Pure Mathematics 理论数学, 2018, 8(5), 491-498
Published Online September 2018 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2018.85066

Hans汉斯

Cluster Analysis of Students' Employment Potential Based on SPSS

Luoman Li[1], Yalan Gao[1], Jiabo Xu[2]

[1]Xinjiang University of Finance and Economics, Urumqi Xinjiang
[2]Xinjiang Institute of Technology, Urumqi Xinjiang
Email: 1575900814@.qqcom, 1598724590@.qqcom, xujiabo_math@aliyun.com

Received: Aug. 17th, 2018; accepted: Sep. 4th, 2018; published: Sep. 11th, 2018

Abstract

This paper examines the test scores of selected courses of 15 graduate students in a school's statistics major. The mathematical model of cluster analysis is established. The specific steps of cluster analysis are given. The corresponding results are obtained by using SPSS software. According to the results of statistical analysis, they objectively evaluate their employment potential, and provide some reference opinions for the career direction of the 15 graduates. The analysis process and method can also be applied to other fields, providing some theory for university education and employment management.

Keywords

Cluster Analysis, Pearson Correlation Coefficient, Approximate Matrix, Employment Potential

基于SPSS的学生就业潜能的聚类分析

李罗蔓[1]，高亚兰[1]，徐加波[2]

[1]新疆财经大学，新疆 乌鲁木齐
[2]新疆工程学院，新疆 乌鲁木齐
Email: 1575900814@.qqcom, 1598724590@.qqcom, xujiabo_math@aliyun.com

收稿日期：2018年8月17日；录用日期：2018年9月4日；发布日期：2018年9月11日

摘　要

本文考察了某学校统计学专业15名研究生在校期间的所选课程的考试成绩，建立了聚类分析的数学模型，给出了聚类分析的具体步骤，运用SPSS软件得到相应结果，并根据统计分析的结果对他们的就业潜能做

文章引用: 李罗蔓, 高亚兰, 徐加波. 基于 SPSS 的学生就业潜能的聚类分析[J]. 理论数学, 2018, 8(5): 491-498.
DOI: 10.12677/pm.2018.85066

出客观评价，为该15名毕业生职业方向提供一些参考意见，该分析过程和方法还可以应用到其他领域，为大学教育教学和就业管理提供一些理论依据。

关键词

聚类分析，皮尔逊相关系数，近似矩阵，就业潜能

1. 背景

如今，就业人才供求的多元化，高校扩招政策的不断推进，我们国家的经济速度增长逐渐变缓，使得大学生的就业形势日益严峻，国家相继出台了一系列政策促进大学生就业。由于社会经济发展变化性导致高校所开设专业与社会的供求不完全匹配和教育质量的下滑，就会带来多多少少的就业压力。从学历上看，研究生初次就业率最高，本科初次就业率略低，高职高专初次就业率最低。从专业看，工科毕业生就业率较高，理科和文史哲类毕业生就业率较低。从毕业院校看，重点大学就业率较高，普通本科和独立学院就业率较低。

研究生的就业方向与所学的专业，以及自己的兴趣、爱好和性格特点密切相关。不同的人能力倾向是有差异的了解并分析自己的能力倾向，十分有助于毕业后的就业选择。本文通过收集某学校的统计学专业的 2015 级研究生的成绩，利用 SPSS 软件，进行聚类分析，客观分析其就业潜能，为其提供就业方向指导性意见。

通过聚类分析对学生潜能进行评价，是提高学生就业能力和就业质量的一种定量分析方法。所谓“聚类分析”，它是一种教育统计分析方法，是数据挖掘技术的重要组成部分，它能够在不同的潜在数据中发现数据分布模式，从而找出修正这一模式的方法[1]。换言之，通过采集学生的学习成绩数据，对学生的能力倾向进行“聚类分析”，从而有的放矢地对学生进行就业指导。具体操作是：首先，对学生的学业成绩进行“聚类分析”，把学生按某种能力属性分成若干小组(类)；再根据各门课程的特点分析每类学生的能力倾向；最后，根据每类学生的能力倾向和不同职业特点进行科学的就业指导。

