

# 基于 R 语言的高校学生成绩的分析研究

## ——以阿坝师范学院为例

刘明月 胡欣 王春垚 吴莉

阿坝师范学院 数学与计算机科学学院, 中国·四川 汶川 623002

**【摘要】**采用因子分析与聚类分析对阿坝师范学院 2016 级数学与应用数学专业 116 名学生的 28 门课程成绩进行分析。更能客观地、有效地评价学生各学科的优势和劣势, 从而更深刻了解学生各方面能力高低, 并可以给予学生在就业和创业以及考研上的科学建议。对于老师而言可以提供更好的授课方式, 在管理等方面也可以提供更科学的依据。

**【关键词】**R 软件; 因子分析; 聚类分析; 成绩评价

**【基金项目】**四川省高校科研创新团队建设计划课题编号(18TD0047); 阿坝师范学院质量工程项目(201904020, 201908007, 201909002); 学生科研项目(S201910646060)。

### 1 引言

当今, 学校以及老师对学生的成绩都十分的重视, 也会对学生的成绩进行不同等级的划分。例如, 成绩在 90 分及以上的为 A, 小于 90 大于等于 80 为 B, 小于 80 大于等于 70 为 C, 小于 70 大于等于 60 为 D, 小于 60 为 E。虽然这种方法简单易行, 但还是存在一定的弊端, 因为这样不利于老师对学生成绩作出标准的评定, 而且还存在很大的误差, 当考试科目增多时, 学生成绩波动较大, 此方法便不再适用。因此我们利用自己的专业课程——《多元统计分析》来对学生成绩进行分析。而多元统计分析常用的方法有因子分析法和极大似然法。

通过极大似然估计与因子分析比较, 发现因子分析法的方差贡献率大于极大似然法, 因此我们将利用因子分析法来分析学生的各科成绩。因子分析法是主成分分析法的推广与发展, 因此利用聚类分析和因子分析能更加合理的对学生的成绩进行分析, 为教师教学水平的提高提供参考。

### 2 模型建立与分析

以阿坝师范学院 2016 级数学与应用数学专业 116 名学生从大一到大三主修课程为原始数据, 对于两学期及以上的课程, 则以平均成绩作为该科成绩, 每位学生记录 28 科成绩: X1-C 语言程序设计, X2-常微分方程, X3-抽象代数, X4-初等数学研究, X5-大学计算机基础, X6-大学体育, X7-大学物理 B, X8-大学英语, X9-大学语文, X10-多元统计分析, X11-概率统计, X12-高等代数, X13-公益劳动, X14-教育学, X15-军事理论与训练, X16-空间解析几何, X17-马克思基本原理概论, X18-毛泽东思想和中国特色社会主义理论, X19-普通话, X20-数学分析, X21-数学建模方法及应用, X22-数学实验, X23-思想道德修养与法律基础, X24-特色活动(文化遗产), X25-统计预测与决策, X26-心理学, X27-中国近现代史纲, X28-中学数学学科教学论(含微格教学训练)。因此得到一个  $116 \times 28$  的矩阵, 将数据存为 Excel 表。利用 R 软件, 先将原始数据标准化, 计算其相关系数矩阵, 运用主成分分析计算出初等载荷矩阵, 接着确定主因子, 利用方差旋转得到因子载荷矩阵及因子得分函数与综合因子得分, 按照因子得分对学生进行排名, 同时将学生的平均成绩排序。

