

大学生学习情况调查统计分析

——以平顶山学院为例

刘萍汝, 谢强*, 王晗笑

平顶山学院数学与统计学院, 河南 平顶山

收稿日期: 2023年7月11日; 录用日期: 2023年8月1日; 发布日期: 2023年8月15日

摘要

本文通过设计调查问卷, 统计学校同学对所研究因素的评分, 整理问卷数据, 建立以研究因素为自变量, 学生平均绩点为因变量的多元线性回归模型, 在建立该回归模型的过程中, 发现原始变量间存在共线性, 因此采用主成分分析法对研究因素降维, 提取主成分, 将学生成绩与提取的四个主成分做主成分回归分析。再将四个主成分与研究因素做线性回归, 代入整理得到学生成绩与研究因素的最终回归方程。通过比较回归方程系数绝对值大小, 可以知道在学习成绩和原始24个因素的回归表达式中, 课内认真程度, 课外学习时间, 知识基础, 阅读浏览量, 社会实践经历, 考前准备在所有因素中系数绝对值较大, 即学习行为因子对我校学生学习程度的影响较大。

关键词

多元线性回归, 主成分分析, 主成分回归

Statistical Analysis of the Survey of College Students' Study Situation

—Taking Pingdingshan University as an Example

Pingru Liu, Qiang Xie*, Hanxiao Wang

School of Mathematics and Statistics, Pingdingshan University, Pingdingshan Henan

Received: Jul. 11th, 2023; accepted: Aug. 1st, 2023; published: Aug. 15th, 2023

Abstract

This paper designed a questionnaire, collected scores of students on the factors studied, sorted

*通讯作者。

out the questionnaire data, and established a multiple linear regression model with the research factors as independent variables and students' average score-point as dependent variables. In the process of establishing the regression model, it was found that there was collinearity among the original variables. Therefore, principal component analysis method was adopted to reduce the dimension of the research factors, extract the principal components, and perform principal component regression analysis between student achievement and the four principal components extracted. Then, the four principal components and research factors are used for linear regression, and the final regression equation of student achievement and research factors is obtained. By comparing the absolute value of the coefficient of the regression equation, it can be seen that in the regression expression of the academic performance and the original 24 factors, the absolute value of the coefficient of all factors is larger, namely, the learning behavior factor has a greater influence on the learning level of students in our school.

Keywords

Multiple Linear Regression, Principal Component Analysis, Principal Component Regression

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

世界上具有代表性的大学生调查研究项目有很多,例如英国的“全国大学生调查”,澳大利亚的“课程体验调查”与“大学就读经验调查”,及美国的“全国大学生学习性投入调查[1]”。

近年以来,社会越来越期待能够培养出更多高等教育人才,因此国内学术界也开始关注大学生的学习和发展,并且已取得瞩目成果:清华大学——“中国大学生学习与发展追踪研究”,北京师范大学——“中国大学生就读经验调查”,北京大学——“首都高校质量检测项目”,除此外湖南大学、西安交通大学等参加了美国伯克利大学主持的“国际研究型大学学生就读经验调查”。诸多高校也曾使用调查问卷对本校生进行调查,例如中山大学[2]。

总的来说,国内大学生学习情况调查项目既有改进学校教育质量的作用,又有充实和推进高等教育学科研究领域的功能,非常符合中国教育界发展和研究需要,应当坚持并且不断完善[3]。

2. 学生学习状况调查

2.1. 设计调查问卷

本次数据来源于自己制作的调查问卷,并用调查问卷收集得到的抽样数据估计学院学生的总体状况,即对平顶山学院学生做抽样调查,并且为简单随机抽样调查。

问卷共回收了 431 份,将原始问卷中 A, B, C, D, E 选项替换成-5, -3, 0, 3, 5 等数字来表示所研究因素的状态和水平,导出后删除不合理数据并对剩余问卷数据进行统计。最后得到实际有效数据 414 份。将这 414 份问卷录入 SPSS 软件中,以便对数据进行处理。

本次问卷共设了 25 道题,包括可能会影响学生学习成绩的 24 个因素:对任课老师的印象(X_1),课内认真程度(X_2),好胜心(X_3),校园学习环境(X_4),对科目没有兴趣(X_5),经济水平(X_6),自信心(X_7),没有学习方法(X_8),人际关系(X_9),愉快的情绪(X_{10}),课外学习时间(X_{11}),自制力(X_{12}),求知欲(X_{13}),知识

基础(X_{14}), 没有好的学习习惯(X_{15}), 家庭氛围(X_{16}), 阅读浏览量(X_{17}), 社会实践经历(X_{18}), 幸福感(X_{19}), 知识衔接差(X_{20}), 兴趣爱好(X_{21}), 家庭学习环境(X_{22}), 学习目标不明确(X_{23}), 考前准备(X_{24}), 和他们上学期的平均绩点, 因为若要研究学生的学习情况, 学习成绩即平均绩点是最直观简洁的参考方式。

