

基于自我效能感的数学成绩的预测模型

王志祥^{1*}, 赵佩佩², 张加明³, 周友士⁴

¹淮阴师范学院数学与统计学院, 江苏 淮安

²新淮高级中学, 江苏 淮安

³江苏省淮海中学, 江苏 淮安

⁴淮阴师范学院教育科学学院, 江苏 淮安

Email: *wzqiang@hytc.edu.cn

收稿日期: 2021年7月21日; 录用日期: 2021年8月13日; 发布日期: 2021年8月20日

摘要

本文以淮安市淮海中学的实际调查数据为样本进行实证研究。基于数学学习自我效能感, 本文给出了数学成绩的两个预测模型: 线性回归模型和Logistic回归模型。结果表明, 利用加权最小二乘法得到的线性回归模型具有较高的拟合优度; 而Logistic回归模型对“优”、“良”、“差”有较高的预测正确率, 对“中”的预测正确率最差。

关键词

自我效能感, 线性回归, Logistic回归, 实证研究

Prediction Models of Academic Performance in Math Based on the Self-Efficacy

Zhixiang Wang^{1*}, Peipei Zhao², Jiaming Zhang³, Youshi Zhou⁴

¹School of Mathematics and Statistics, Huaiyin Normal University, Huai'an Jiangsu

²Xinhuai Senior High School, Huai'an Jiangsu

³Jiangsu Huaihai Senior High School, Huai'an Jiangsu

⁴School of Education Science, Huaiyin Normal University, Huai'an Jiangsu

Email: *wzqiang@hytc.edu.cn

Received: Jul. 21st, 2021; accepted: Aug. 13th, 2021; published: Aug. 20th, 2021

Abstract

In this paper, an empirical study is conducted by taking samples from survey in Jiangsu Huaihai

*通讯作者。

文章引用: 王志祥, 赵佩佩, 张加明, 周友士. 基于自我效能感的数学成绩的预测模型[J]. 统计学与应用, 2021, 10(4): 706-713. DOI: 10.12677/sa.2021.104072

senior high school. Two prediction models of academic performance in math are obtained based on the self-efficacy, which are linear regression model and Logistic regression model. The results suggest that the linear regression model obtained by weighted least square method has high goodness of fit, and the Logistic regression model has a high prediction accuracy for “excellent”, “good” and “poor”, but has the lowest prediction accuracy for “medium”.

Keywords

Self-Efficacy, Linear Regression, Logistic Regression, Empirical Study

Copyright © 2021 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

学生的学业成绩是一直教育工作者、心理学家研究的热点问题。许多学者认为学业成绩主要受到智力的和非智力的因素的影响[1]，其中主要的非智力因素包括自我效能感、归因、成就目标定向等[2][3][4][5]。

班杜拉认为自我效能感是人类动因的中心机制，是人们行动的重要基础[6]。作为一种重要的非智力因素，自我效能感通过影响个体的认知过程、动机过程、情感过程 and 选择过程而调节人类活动[7]。Schunk 的实验结果表明：无论学习任务处于何种难度水平或对待被试的方式如何，自我效能是一个对学业成绩的良好预测指标[8]。

高中学生的数学学业成绩无疑与学生的自我效能感有关。在前期的研究中，我们参照边玉芳博士的《学习自我效能感量表》[9]、倪颖丽的《数学学习自我效能感调查问卷》[10]，结合高中数学学科特点，编制了《数学学习自我效能感量表》进行了实证研究。本文我们将在此基础上，就基于数学学习的自我效能感对数学学业成绩的预测开展研究，给出合理的学业成绩的预测模型。为方便计，下文中所说的自我效能感指的是数学学习的自我效能感。

2. 数据来源

以淮安市淮海中学高一年级的学生为调查对象，共获得了 108 名同学的自我效能感与数学成绩数据，其中女生占比为 47.2%，男生占比为 52.8%。部份观测数据参见文末附录。处理数据所用的软件是 SPSS19.0 [11]。

调查问卷采用四点李克特量表，其有 35 个变量，以平均得分来衡量学生的数学学习的自我效能感水平，并以平均得分参与建模，变量名为“自我效能感得分”。参与建模的数学成绩分两种级制，其一是百分制，总分为 150 分，变量名为“数学成绩”；其二是等级制，共分优、良、中、差四级，变量名为“学业等级”。

3. 基本统计分析

表 1 给出了数学成绩与自我效能感得分的基本统计量，其中数学成绩的最高分为 146 分，最低分为 47 分，均值为 97.53 分，标准差为 21.393；而自我效能感得分的最低分为 1.63，最高分为 3.43，均值为 2.496，标准差为 0.33059。

Table 1. Descriptive statistics of academic performance and self-efficacy in math