2. 建立数学模型及分析步骤

1、聚类分析原理

聚类分析又称群分析，聚类是将相近的样本合并为同一类，根据“物以类聚”的思想，研究样本分类问题的一种多元统计方法。所谓类，就是指相似元素的集合[2]。聚类分析的研究目的是把相似的东西归成类，根据相似的程度将研究目标进行分类，聚类分析的研究对象分为两种，一是对变量分析(R 型)，二是对样品进行分类(Q 型)。聚类分析中，个体之间的“亲疏程度”是极为重要的，它将直接影响最终的聚类结果。对“亲疏程度”的测度一般有两个角度：第一，样品之间的亲疏程度；第二，变量之间的亲疏程度。衡量样品之间的亲疏程度通常通过某种距离来测度。变量之间的亲疏程度通常可采用简单相关系数或等级相关系数等[3]。定义样品间距离的方法也有很多，比如：欧氏距离、绝对值距离、切比雪夫距离、明考斯基距离等。常见的聚类分析方法有层次聚类和 K-Means 聚类。

2、数据来源

通过收集某学校统计专业的 15 名 2015 级研究生的 16 门主修课程的成绩，分别有经典著作选读、现代西方经济学、中国特色社会主义理论与实践研究、经济应用数学、应用数理统计、金融数学、英语、英语口语与听力、马克思主义与社会科学方法论、测度论与概率论基础、社会科学软件 spss、MATLAB 软件包、计量经济学的方法与应用、非参数统计、应用多元统计方法、时间序列分析，如表 1 所示。

将这 16 门主修课程的成绩输入 SPSS17.0 软件对其进行聚类分析。还有一些选修课程在表 1 中没有体现出来，选修科目是根据个人爱好选择的，包括数据模型与决策有 13 人选，空间统计学有 14 人选，随机过程与应用有 13 人选，现代统计模型有 11 人选，精算数学有 5 人选等等。还有一些学生的科研成果，综合成绩在表 1 中也没有体现出。

本文首先是对 15 名研究生主修成绩聚类分析做客观分析，其次是通过他们的三年来的科研成果和综合成绩做出一些主观分析。

3、建立数学模型

1) 本文所采用的是聚类分析数学模型：

样本间的距离用的是欧氏距离平方

$$d_{ij}=\sqrt{\sum_{l=1}^{p}\left(x_{il}-x_{jl}\right)^{2}}$$

其中 d_{ij} 表示样品 x_i 与 x_j 的距离。

Table 1. Comprehensive results
表 1. 综合成绩

姓名	经典著作选读	现代西方经济学	中国特色社会主义理论与实践研究	经济应用数学	应用数理统计	金融数学	英语	英语口语、听力	马克思主义与社会科学方法论	测度论与概率论基础	社会科学软件 SPSS	MATLAB 软件包	计量经济学的方法与应用	非参数统计	应用多元统计方法	时间序列分析
1	78	77	83	83	90	85	85	89	91	91	94	91	83	84	88	50
2	85	87	83	93	94	92	80	89	88	84	93	96	86	81	92	60
3	84	87	84	84	94	90	80	83	91	80	94	83	76	80	91	60
4	86	89	81	88	94	94	83	88	87	88	93	97	86	80	91	68
5	77	86	83	95	94	89	78	89	86	84	93	92	83	79	92	60
6	87	87	90	85	93	92	80	91	91	86	93	98	88	87	93	78
7	84	86	85	92	93	93	81	87	83	79	94	90	83	81	91	60
8	89	87	85	88	94	93	82	86	89	83	94	93	86	83	92	83
9	77	90	88	90	94	94	78	93	86	91	94	93	76	79	92	60
10	87	86	89	82	93	94	82	80	88	79	94	91	84	79	90	71
11	79	85	88	95	94	92	84	94	89	88	93	97	82	83	91	64
12	79	86	88	85	93	95	86	92	92	88	93	98	90	83	92	92
13	77	88	85	84	93	93	81	85	83	81	94	82	84	75	91	78
14	79	85	86	82	94	94	80	91	88	79	94	91	80	80	90	78
15	84	89	87	100	94	97	85	97	90	88	94	91	76	75	91	47