2.2. 数据的信度检验

Cronbach- α 信度系数是最普遍被使用的关于问卷信度的估计方法, 通常根据 Cronbach- α 系数判定调查问卷的信息是否可靠, 通常其判定标准在 0~1 之间, 系数越大, 问卷信息越可靠。

将 Cronbach- α 信度系数范围列表为表 1:

Table 1. Cronbach- α Coefficients table

表 1. Cronbach- α 系数表

Cronbach- α 系数	结论
>0.9	信度极高
>0.8	信度非常好
>0.7	可以接受
>0.6	应该进行修订, 但仍有价值
<0.6	需要重新设置题项

对数据进行信度检验结果:

Table 2. Case processing summary

表 2. 案例处理汇总

	N	%
有效	414	100.0
案例 已排除 a	0	.0
总计	414	100.0

Table 3. Reliability statistics

表 3. 可靠性统计量

Cronbach's Alpha	项数
0.902	24

由案例处理汇总表(表 2)可知, 收集到的 414 个回答均有效, 并且由可靠性统计表(表 3)可以看出 Cronbach- α 系数的值为 0.902, 则证明所收集到的数据信度很好, 真实、可靠、有效, 可以用于本篇文章的统计分析基础数据。

2.3. 多重共线性检验

VIF 值是检验回归方程多重共线性的常规方法, 通常认为 VIF 值以 10 为界限(严格定义为 5), 在此

基础上 VIF 值小于 10, 且其越趋近 1, 多重共线性的程度接近无; 若 VIF 的值大于 10, 则认为其不适合做回归分析。

Table 4. Coefficient
表 4. 系数

模型	非标准化系数		标准系数	t	Sig.	共线性统计量	
	B	标准误差	试用版			容差	VIF
(常量)	-0.221	0.000		.	.		
X ₁	0.051	0.000	0.159	.	.	0.040	24.959
X ₂	-0.003	0.000	-0.008	.	.	0.066	15.234
X ₃	0.055	0.000	0.166	.	.	0.023	43.147
X ₄	-0.005	0.000	-0.017	.	.	0.324	3.084
X ₅	-0.005	0.000	-0.016	.	.	0.306	3.265
X ₆	-0.003	0.000	-0.011	.	.	0.234	4.268
X ₇	-2.187E-16	0.000	0.000	.	.	0.591	1.692
X ₈	-0.004	0.000	-0.013	.	.	0.249	4.010
X ₉	0.000	0.000	-0.003	.	.	0.249	4.022
X ₁₀	0.055	0.000	0.167	.	.	0.024	41.750
X ₁₁	-0.003	0.000	-0.010	.	.	0.063	15.749
1 X ₁₂	0.050	0.000	0.154	.	.	0.226	4.423
X ₁₃	0.053	0.000	0.160	.	.	0.143	7.003
X ₁₄	-0.007	0.000	-0.020	.	.	0.247	4.046
X ₁₅	-0.010	0.000	-0.031	.	.	0.298	3.358
X ₁₆	-0.001	0.000	-0.004	.	.	0.254	3.943
X ₁₇	-0.004	0.000	-0.012	.	.	0.248	4.026
X ₁₈	-0.006	0.000	-0.018	.	.	0.266	3.758
X ₁₉	0.051	0.000	0.157	.	.	0.046	21.939
X ₂₀	-0.009	0.000	-0.026	.	.	0.187	5.341
X ₂₁	0.051	0.000	0.157	.	.	0.218	4.582
X ₂₂	-0.003	0.000	-0.008	.	.	0.200	4.995
X ₂₃	-0.008	0.000	-0.023	.	.	0.236	4.239
X ₂₄	-0.004	0.000	-0.013	.	.	0.248	4.033

由系数表(表 4)可知, $X_1, X_2, X_3, X_{10}, X_{11}, X_{19}$ 的 VIF 系数均远远大于 10, 表明这些变量有非常明显的共线性, 为使得模型预测更为精确, 使用主成分分析法, 消除原始变量中的多重共线性, 在诸多变量中提取主成分。进而采用主成分回归法协助建立学生成绩和 24 个研究变量的回归方程[4]。

3. 数据的主成分分析

3.1. KMO 检验和巴特利球体检验

做主成分分析前要先做 KMO 检验, KMO 检验可以检验变量间的相关性和偏相关性。KMO 的检验系数取值范围应在 0 到 1, KMO 值越接近 1, 说明变量间的相关性越强, 反之越弱。