表 1. 数学成绩与自我效能感得分的基本统计量

	N	Minimum	Maximum	Mean	Std. Deviation
数学成绩	108	47	146	97.53	21.939
自我效能感得分	108	1.63	3.43	2.4960	0.33059

为了研究数学成绩与自我效能感得分是否存在线性关系，我们计算了二者之间的相关系数，对是否线性相关进行了假设检验，结果如表 2。假设检验的 p -值几近为 0，若取显著性水平为 0.05，则拒绝原假设，认为数学成绩与自我效能感是相关的。另一方面 Pearson 相关系数为 0.795，表明二者之间呈正相关且具有较高的线性相关性，因此可以尝试进行线性回归分析。为方便计，在接下来的所有假设检验问题中，所取的显著性水平皆为 0.05。

Table 2. The correlation between academic performance and self-efficacy in math

表 2. 数学成绩与自我效能感得分的相关系数

		数学成绩	自我效能感得分
数学成绩	Pearson Correlation	1	0.795
	Sig. (2-tailed)		0.000
	N	108	108

关于性别及家庭教育方式是否对数学成绩产生显著影响，我们进行了双因素方差分析，结果如表 3 所示，其中家庭教育方式分为“民主”、“放任”、“严厉”与“专制”四种类别。这里我们仍然取显著性水平为 0.05，不难发现无论是“性别”、“家庭教育方式”还是这两者交互都不会对学生的数学成绩产生显著影响，因为它们所对应的 p -值分别为 0.720、0.843 和 0.640 都大于显著性水平。因此，在进行线性回归分析以及 Logistic 回归时都不考虑这两个因素的影响。

Table 3. Two-way ANOVA of the gender and family education on the math academic performance

表 3. 性别与家庭教育方式对数学成绩影响的双因素方差分析

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1678.157 ^a	7	239.737	0.481	0.846
Intercept	491496.607	1	491496.607	986.530	0.000
性别	64.314	1	64.314	0.129	0.720
家庭教育方式	411.746	3	137.249	0.275	0.843
性别 * 家庭教育方式	842.657	3	280.886	0.564	0.640
Error	49820.760	100	498.208		
Total	1078759.000	108			
Corrected Total	51498.917	107			

4. 基于自我效能感的数学成绩的线性回归模型

这一节我们采用加权最小二乘法[12]，以数学成绩作为被解释变量，以自我效能感作为解释变量建立线性回归模型，其中权重为残差的平方的倒数。

统计计算的结果依次呈现在表 4-8 中。表 4 表明采用加权最小二乘法可以获得很高的拟合优度: 0.987, 表明利用加权最小二乘法得到的回归直线很好地拟合了数据点。

Table 4. Goodness of fit

表 4. 拟合优度

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.993	0.987	0.987	0.998

表 5 给出了利用方差分析的方法对回归方程的显著性进行检验的结果。可以看出 F 检验的 p -值几近为 0 (<显著性水平 0.05), 表明被解释变量与所有的解释变量间存在线性关系。

Table 5. Significance test of regression equation

表 5. 回归方程的显著性检验

Model	Sum of Squares	df	Mean Square	F	Sig.
1					
Regression	7840.997	1	7840.997	7877.736	0.000
Residual	105.506	106	0.995		
Total	7946.502	107			

表 6 给出加权线性回归方程的系数以及对回归系数是否为 0 进行了检验, 检验的 p -值也几近为 0 (<0.05), 表明回归方程中变量“自我效能感得分”的系数与 0 有显著差异, 也即“自我效能感得分”与被解释变量“数学成绩”间有显著的线性关系。由表 6 可得回归方程为

$$\text{数学成绩} = -32.011 + 51.851 \times \text{自我效能感得分} \quad (1)$$

Table 6. Significance test of regression coefficients

表 6. 回归系数的显著性检验

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1					
(Constant)	-32.011	1.437		-22.270	0.000
自我效能感得分	51.851	0.584	0.993	88.757	0.000

接下来我们检验模型的合理性。表 7 给出了残差的正态性检验, 结果表明 K-S 检验的 p -值为 0.280, 在显著性水平为 0.05 的条件下接受原假设, 认为残差呈正态分布。而表 8 中的 t 检验表明, 残差的均值在显著性水平为 0.05 的条件下与 0 无显著差异。这里的两个检验表明模型的残差是满足线性回归分析的先决条件的。

Table 7. K-S test of residuals

表 7. 残差的 K-S 检验

		Unstandardized Residual
N		108
Normal Parameters	Mean	0.1158793
	Std. Deviation	13.31413036
Kolmogorov-Smirnov Z		0.991
Asymp. Sig. (2-tailed)		0.280