注：为方便运算把英语一和英语二合并取平均数。

度量变量间的亲疏程度关系用的是 Pearson correlation 皮尔逊相关系数，在统计学中，皮尔逊积矩相关系数用于度量两个变量 X 和 Y 之间的相关(线性相关)，其值介于−1 与 1 之间。在自然科学领域中，该系数广泛用于度量两个变量之间的相关程度。它是由卡尔·皮尔逊从弗朗西斯·高尔顿在 19 世纪 80 年代提出的一个相似却又稍有不同的想法演变而来的。这个相关系数也称作“皮尔森相关系数 r”。

4、分析步骤

聚类分析有很多，比如系统聚类、动态聚类等，本文用的是系统聚类，所谓的系统聚类先将 n 个样品各自看成一类，然后规定样品之间的“距离”和类与类之间的距离。选择距离最近的两类合并成一个新类，计算新类和其它类(各当前类)的距离，再将距离最近的两类合并。这样，每次合并减少一类，直至所有的样品都归成一类为止。

第一步，建立 15 名学生成绩的 Excel 样本文档如上述(表 1)。

第二步，打开并进入 SPSS 系统，在“文件”菜单中，将上述表中的数据导入系统中。

第三步，进入系统分析功能，在“分析”菜单“分类”中选择“系统聚类”命令。

第四步，在弹出的系统聚类分析对话框中，从对话框左侧的变量列表中选择“经典著作选读、现代西方经济学、中国特色社会主义理论与实践研究、经济应用数学、应用数理统计、……”等变量，使之添加到右边的变量框中。

第五步，确定变量的 R 型聚类，单击统计量按钮，选择相似性矩阵，聚类成员中选择单一方案，聚类成员写 3。

第六步，单击绘制按钮，选择树状图，选择冰柱中的所有聚类，方向为垂直。

第七步，单击方法按钮，选中聚类方法项，并选择组间聚类，度量标准中的区间项选择 Pearson 相关性。

第八步，输出统计量和图。

第九步，单击确定按钮，SPSS 自动完成分析过程。

3. 结果分析

1、第一部分输出的是系统 Q 型聚类的分析结果(见表 2)，从结果中可以看出 15 个样本都进入了聚类分析。

2、输出 SPSS 系统聚类分析各变量的距离矩阵(见表 3)。这部分输出的是系统 R 型聚类分析结果。从中可以看出各个变量之间的距离。

3、聚类表(见表 4)。在表格中的第一列表示的是聚类分析的第几步；第二列、第三列表示聚类中哪两个样本或小类聚成一类；第四列的系数表示的是相应的样本距离或小类距离；第五列、第六列表示本步聚类中，参与聚类的是样本还是小类。0 表示的是样本，数据 n (非 0)表示由第几步聚类产生的小类参与本步聚类；第七列表示本步聚类的结果将在下面聚类的第几步中用到。如表 4 所示。

表 5 是 R 型系统聚类分析后，聚类成 3 个类时变量的分类情况。

Table 2. Summary of case handling
表 2. 案例处理摘要

案例处理摘要 [a]

案例					
有效		缺失		合计	
N	百分比	N	百分比	N	百分比
15	100.0%	0	0%	15	100.0%

[a] 值向量间的相关性，已使用。

如图 1，聚类分析后的冰柱图，不同的群集数可以分不同的几类。

4. 结果及讨论

根据上述分析所得到的 3 个分类以及我们分别对每个类的定义，可以看到(表 6)，第一类学生在“经典著作选读、现代西方经济学、中国特色社会主义理论与实践研究、经济应用数学、英语口语、听力、应用数理统计、测度论与概率论基础、金融数学、应用多元统计方法”等方面学习能力比较强，具有较