Table 5. KMO and Bartlett coefficients
表 5. KMO 和 Bartlett 系数

检测类别	值的范围	主成分分析适合情况
KMO 值	大于 0.9	非常适合
	0.8~0.9	很适合
	0.7~0.8	适合
	0.6~0.7	勉强适合
	0.5~0.6	不太适合
	小于 0.5	不适合
BartlettP 值	小于或者等于 0.01	适合

KMO 和 Bartlett 系数表(表 5)给出了 KMO 系数代表的意义, 调查问卷数据 KMO 和 Bartlett 系数检验输出结果如下:

Table 6. KMO test and Bartley sphere test
表 6. KMO 检验和巴特利球体检验

取样足够度的 Kaiser-Meyer-Olkin 度量。	0.843
近似卡方	11774.734
Bartlett 的球形度检验	df
	276
	Sig.
	0.000

由 KMO 检验和巴特利球体检验表(表 6)可知, 问卷数据的 KMO 统计量为 0.843, 超过 0.8, 即原始变量 KMO 检验通过, 原始数据的 Bartlett 的球形度检验 Sig. 值为 $0 < 0.05$, 说明调查问卷数据适合做主成分分析。

3.2. 寻找公因子的个数

公因子方差表(表 7)表示在被提取的公因子中, 原数据被表达的比例, 由图可知几乎所有原始变量原始数据提取比例都在 80% 左右, 所以提取出的 4 个主成分对于原始变量的阐释能力是较好的。

Table 7. Common factor variance
表 7. 公因子方差

	初始	提取
X ₁	1.000	0.833
X ₂	1.000	0.857
X ₃	1.000	0.895
X ₄	1.000	0.751
X ₅	1.000	0.770
X ₆	1.000	0.810
X ₇	1.000	0.431
X ₈	1.000	0.820
X ₉	1.000	0.788
X ₁₀	1.000	0.894
X ₁₁	1.000	0.866
X ₁₂	1.000	0.817
X ₁₃	1.000	0.802
X ₁₄	1.000	0.819
X ₁₅	1.000	0.795
X ₁₆	1.000	0.749
X ₁₇	1.000	0.777
X ₁₈	1.000	0.805
X ₁₉	1.000	0.814
X ₂₀	1.000	0.865
X ₂₁	1.000	0.812
X ₂₂	1.000	0.835
X ₂₃	1.000	0.821
X ₂₄	1.000	0.787

由解释的总方差表(表 8)可以看出主成分 F₁ 特征值为 7.651, 即 F₁ 表达了 7.651 个原始因素的信息, 主成分 F₂ 和主成分 F₃ 主成分 F₄ 特征值分别为 4.628, 3.787 和 3.147, 而 80.053% 的贡献率是由前四个主成分的累计贡献达到的, 因此使用前四个组成成分足以描述原始变量的大部分信息。

Table 8. The total variance of the interpretation
表 8. 解释的总方差

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的%	累积%	合计	方差的%	累积%	合计	方差的%	累积%
1	7.651	31.878	31.878	7.651	31.878	31.878	6.292	26.215	26.215
2	4.628	19.284	51.162	4.628	19.284	51.162	4.910	20.457	46.671
3	3.787	15.780	66.942	3.787	15.780	66.942	4.083	17.014	63.685
4	3.147	13.111	80.053	3.147	13.111	80.053	3.928	16.368	80.053
5	0.645	2.688	82.741						
6	0.547	2.280	85.021						
7	0.416	1.731	86.752						
8	0.408	1.698	88.451						
9	0.328	1.365	89.815						
10	0.285	1.188	91.003						
11	0.268	1.115	92.118						
12	0.259	1.080	93.198						
13	0.249	1.040	94.238						
14	0.218	0.908	95.146						
15	0.211	0.881	96.027						
16	0.183	0.761	96.787						
17	0.172	0.717	97.504						
18	0.155	0.645	98.149						
19	0.140	0.581	98.730						
20	0.122	0.510	99.240						
21	0.090	0.376	99.615						
22	0.046	0.192	99.808						
23	0.037	0.154	99.961						
24	0.009	0.039	100.000						

由方差解释表(表 8)和碎石图(图 1)可以看出, 本例保留了大于 1 的特征根, 即提取的 4 个主成分, 这样由分析原来的 24 个变量转化为仅需分析 4 个综合变量, 极大地起到了降维的作用。