Table 8. One sample t-test of residuals

表 8. 残差的单样本 t 检验

Test Value = 0				
Unstandardized Residual	t	df	Sig.(2-tailed)	Mean Difference
	0.090	107	0.928	0.11587925

进一步地，我们对残差进行了异方差分析，通过计算得残差的绝对值与解释变量的 Spearman 等级相关系数为-0.082，而检验的 p -值为 0.402，这两个数据表明，在显著性水平为 0.05 的条件下，残差序列不存在异方差情形(见表 9)。以上结论表明，这里所建立的线性回归模型是一个合理的预测模型。

Table 9. Spearman correlation coefficient

表 9. Spearman 等级相关系数

		自我效能感得分	残差
Spearman's rho	自我效能感得分	Correlation Coefficient	1.000
		Sig. (2-tailed)	0.0
		N	108
			-0.082
			0.402
			108

5. 基于自我效能感的数学成绩的 Logistic 回归模型

我们将数学成绩按等级计分，127~150 分记为“优”，105~126 分记为“良”，90~104 分记为“中”，90 分以下记为“差”，从而将成绩分成四个等级，在模型中的变量名为“学业等级”，且规定变量“学业等级”取值 1、2、3、4 分别代表“优”、“良”、“中”、“差”。本节我们以“自我效能感得分”作为解释变量，以“学业等级”作为被解释变量，将“差”级作为参考类别，建立多项 Logistic 回归模型：

$$\begin{aligned} \text{Logit } P_i &= \ln(P(\text{学业等级} = i) / P(\text{学业等级} = 4)) \\ &= \beta_0^i + \beta_i \times \text{自我效能感得分}, i = 1, 2, 3 \end{aligned} \tag{2}$$

表 10 给出了观测数据中学业等级的分布情况，其中“优”、“良”、“中”、“差”占比分别为 8.3%、26.9%、28.7%和 36.1%。“优”级最少，“差”级最多。

Table 10. Case processing summary

表 10. 个案处理摘要

		N	Marginal Percentage
学业等级	优	9	8.3%
	良	29	26.9%
	中	31	28.7%
	差	39	36.1%
Valid		108	100.0%

三个测度拟合优度的指标在表 11 中给出，其中 Cox and Snell R^2 、Nagelkerke R^2 和 McFadden R^2 分别为 0.535、0.580 和 0.298，显然都不是很高，但是处于可接受水平。从表 12 中可以发现拟合度的 Pearson 卡方检验的 p -值为 0.999，大于显著性水平 0.05，表明模型对原始数据的拟合通过检验，即被解释变量的实际类别值的与预测类别值的分布无显著差异。

Table 11. Pseudo R-square
表 11. 伪 R 方

Cox and Snell	0.535
Nagelkerke	0.580
McFadden	0.298

Table 12. Goodness of fit
表 12. 拟合度检验

	Chi-Square	df	Sig.
Pearson	81.169	123	0.999
Deviance	78.037	123	0.999

表 13 表明在回归方程中引入变量“自我效能感得分”后，其卡方检验的 p -值近似为 0，在显著性水平是 0.05 的条件下，变量“自我效能感得分”对 Logit P 的贡献是显著的。

Table 13. Likelihood ratio tests
表 13. 似然比检验

Effect	Model Fitting Criteria		Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.	
Intercept	208.995	89.391	3	0.000	
自我效能感得分	202.413	82.809	3	0.000	

回归模型的参数估计及 Wald 检验的结果见表 14，或以发现在三个模型中，变量“自我效能感得分”所对应的 p -值均小于显著性水平 0.05，因此，在三个模型中变量“自我效能感得分”的回归系数与 0 均有显著性差异，也即变量“自我效能感得分”与 Logit $P_i (i = 1, 2, 3)$ 有显著的线性关系。三个模型的回归方程依次为：

$$\begin{aligned} \text{Logit } P_1 &= \ln(P(\text{学业等级} = 1)/P(\text{学业等级} = 4)) \\ &= -50.191 + 18.058 \times \text{自我效能感得分} \end{aligned} \quad (3)$$

$$\begin{aligned} \text{Logit } P_2 &= \ln(P(\text{学业等级} = 2)/P(\text{学业等级} = 4)) \\ &= -21.082 + 8.420 \times \text{自我效能感得分} \end{aligned} \quad (4)$$

$$\begin{aligned} \text{Logit } P_3 &= \ln(P(\text{学业等级} = 3)/P(\text{学业等级} = 4)) \\ &= -8.993 + 3.715 \times \text{自我效能感得分} \end{aligned} \quad (5)$$

Table 14. Parameter estimates
表 14. 参数估计

学业等级		B	Std. Error	Wald	df	Sig.	Exp(B)
优	Intercept	-50.191	10.631	22.291	1	0.000	
	自我效能感得分	18.058	3.704	23.769	1	0.000	69,575,151.227
良	Intercept	-21.082	4.220	24.957	1	0.000	
	自我效能感得分	8.420	1.696	24.658	1	0.000	4537.867