Table 3. Approximate matrix
表 3. 近似矩阵

	近似矩阵															
	矩阵文件输入															
案例	经典著作选读	现代西方经济学	中国特色社会主义理论与实践研究	经济应用数学	应用数理统计	金融数学	英语口语、听力	马克思主义与社会科学方法论	测度论与概率论基础	社会科学软件 SPSS	英语	MATLAB 软件包	计量经济学的方法与应用	非参数统计	应用多元统计方法	时间序列分析
经典著作选读	1.000	0.262	0.036	−0.018	0.210	0.268	−0.344	0.216	−0.288	−0.007	0.065	0.203	0.254	0.242	0.236	0.114
现代西方经济学	0.262	1.000	0.172	0.329	0.821	0.755	0.084	−0.318	−0.122	−0.095	−0.360	0.010	−0.164	−0.454	0.716	0.180
中国特色社会主义理论与实践研究	0.036	0.172	1.000	−0.084	0.052	0.430	0.241	0.253	0.038	0.054	0.080	0.226	−0.004	0.189	0.269	0.326
经济应用数学	−0.018	0.329	−0.084	1.000	0.434	0.229	0.582	−0.166	0.302	−0.269	0.026	0.246	−0.266	−0.269	0.338	−0.531
应用数理统计	0.210	0.821	0.052	0.434	1.000	0.604	0.145	−0.190	−0.244	−0.214	−0.406	0.095	−0.239	−0.332	0.638	0.142
金融数学	0.268	0.755	0.430	0.229	0.604	1.000	0.257	−0.136	−0.118	0.040	0.125	0.175	−0.041	−0.399	0.406	0.333
英语口语、听力	−0.344	0.084	0.241	0.582	0.145	0.257	1.000	0.242	0.666	−0.302	0.252	0.499	−0.178	0.065	0.208	−0.204
马克思主义与社会科学方法论	0.216	−0.318	0.253	−0.166	−0.190	−0.136	0.242	1.000	0.374	−0.213	0.481	0.365	0.097	0.517	−0.049	0.083
测度论与概率论基础	−0.288	−0.122	0.038	0.302	−0.244	−0.118	0.666	0.374	1.000	−0.339	0.384	0.530	0.008	0.237	0.025	−0.260
社会科学软件 spss	−0.007	−0.095	0.054	−0.269	−0.214	0.040	−0.302	−0.213	−0.339	1.000	−0.056	−0.721	−0.582	−0.416	−0.498	−0.207
英语	0.065	−0.360	0.080	0.026	−0.406	0.125	0.252	0.481	0.384	−0.056	1.000	0.248	0.255	0.126	−0.393	0.049
MATLAB 软件包	0.203	0.010	0.226	0.246	0.095	0.175	0.499	0.365	0.530	−0.721	0.248	1.000	0.522	0.599	0.355	0.206
计量经济学的方法与应用	0.254	−0.164	−0.004	−0.266	−0.239	−0.041	−0.178	0.097	0.008	−0.582	0.255	0.522	1.000	0.533	0.255	0.639
非参数统计	0.242	−0.454	0.189	−0.269	−0.332	−0.399	0.065	0.517	0.237	−0.416	0.126	0.599	0.533	1.000	0.129	0.289
应用多元统计方法	0.236	0.716	0.269	0.338	0.638	0.406	0.208	−0.049	0.025	−0.498	−0.393	0.355	0.255	0.129	1.000	0.372
时间序列分析	0.114	0.180	0.326	−0.531	0.142	0.333	−0.204	0.083	−0.260	−0.207	0.049	0.206	0.639	0.289	0.372	1.000