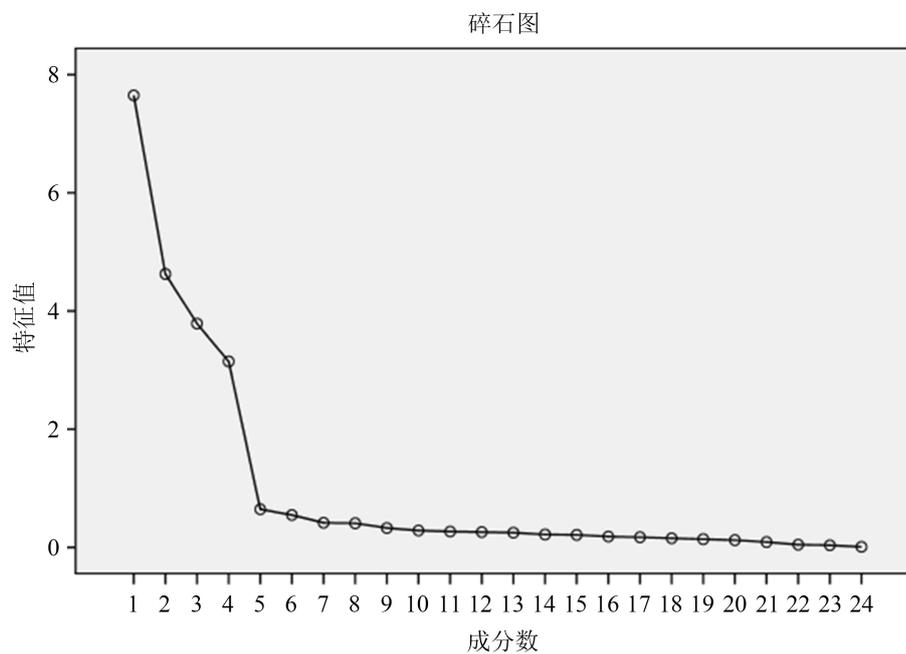


Figure 1. Gravel diagram

图 1. 碎石图

3.3. 主成分表达式及分析

Table 9. Composition matrix

表 9. 成份矩阵

	成份			
	1	2	3	4
X_1	0.775	-0.439	-0.151	0.127
X_2	0.528	0.617	-0.386	-0.221
X_3	0.792	-0.464	-0.190	0.127
X_4	0.356	0.381	0.466	0.512
X_5	0.483	0.006	0.604	-0.415
X_6	0.354	0.358	0.479	0.572
X_7	0.551	-0.316	-0.167	0.022
X_8	0.474	-0.048	0.613	-0.467
X_9	0.383	0.380	0.381	0.593
X_{10}	0.791	-0.461	-0.199	0.129
X_{11}	0.527	0.628	-0.384	-0.216
X_{12}	0.760	-0.466	-0.138	0.066
X_{13}	0.744	-0.424	-0.222	0.142

Continued

X ₁₄	0.483	0.648	-0.349	-0.212
X ₁₅	0.426	0.037	0.604	-0.498
X ₁₆	0.348	0.404	0.319	0.602
X ₁₇	0.495	0.625	-0.341	-0.161
X ₁₈	0.457	0.636	-0.385	-0.209
X ₁₉	0.769	-0.429	-0.146	0.131
X ₂₀	0.463	0.014	0.628	-0.506
X ₂₁	0.760	-0.434	-0.196	0.088
X ₂₂	0.384	0.415	0.405	0.593
X ₂₃	0.454	-0.001	0.592	-0.514
X ₂₄	0.475	0.627	-0.378	-0.161

成分矩阵表(表 9)给出了标准化后 24 个因素与 4 个主成分之间的线性关系系数, 为了得到含义更加明确, 实际意义也更为明显的主成分, 将原来的主成分进行旋转, 即将因子分析中的因子旋转后得到一组新的因子载荷矩阵。

Table 10. Rotate the component matrix
表 10. 旋转成份矩阵

	成份			
	1	2	3	4
X ₁	0.901	0.058	0.104	0.077
X ₂	0.116	0.914	0.061	0.063
X ₃	0.939	0.066	0.082	0.054
X ₄	0.020	0.072	0.149	0.850
X ₅	0.140	0.045	0.856	0.129
X ₆	0.038	0.032	0.121	0.890
X ₇	0.647	0.086	0.073	-0.024
X ₈	0.153	0.012	0.889	0.072
X ₉	0.077	0.101	0.050	0.877
X ₁₀	0.939	0.072	0.075	0.052
X ₁₁	0.109	0.919	0.058	0.071
X ₁₂	0.891	0.043	0.145	0.024
X ₁₃	0.890	0.085	0.034	0.051

Continued

X ₁₄	0.055	0.897	0.064	0.085
X ₁₅	0.065	0.065	0.884	0.064
X ₁₆	0.054	0.129	-0.012	0.854
X ₁₇	0.084	0.868	0.043	0.120
X ₁₈	0.052	0.893	0.027	0.058
X ₁₉	0.890	0.059	0.103	0.084
X ₂₀	0.099	0.056	0.920	0.072
X ₂₁	0.892	0.087	0.090	0.025
X ₂₂	0.051	0.115	0.066	0.903
X ₂₃	0.108	0.060	0.897	0.041
X ₂₄	0.077	0.879	0.010	0.098