由因子分析结果, 共有 12 个因子, 累计贡献率达到 80。

41%, 由旋转之后的因子载荷矩阵得出: Factor1 主要包括教育学、毛泽东思想和中国特色社会主义理论体系概论、常微分方程、多元统计分析, 反映了学生的理论思维能力和逻辑思维能力; Factor2 主要包括空间解析几何、高等代数、数学分析, 这三门是数学的基础课程, 反映了学生的数学基础能力; Factor3 主要包括抽象代数、马克思主义基本原理概论, 反映了学生的抽象思维能力; Factor4 主要包括公益劳动、军事理论与训练, 反映了学生实践与应用的能力; Factor5 主要包括 C 语言程序设计、数学实验, 反映了学生运用计算机软件编程的能力; Factor6 主要包括数学建模方法及应用、统计预测与决策, 反映了学生模型建立与求解的能力; Factor7 主要包括普通话, 反映了学生的语言表达能力; Factor8 主要包括中国近现代史纲要, 反映了学生对历史的了解情况; Factor9 主要包括大学体育, 主要反映了学生运动方面的能力; Factor10 主要包括大学计算机基础, 反映了学生对计算机的基本运用的能力; Factor11 主要包括中学数学学科教学论(含微格教学训练), 反映了学生数学教育方面的能力; Factor12 主要包括特色活动(文化遗产), 反映了学生对学校当地的特色文化的了解情况。

旋转后的因子载荷矩阵

因子得分函数:

$$\text{Factor1} = 0.433X_1 + 0.657X_2 + \dots + 0.307X_{28}$$

$$\text{Factor2} = 0.197X_1 + 0.3881609X_2 + \dots + 0.307X_{28}$$

||

$$F = 0.148F_1 + 0.106F_2 + 0.092F_3 + 0.647F_4 + 0.639F_5 + 0.592F_6 + 0.528F_7 + 0.502F_8 + 0.442F_9 + 0.430F_{10} + 0.400F_{11} + 0.395F_{12}$$

由此得到了每位同学的综合得分与排名, 前五名的学生总分如下由此得到了每位同学的综合得分与排名, 前五名的学生总分如下:

51 号 0.606, 28 号 0.580, 114 号 0.517, 26 号 0.409, 68 号 0.488. 学生学号排名为: [51, 28, 114, 26, 68, 13, 44, 95, 74, 63, 23, 7, 110, 6, 82, 88, 90, 69, 98, 5, 72, 41, 59, 22, 43, 50, 62, 33, 113, 18, 109, 102, 58, 8, 83, 36, 65, 101, 105, 52, 55, 96, 79, 64, 86, 108, 77, 47, 100, 46, 9, 66, 106, 21, 12, 111, 48, 3, 97, 87, 78, 32, 81, 19, 61, 10, 73, 54, 93, 40, 115, 16, 25, 45, 42, 70, 104, 75, 29, 38, 91, 49, 37, 39, 14, 24, 60, 17, 34, 94, 71, 20, 56, 92, 103, 89, 107, 11, 116, 85, 80, 67, 57, 99,

4, 112, 84, 15, 35, 76, 31, 1, 30, 2, 27, 53]与平均成绩排名:[28, 51, 63, 44, 114, 26, 110, 23, 95, 6, 7, 69, 68, 13, 74, 5, 90, 55, 108, 113, 43, 8, 82, 22, 41, 72, 88, 52, 62, 64, 59, 96, 83, 98, 50, 33, 58, 65, 36, 18, 102, 101, 9, 2, 47, 16, 77, 105, 46, 29, 32, 79, 109, 86, 3, 19, 100, 54, 97, 42, 40, 14, 17, 45, 93, 21, 12, 34, 61, 25, 10, 66, 81, 106, 78, 87, 24, 39, 91, 111, 49, 115, 73, 75, 48, 37, 70, 38, 103, 60, 56, 104, 89, 20, 71, 112, 1, 31, 94, 11, 99, 92, 4, 85, 116, 107, 80, 84, 76, 57, 35, 30, 15, 67, 27, 53]有很明显的区别。

因子分析中变量之间是相互独立的,即各个主因子所代表的信息是不重叠的,也是分开的。因此,采用主因子作为变量再对学生进行分类分析,主要参考离差平方和法,之后用 kmeans()函数对因子得分进行动态聚类分析,依此将学生分为四类,如下表所示。

表1 聚类类别序号

类别	序号
1	1, 2, 28, 59, 60, 61, 62, 68, 72, 80, 83, 85, 87, 90, 95, 98, 104, 106, 107, 109, 112, 113, 114, 115, 116
2	11, 12, 15, 18, 31, 34, 35, 36, 37, 38, 40, 45, 47, 49, 53, 54,
3	8, 13, 14, 17, 25, 26, 30, 32, 39, 41, 44, 46, 50, 52, 71, 73, 81, 82, 88, 89, 91, 102, 110
4	19, 63, 64, 65, 67, 69, 70, 74, 75, 76, 77, 78, 79, 84, 86, 96, 97, 99, 93, 92, 94, 100, 101, 103, 105, 108, 111
5	3, 4, 5, 6, 7, 9, 10, 16, 20, 21, 22, 23, 24, 27, 29, 33, 42, 43, 48, 51, 55, 56, 57, 58