Table 4. Clustering table
表 4. 聚类表

聚类表						
阶	群集组合			首次出现阶群集		
	群集 1	群集 2	系数	群集 1	群集 2	下一阶
1	2	5	0.821	0	0	2
2	2	6	0.679	1	0	6
3	7	9	0.666	0	0	8
4	13	16	0.639	0	0	9
5	12	14	0.599	0	0	9
6	2	15	0.587	2	0	10
7	8	11	0.481	0	0	11
8	4	7	0.442	0	3	13
9	12	13	0.388	5	4	11
10	1	2	0.244	0	6	12
11	8	12	0.217	7	9	14
12	1	3	0.192	10	0	13
13	1	4	0.062	12	8	14
14	1	8	0.034	13	11	15
15	1	10	0-.255	14	0	0

Table 5. Cluster members
表 5. 群集成员

群集成员	
案例	3 群集
经典著作选读	1
现代西方经济学	1
中国特色社会主义理论与实践研究	1
经济应用数学	1
应用数理统计	1
金融数学	1
英语口语、听力	1
马克思主义与社会科学方法论	2
测度论与概率论基础	1
社会科学软件 spss	3
英语	2
MATLAB 软件包	2
计量经济学的方法与应用	2
非参数统计	2
应用多元统计方法	1
时间序列分析	2