由旋转成分矩阵表(表 10)可以得到方差最大化正交旋转后原始变量与提取的 4 个主成分之间的因子载荷,即原始变量与提取的 4 个主成分之间的线性方程系数,由此可得原始变量与提取的主成分之间的回归表达式:

$$X_1^* = 0.901F_1 + 0.058F_2 + 0.104F_3 + 0.077F_4$$

$$X_2^* = 0.116F_1 + 0.914F_2 + 0.061F_3 + 0.063F_4$$

$$X_3^* = 0.939F_1 + 0.066F_2 + 0.082F_3 + 0.054F_4$$

...

$$X_{24}^* = 0.077F_1 + 0.879F_2 + 0.010F_3 + 0.098F_4$$

由成分矩阵表进一步计算,得出平均绩点和 24 个变量的得分系数矩阵表。

Table 11. Component score coefficient matrix

表 11. 成份得分系数矩阵

	成份			
	1	2	3	4
X ₁	0.148	-0.016	-0.015	0.003
X ₂	-0.008	0.193	-0.003	-0.025
X ₃	0.156	-0.014	-0.022	-0.004
X ₄	-0.016	-0.022	0.005	0.223
X ₅	-0.016	-0.010	0.216	-0.001
X ₆	-0.011	-0.033	-0.004	0.236
X ₇	0.106	0.001	-0.009	-0.021

Continued

X ₈	-0.013	-0.016	0.227	-0.017
X ₉	-0.003	-0.017	-0.025	0.232
X ₁₀	0.156	-0.013	-0.023	-0.004
X ₁₁	-0.010	0.194	-0.004	-0.023
X ₁₂	0.146	-0.018	-0.001	-0.013
X ₁₃	0.149	-0.008	-0.033	-0.002
X ₁₄	-0.019	0.190	0.000	-0.017
X ₁₅	-0.030	-0.002	0.229	-0.020
X ₁₆	-0.005	-0.009	-0.040	0.227
X ₁₇	-0.013	0.181	-0.008	-0.006
X ₁₈	-0.017	0.190	-0.009	-0.023
X ₁₉	0.146	-0.016	-0.015	0.005
X ₂₀	-0.025	-0.005	0.237	-0.019
X ₂₁	0.147	-0.007	-0.017	-0.012
X ₂₂	-0.009	-0.015	-0.021	0.238
X ₂₃	-0.022	-0.003	0.231	-0.027
X ₂₄	-0.012	0.185	-0.016	-0.011

由得分系数矩阵表(表 11)输出结果可以得到 4 个主成分与 24 个因素之间的回归方程系数, 即回归方程可以表达为:

$$F_1 = 0.148X_1^* - 0.008X_2^* + 0.156X_3^* - 0.016X_4^* - 0.016X_5^* - 0.011X_6^* + 0.106X_7^* - 0.013X_8^* - 0.003X_9^* \\ + 0.156X_{10}^* - 0.010X_{11}^* + 0.146X_{12}^* + 0.149X_{13}^* - 0.019X_{14}^* - 0.030X_{15}^* - 0.005X_{16}^* - 0.013X_{17}^* \\ - 0.017X_{18}^* + 0.146X_{19}^* - 0.025X_{20}^* + 0.147X_{21}^* - 0.009X_{22}^* - 0.022X_{23}^* - 0.012X_{24}^*$$

$$F_2 = -0.016X_1^* + 0.193X_2^* - 0.014X_3^* - 0.022X_4^* - 0.010X_5^* - 0.033X_6^* + 0.001X_7^* - 0.016X_8^* - 0.017X_9^* \\ - 0.013X_{10}^* + 0.194X_{11}^* - 0.018X_{12}^* - 0.008X_{13}^* + 0.190X_{14}^* - 0.002X_{15}^* - 0.009X_{16}^* + 0.181X_{17}^* \\ + 0.190X_{18}^* - 0.016X_{19}^* - 0.005X_{20}^* - 0.007X_{21}^* - 0.015X_{22}^* - 0.003X_{23}^* + 0.185X_{24}^*$$

$$F_3 = -0.015X_1^* - 0.003X_2^* - 0.022X_3^* + 0.005X_4^* + 0.216X_5^* - 0.004X_6^* - 0.009X_7^* + 0.227X_8^* - 0.025X_9^* \\ - 0.023X_{10}^* - 0.004X_{11}^* - 0.001X_{12}^* - 0.033X_{13}^* + 0.000X_{14}^* + 0.229X_{15}^* - 0.040X_{16}^* - 0.008X_{17}^* \\ - 0.009X_{18}^* - 0.015X_{19}^* + 0.237X_{20}^* - 0.017X_{21}^* - 0.021X_{22}^* + 0.231X_{23}^* - 0.016X_{24}^*$$

$$F_4 = 0.003X_1^* - 0.025X_2^* - 0.004X_3^* + 0.223X_4^* - 0.001X_5^* + 0.236X_6^* - 0.021X_7^* - 0.017X_8^* + 0.232X_9^* \\ - 0.004X_{10}^* - 0.023X_{11}^* - 0.013X_{12}^* - 0.002X_{13}^* - 0.017X_{14}^* - 0.020X_{15}^* + 0.227X_{16}^* - 0.006X_{17}^* \\ - 0.023X_{18}^* + 0.005X_{19}^* - 0.019X_{20}^* - 0.012X_{21}^* + 0.238X_{22}^* - 0.027X_{23}^* - 0.011X_{24}^*$$