Continued

中	Intercept	-8.993	3.072	8.570	1	0.003	
	自我效能感得分	3.715	1.292	8.267	1	0.004	41.065

表 15 给出了利用上述模型进行预测的各级别的正确率, 结果表明, 对于学业等级为“优”、“良”、“中”、“差”的预测正确率分别为 77.8%、62.1%、22.6%及 71.8%, 总正确率为 55.6%。尽管总正确率不是太理想, 但是对于“优”、“良”、“差”预测正确率较高, 因此该模型有一定的实用价值。

Table 15. Classification

表 15. 分类表

Observed	Predicted				Percent Correct
	优	良	中	差	
优	7	2	0	0	77.8%
良	2	18	3	6	62.1%
中	0	9	7	15	22.6%
差	0	4	7	28	71.8%
Overall Percentage	8.3%	30.6%	15.7%	45.4%	55.6%

6. 结论

1) 学生的性别、家庭教育方式以及它们的交互对学生的数学成绩没有显著影响。

2) 学生的数学成绩与数学学习的自我效能感呈正相关, 用线性回归模型对数学成绩进行预测, 且具有较高的拟合优度。

3) Logistic 回归模型对以等级制给出的数学学业等级进行预测时, 并不能对所有的类别都得到较高的预测准确率, 在本文的研究中, Logistic 回归模型对于“优”、“良”、“差”预测正确率较高, 但是对于“中”的预测正确率不高。

基金项目

江苏省教育科学“十三五”规划 2017 年度课题(D/2016/01/98)。

参考文献

- [1] 蔡廷栋, 高光珍. 大学生自我效能感, 归因与学习成绩的结构方程模型[J]. 牡丹江师范学院学报: 哲学社会科学版, 2011(5): 121-123.
- [2] 王凯荣, 辛涛, 李琼. 中学生自我效能感、归因与学习成绩关系的研究[J]. 心理发展与教育, 1999(4): 22-25.
- [3] 胡桂英, 许百华. 初中生学习归因、学习自我效能感、学习策略和学业成就关系的研究[J]. 心理科学, 2002(25): 757-758.
- [4] 王雁飞, 凌文轻, 朱瑜. 成就目标定向、自我效能与反馈寻求行为的关系[J]. 心理科学, 2004(1): 204-207.
- [5] 王志祥, 柏传志, 赵佩佩. 高中生数学学习的自我效能感与成就目标定向的实证研究—以淮安市为例[J]. 统计学与应用, 2019, 8(6): 887-894.
- [6] Bandura, A. (1977) Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, **84**, 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>
- [7] 边玉芳. 学习的自我效能[M]. 杭州: 浙江教育出版社, 2004: 1-2.
- [8] Schunk, D.H. (1989) Self-Efficacy and Achievement Behaviors. *Education Psychology Review*, **12**, 173-208.

- [9] 边玉芳. 学习自我效能感量表的编制与应用[D]: [博士学位论文]. 上海: 华东师范大学, 2003.
- [10] 倪颖丽. 高中生数学学习自我效能感的相关研究[D]: [博士学位论文]. 长春: 东北师范大学, 2012.
- [11] 薛薇. 基于 SPSS 的数据分析[M]. 北京: 中国人民大学出版社, 2015.
- [12] Jeff Gill(著), 王彦蓉(译). 广义线性模型: 一种统一的方法[M]. 上海: 格致出版社、上海人民出版社, 2015.

附录

Table A1. A part of the observation data of the self-efficacy in mathematics learning and academic performance in math
表 A1. 数学学习的自我效能感与数学学业成绩的部分观测数据

年级	班级	数学成绩	学业等级	考试号	性别	教育方式	v1	v2	v3	...	v33	v34	v35
高一	8	86	差	000010801	女	民主	1	2	3	...	2	2	3
高一	8	112	良	000010802	男	民主	1	4	2	...	4	4	4
高一	8	101	良	000010803	男	民主	1	3	3	...	3	3	3
高一	8	108	良	000010804	女	民主	1	3	3	...	4	3	4
高一	8	124	良	000010805	男	民主	1	2	4	...	2	3	4
高一	8	94	中	000010806	男	严厉	3	3	4	...	1	2	2
高一	8	91	中	000010807	女	专制	4	2	2	...	2	3	2
高一	8	68	差	000010808	女	民主	1	3	1	...	1	4	4
高一	8	103	良	000010809	女	民主	1	1	2	...	4	4	3
高一	8	140	优	000010810	男	民主	1	2	3	...	1	4	3
高一	8	110	良	000010811	女	严厉	3	3	3	...	3	2	3