots - 0.34 \text{作业批改情况})$$

$$\ln\left(\frac{\gamma_3}{1-\gamma_3}\right) = 12.254 - 0.079 \text{高考成绩} - (0.54 \text{文理科} - 2.589 \text{考研} + \dots - 0.34 \text{作业批改情况})$$

上述三个方程参数只在阈值上有差别。在控制了高考成绩的条件下计算出 $\gamma_1, \gamma_2, \gamma_3$ 。可以由此得出取得优秀、良好、及格、不及格的可能性，从而我们可以对大学生的等级数学成绩进行预测。例如，由第一位学生各项数据我们计算出 $\gamma_1 = 0.09, \gamma_2 = 0.35, \gamma_3 = 0.46, \gamma_4 = 0.1$ ，可见，该学生数学成绩等级为 5 (不及格) 的可能性为 9%，数学成绩等级为 6 (及格) 的可能性为 35%，数学成绩等级为 7 (良好) 的可能性为 46%，数学成绩等级为 8 (优秀) 的可能性为 10%，综合分析，该生的成绩等级预测结果为 7 (良好)。利用该逻辑回归模型对大学生数学成绩等级进行预测，图 1 为对大学生数学成绩等级的预测结果部分图。

等级	EST1_1	EST2_1	EST3_1	EST4_1	PRE_1
7	0.09	0.35	0.46	0.10	7
7	0.02	0.13	0.52	0.33	7
8	0.01	0.05	0.34	0.61	8
6	0.11	0.39	0.42	0.08	7

Figure 1. Partial diagram of the prediction results of college students' mathematics grades
图 1. 大学生数学成绩等级的预测结果部分图

EST1_1、EST2_1、EST3_1、EST4_1 代表四个成绩等级的预测概率。其中第一行中 EST3_1 概率最大，为 0.46，对应的 PRE_1 为 7 (良好)，因变量成绩等级的观察值也为 7 (良好)，此时模型准确地预测了因变量。而第四行中 EST3_1 最大，预测的 PRE_1 为 7 (良好)，而因变量的实际的观察值为 6 (及格)，此时模型没有准确地预测因变量。

为明确该模型的预测效果，建立等级*预测响应类别交叉制表。表 4 显示，成绩等级为 5 (不及格) 时，

Table 4. Rank*PRE_1 Predicted response category cross tabulation
表 4. 等级*预测响应类别交叉制表

		预测响应类别				合计
		5	6	7	8	
5	计数	1	2	3	0	6
	等级中的%	16.7%	33.3%	50.0%	0.0%	100.0%
6	计数	0	3	16	3	22
	等级中的%	0.0%	13.6%	72.7%	13.6%	100.0%
7	计数	1	4	34	12	51
	等级中的%	2.0%	7.8%	66.7%	23.5%	100.0%
8	计数	0	0	16	23	39
	等级中的%	0.0%	0.0%	41.0%	59.0%	100.0%
合计	计数	2	9	69	38	118
	等级中的%	1.7%	7.6%	58.5%	32.2%	100.0%

模型预测正确的有 1 人(16.7%)，同理，成绩等级为 6 (及格)、7 (良好)、8 (优秀)时，模型预测正确的分别有 3 人(13.6%)、34 人(66.7%)、23 人(59%)。根据预测响应类别交叉制表可计算出模型预测的总准确度为 $(1 + 3 + 34 + 23) \div 118 = 51.69\%$ 。

4. 结论

综上所述，是否有考研计划、听课认真情况及高考数学成绩对大学生数学成绩是否趋于优秀有较大的影响，自觉做作业和整理笔记对大学生期末数学成绩没有较大的促进作用，课前适当程度的预习对大学生数学成绩有一定的帮助，且适当程度的预习比极为细致的预习有更好的效果，预习过于细致可能会在课堂上稍加松懈，对老师课堂上的传授不会过多在意，该行为会导致学习效果不佳。

参考文献

- [1] 肖婕, 宁静, 缪柏其, 等. 影响学生成绩各因素的评价[J]. 中国高等教育评估, 2001(1): 40-43.
- [2] 王亮红, 宋代清, 徐娜. 聚类分析在学生成绩分析中的应用[J]. 东北电力大学学报, 2009, 29(4): 35-38.
- [3] 白春玲, 樊顺厚, 刘军利, 等. 学生学习成绩相关性的研究[J]. 数学的实践与认识, 2009, 39(23): 48-55.
- [4] 宁静, 肖婕, 缪柏其, 等. 高考成绩与大学成绩的相关性研究[J]. 高等理科教育, 2001(3): 46-50.

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2169-2556, 即可查询。
2. 通过知网首页 <http://cnki.net/> 顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: ass@hanspub.org