由四个主成分的表达式可看出:

第一个主成分 F_1 主要由 X_1^* 、 X_3^* 、 X_7^* 、 X_{10}^* 、 X_{12}^* 、 X_{13}^* 、 X_{19}^* 、 X_{21}^* 即对任课老师印象, 好胜心, 自信心, 愉快的情绪, 自制力, 求知欲, 幸福感, 兴趣爱好决定, 由于大多数情况都与学生心理方面相关, 因此我们把它统称成为心理因子。

第二个主成分 F_2 主要由 X_4^* 、 X_6^* 、 X_9^* 、 X_{16}^* 、 X_{22}^* 即由校园学习环境, 经济水平, 人际关系, 家庭氛围, 家庭学习环境决定, 我们把它统称为环境因子。

第三个主成分 F_3 主要由 X_5^* 、 X_8^* 、 X_{15}^* 、 X_{20}^* 、 X_{23}^* 即对科目的没有兴趣, 没有好的学习习惯, 知识衔接差, 学习目标不明确决定, 我们把它命名为缺陷因子。

第四个主成分 F_4 主要由 X_2^* 、 X_{11}^* 、 X_{14}^* 、 X_{17}^* 、 X_{18}^* 、 X_{24}^* 即对课内认真程度, 课外学习时间, 知识基础, 阅读浏览量, 社会实践经历, 考前准备决定, 我们把它统称为学习行为因子。

旋转空间中的成分图

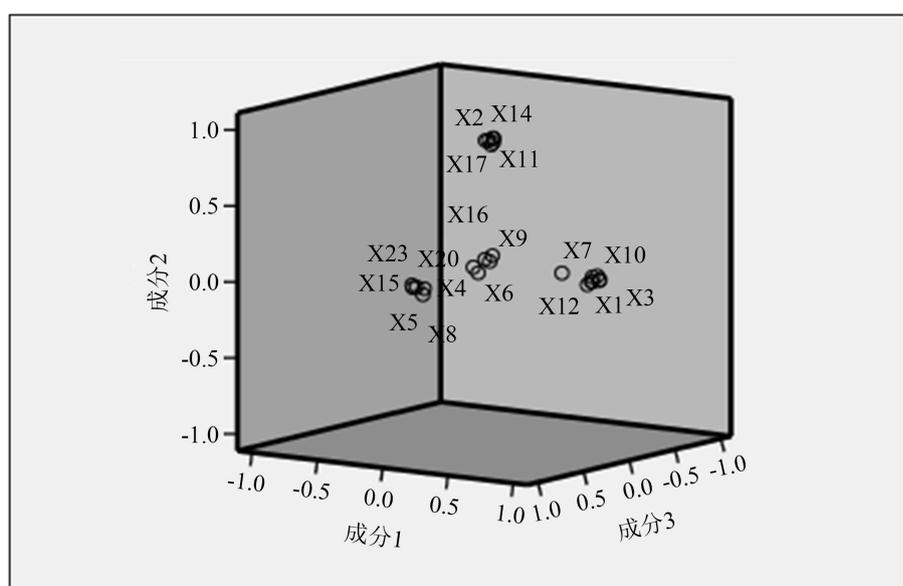


Figure 2. Rotate the spatial matrix diagram
图 2. 旋转空间矩阵图

旋转空间矩阵图(图 2)代表提取四个主成分后, 原始数据在四维空间上的分布。

3. 主成分回归分析

3.1. 学习成绩关于主成分的回归

由系数表(表 12)可知, 四个主成分的 VIF 检验值都为 1, 小于 10, 四个主成分之间不存在共线性, 则通过主成分分析提取主成分之后, 由 4 个主成分作为学生平均成绩的自变量是合适的。

由模型汇总表(表 13)可以看出: $R^2 = 0.830$, 调整后的 $R^2 = 0.828$, 可以知道, 回归后自变量对因变量的解释率为 82.8% [5]。

且根据方差分析表(表 14)可知: 回归模型中 4 个主成分关于因变量即学生成绩的检验 P 值 = $0.00 < 0.05$, 拒绝方程不显著的假设, 即 P 检验说明回归方程显著; 每个自变量对因变量都有显著的影响[6]。