动态聚类分析法聚类的特点为:类内的相似度高,而类间的相似度较低。将分数差异很小的学生聚在同一类中,分数差异很大的学生聚在不同类中。第一类学生的因子1, 10能力强,说明善于寻找不同事物的关联性与对待事物的专注度高,理论思维能力强,而在因子3的学习中存在劣势,逻辑思维能力差。这类学生应该多锻炼自己的逻辑思维能力;第二类学生在因子8学习能力强,说明对历史很感兴趣,而在因子3,的学习能力不足,因子10的能力很弱,需要更加努力去学习,去提升理论思维能力,去提高对其他学科的兴趣。第三类学生在因子2, 3的能力很强,抽象思维很发达,专业学科掌握能力与应用方面很强,然而在这种需要不断识记的学科上能力弱。他们应该多花时间去学习非专业学科的知识。第四类学生在因子3有优势,有很好的抽象思维能力,在因子2的能力弱,在创新和仔细方面的能力较弱。第五类学生在所有的因子中的能力都不太强,特别是在因子1, 3, 9的能力较弱,因子11的能力非常弱,即对专业知识的应用能力很

弱。这类学生应该更加的努力,需要去思考自己读大学的意义。

运用R软件进行因子分析和聚类分析对学生成绩进行合理的评价,很好的解决了学生数量多,学科科目多,评价指标多等带来的计算量大的问题,也让学生了解到多元统计分析的复杂性。通过这两种方法的结合,我们可以得到较客观合理的结果,突破了传统的计算平均数等统计方法,知道学生对学科知识的掌握情况,提高了综合测评的有效性与合理性,也能更加全面的了解学生学习能力的差异。对于老师而言,可以找到更好的教学方式,教学研究,教学管理以及为学生就业提供科学的参考依据。对于教学管理部门而言,可以作出合理的决策,来进一步提升学生的综合素养。对学生而言,促进了自己的发展,专业能力也得到了提高,同时也了解自己在学习中存在的问题。

#### 参考文献:

- [1]王斌会.多元统计分析及建模[M].暨南大学出版社,2014:186-225.
- [2]管宇.实用多元统计分析[M].浙江大学出版社,2001:159-214.
- [3]李珊珊,李权.基于R软件因子分析与聚类分析法的学生成绩评价[J].电脑编程技巧与维护,2016.
- [4]孙文兴,王俊英,蒋励.因子分析法在高职院校学生成绩评价中的应用研究[J].昆明冶金高等专科学校学报,2019,35(01):101-106.
- [5]张晓军,李珊珊,杨树生.基于MATLAB的因子分析与聚类分析在学生成绩评价中的应用[J].聊城大学学报(自然科学版),2015,28(02):71-75.
- [6]刘平清,朱桂玲,张慧愿.基于因子分析与TOPSIS模型的医疗水平综合评价[J].电脑知识与技术,2020,16(29):10-12.
- [7]陈喜华,黄海宁,黄沛杰.基于聚类分析的学生成绩分析[J].清远职业技术学院学报,2018,11(02):64-70.
- [8]电子资料,<http://wenku.baidu.com>,<http://www.doc88.com>.

#### 作者简介:

- 刘明月(1998.04-),女,民族:汉,籍贯:四川雅安,学历:本科,职称:在读学生,研究方向:代数与统计。
- 胡欣(1997.10-),女,民族:汉,籍贯:四川达州,学历:本科,职称:在读学生,研究方向:代数与统计。
- 王春垚(1997.12-),女,民族:汉,籍贯:四川达州,学历:本科,职称:在读学生,研究方向:代数与统计。
- 吴莉(1982-),女,民族:汉族,籍贯:四川崇州人,学历:硕士,副教授,研究方向:代数数论及统计学。