根据系数表可以写出 y 关于 4 个主成分的回归方程, 即:

$$y = 3.316 + 0.285F_1 + 0.301F_2 + 0.251F_3 + 0.230F_4 \quad (3.1)$$

Table 12. Coefficient
表 12. 系数

模型	非标准化系数		标准系数		t	Sig.	共线性统计量	
	B	标准误差	试用版				容差	VIF
(常量)	3.316	0.012			275.613	0.000		
1 REGR factor score 1 for analysis 1	0.285	0.012	0.481		23.564	0.000	1.000	1.000
REGR factor score 2 for analysis 1	0.301	0.012	0.510		24.955	0.000	1.000	1.000
REGR factor score 3 for analysis 1	0.251	0.012	0.426		20.846	0.000	1.000	1.000
REGR factor score 4 for analysis 1	0.230	0.012	0.391		19.120	0.000	1.000	1.000

Table 13. Model rollup
表 13. 模型汇总

模型	R	R 方	调整 R 方	标准估计的误差
1	0.911a	0.830	0.828	0.2445

Table 14. Analysis of variance
表 14. 方差分析

模型	平方和	df	均方	F	Sig.
1 回归	118.804	4	29.701	496.836	0.000a
残差	24.390	408	0.060		
总计	143.195	412			

3.2. 主成分关于原始因素的回归

在四个主成分和平均绩点做回归分析之后，再做 24 个原始变量和四个主成分的回归分析，可得 24 个原始因素与主成分之间的回归表达式：

$$F_1 = -0.221 + 0.051X_1 - 0.003X_2 + 0.055X_3 - 0.005X_4 - 0.005X_5 - 0.003X_6 + 0.000X_7 - 0.004X_8 - 0.001X_9 + 0.055X_{10} - 0.003X_{11} + 0.050X_{12} + 0.053X_{13} - 0.007X_{14} - 0.010X_{15} - 0.001X_{16} - 0.004X_{17} - 0.006X_{18} + 0.051X_{19} - 0.009X_{20} + 0.051X_{21} - 0.003X_{22} - 0.008X_{23} - 0.004X_{24} \quad (3.2)$$

$$F_2 = -0.250 - 0.005X_1 + 0.064X_2 - 0.004X_3 - 0.007X_4 - 0.003X_5 - 0.011X_6 + 0.000X_7 - 0.005X_8 - 0.006X_9 - 0.004X_{10} + 0.065X_{11} - 0.005X_{12} - 0.002X_{13} + 0.062X_{14} - 0.001X_{15} - 0.003X_{16} + 0.060X_{17} + 0.063X_{18} - 0.005X_{19} - 0.002X_{20} - 0.002X_{21} - 0.005X_{22} - 0.001X_{23} + 0.060X_{24} \quad (3.3)$$

$$F_3 = -0.217 - 0.005X_1 + 0.064X_2 - 0.004X_3 - 0.007X_4 - 0.003X_5 - 0.011X_6 + 0.000X_7 - 0.005X_8 - 0.006X_9 - 0.004X_{10} + 0.065X_{11} - 0.005X_{12} - 0.002X_{13} + 0.062X_{14} - 0.001X_{15} - 0.003X_{16} + 0.060X_{17} + 0.063X_{18} - 0.005X_{19} - 0.002X_{20} - 0.002X_{21} - 0.005X_{22} - 0.001X_{23} + 0.060X_{24} \quad (3.4)$$

$$F_4 = -0.240 + 0.000X_1 - 0.008X_2 - 0.002X_3 + 0.071X_4 - 0.001X_5 + 0.076X_6 + 0.000X_7 - 0.006X_8 + 0.075X_9 - 0.002X_{10} - 0.008X_{11} - 0.005X_{12} - 0.002X_{13} - 0.006X_{14} - 0.007X_{15} + 0.074X_{16} - 0.002X_{17} - 0.008X_{18} + 0.001X_{19} - 0.006X_{20} - 0.005X_{21} + 0.076X_{22} - 0.009X_{23} - 0.004X_{24} \quad (3.5)$$

将公式(3.2)、(3.3)、(3.4)、(3.5), 代入公式(3.1)得到自变量 y (平均绩点)关于因变量 X (24 个因素)的回归表达式(3.6):

$$y = 3.068 + 0.012X_1 + 0.033X_2 + 0.013X_3 + 0.011X_4 - 0.004X_5 + 0.011X_6 + 0.000X_7 - 0.005X_8 + 0.014X_9 + 0.013X_{10} + 0.033X_{11} + 0.010X_{12} + 0.013X_{13} + 0.031X_{14} - 0.005X_{15} + 0.015X_{16} + 0.032X_{17} + 0.031X_{18} + 0.012X_{19} - 0.005X_{20} + 0.012X_{21} + 0.014X_{22} - 0.005X_{23} + 0.031X_{24} \quad (3.6)$$

由公式(3.6)可知, 每一个变量前都有它的回归系数, 可以通过比较变量前系数的绝对值大小, 分析每个变量对因变量的影响程度[7], 根据变量前的系数由大到小排序为:

$$y = 0.033X_2 + 0.033X_{11} + 0.032X_{17} + 0.031X_{14} + 0.031X_{18} + 0.031X_{24} + 0.015X_{16} + 0.014X_9 + 0.014X_{22} + 0.013X_{13} + 0.013X_{10} + 0.013X_3 + 0.012X_{19} + 0.012X_1 + 0.012X_{21} + 0.011X_4 + 0.011X_6 + 0.010X_{12} - 0.005X_{20} - 0.005X_{23} - 0.005X_8 - 0.005X_{15} + -0.004X_5 + 0.000X_7 + 3.068$$

在回归方程中系数绝对值较大的为 $X_2, X_{11}, X_{17}, X_{14}, X_{18}, X_{24}$ 。即课内认真程度, 课外学习时间, 知识基础, 阅读浏览量, 社会实践经历, 考前准备是需要重点研究的对象, 就此对平顶山学院学生学习影响较大的因素找到, 为学校提高教学水平改革提供一定基础。

4. 结论

通过制作的调查问卷收集整理数据, 问卷包括影响学生学习的 24 个因素和他们上学期的平均绩点, 然后对数据进行了初步整理, 做了关于数据可靠性的信度分析, 通过信度分析结果发现调查问卷的可靠性极高, 则说明收集的数据对于要做的回归分析问题有基础的依据[8]。然后对调查问卷数据进行了 VIF 检验, 结果显示多个研究因素的 VIF 值远远高于临界值, 则说明数据间存在多重共线性, 而存在共线性会使模型精度大大降低, 为使多重共线性消除。对原始变量进行主成分分析, 提取四个主成分, 并且各个主成分代表的含义各不相同, 用主成分和平均绩点做主成分回归时 VIF 值均小于标准值[9], 则通过共线性检验, 将学生成绩与提取的四个主成分做主成分回归分析。再将四个主成分与研究因素做线性回归, 代入整理得到学生成绩与研究因素的最终回归方程[10]。在仅限于调查问卷所涉及到的 24 个因素里, 将所研究因素做主成分回归后, 24 个研究因素分为了 4 个因子, 分别为心理因子, 环境因子, 缺陷因子, 学习行为因子。通过回归方程系数可知学习成绩关于原始 24 个因素的回归表达式中, 课内认真程度, 课外学习时间, 知识基础, 阅读浏览量, 社会实践经历, 考前准备在所有因素中系数绝对值较大, 即学习行为因子对我校学生学习程度的影响较大。

参考文献

- [1] 中国高等教育学会高等财经教育分会, 西南财经大学高等财经教育研究中心, 编. “双一流”背景下的高等财经教育[M]. 成都: 西南财经大学出版社, 2018.
- [2] 史静寰. 走向质量治理: 中国大学生学情调查的现状与发展[J]. 中国高教, 2016(2): 37-41.
- [3] 樊华强, 朱荣. 大学生学习情况调查研究的现状与瞻望[J]. 煤炭高等教育, 2015, 33(6): 98-101.
- [4] 梁兴堃. 图情档研究中的回归分析: 基本原理[J]. 图书情报知识, 2021, 38(3): 154-164.
- [5] 任升录. 关于线性回归模型的显著性检验[J]. 数学教学, 2012(3): 7-8.
- [6] 常海涛, 蔡静, 温悦, 等. 一种基于主成分分析的 TDLAS 高频噪声滤波[J]. 计量学报, 2022, 43(10): 1285-1290.
- [7] Coto, B. *et al.* (2022) Fast and Simplified Determination of PCA and Aromatic Carbon Content of Treated Distilled Aromatic Extract (TDAE) by NMR. *Analytical and Bioanalytical Chemistry*, **414**, 3109-3119. <https://doi.org/10.1007/s00216-022-03941-8>

- [8] Xiang, C.Y., Wu, L.G. *et al.* (2022) Characteristics of Extreme Rainfall and Rainbands Evolution of Super Typhoon Lekima (2019) during Its Landfall. *Frontiers of Earth Science*, **16**, 64-74. <https://doi.org/10.1007/s11707-021-0871-3>
- [9] 李刚, 梁家卷, 潘建新, 等. 多元统计分析及其应用[J]. 中国科学: 数学, 2020, 50(5): 571-584.
- [10] 苏毓淞, 刘江锐. 统计分析方法与美国政治学研究[J]. 美国研究, 2020, 34(3): 107-125+7-8.