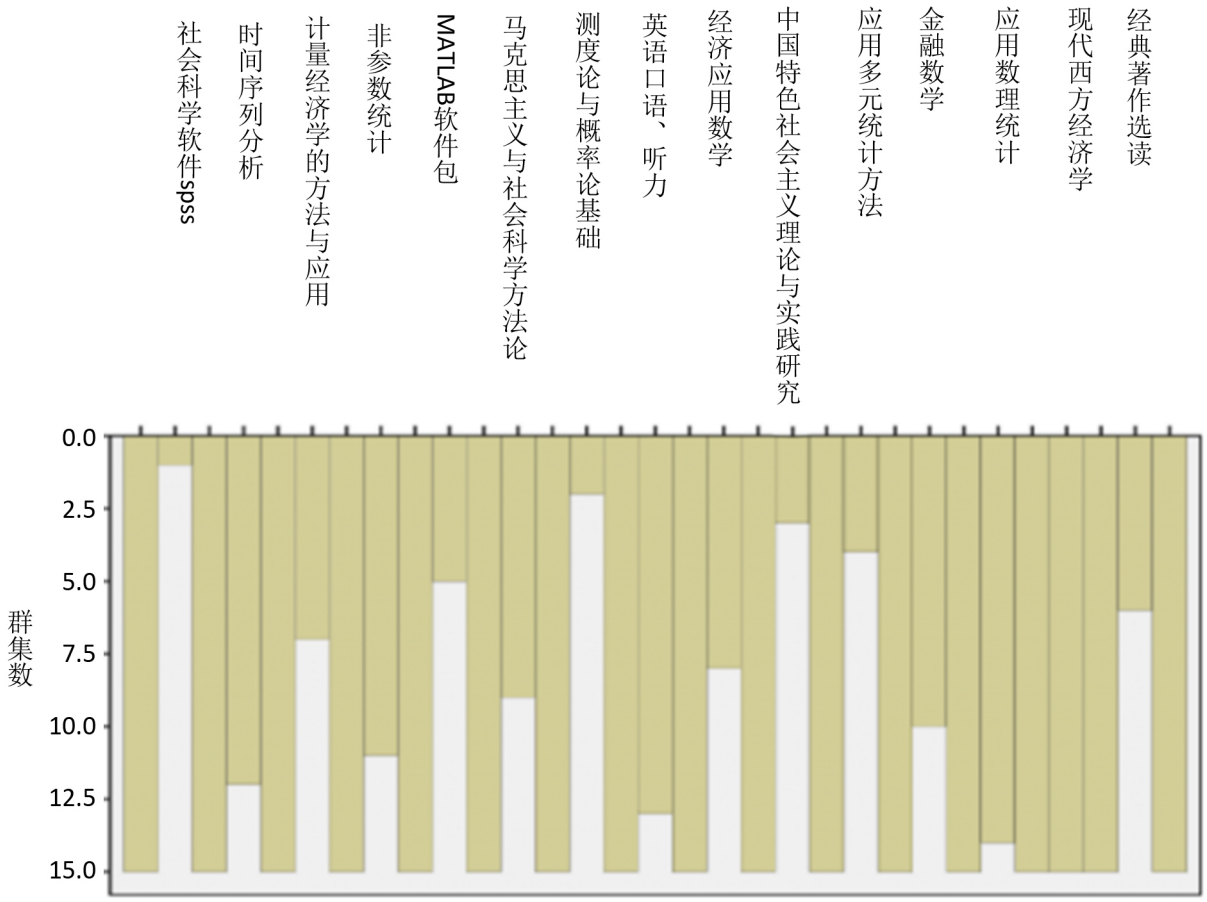

Figure 1. Icicles
图 1. 冰柱图

Table 6. Classification of courses
表 6. 课程分类

课程名称	课程分类
经典著作选读、现代西方经济学、中国特色社会主义理论与实践研究、经济应用数学、英语口语、听力、应用数理统计、测度论与概率论基础、金融数学、应用多元统计方法	第一类我们定义为偏重理论基础、偏经济金融知识
马克思主义与社会科学方法论、英语、MATLAB 软件包、计量经济学的方法与应用、非参数统计、时间序列分析	第二类我们定义为偏实践研究型
社会科学软件 spss	第三类为应用软件，偏应用处理数据

强的抽象思维、逻辑推理和外语基础，适合于从事科学研究工作，因此可以建议这些学生继续深造，有条件报考研究生的应动员其报考研究生，暂时不能考研的也要抓紧学习以争取进一步提高；第二类学生行政管理和政治敏锐力较强，建议他们可以去考公务员，将来可能会在政界取得比较大的成就；第三类学生，处理数据和动手能力较强，他们比较适合从事本专业技术方面的工作，建议他们进一步打好基础，深入掌握实际中的。根据个别同学的选修课程、综合成绩、科研成果进行一些个人的主观评价，仅供参考。15 名学生中，有 5 个学生选修了精算数学，并且成绩在都在 90 分以上，说明他们有扎实的计算能力，可以考虑比如保险方面的工作，只有一个学生选修了英语口语、听力二，成绩为 84，说明该学生在英语表达不错，有学习语言的潜力，可以考虑进一步学习或者在学习一些其他外语。综合成绩大家都在

35 分左右，最低 33.64 分，最高 36.8 分，差距不大，但在科研成果方面差距比较大有的为 0 分，最高为 201 分，大部分在五六十分，可以看出科研成果较高的人应该适合研究型，可以继续深造读博。

本文运用 SPSS 软件对 15 名在校研究生的考试成绩进行了聚类分析，根据聚类分析的结果判断其相应的就业潜能，具有一定的指导意义，相应地分析方法运用到大学生就业指导领域，为大学的教育教学和就业指导提供理论依据[4]。但是需要指出的是，决定学生就业潜能的因素是多方面的，考试成绩只能在一定程度上反映学生的学习能力，不能完全体现其就业潜能，仅仅凭借对课程成绩来判定学生的就业潜能具有片面性。因此，我们需要进一步研究和考察学生就业潜能分析，制定合理的评价体系，运用科学的统计方法，最大程度上准确断定学生的就业潜能。

参考文献

[1] 周蕾. 聚类分析在学生成绩分析中的应用[J]. 农业网络信息, 2010(5): 115-116.

[2] 罗家国, 罗浩, 仲佳嘉. 基于 SPSS 的学生能力倾向聚类分析研究[J]. 高等工程教育研究, 2012(6): 101-104, 135.

[3] 陈宇, 潘莹莹, 王娴, 祖冠群, 孙晓松, 艾玉波. 基于 SPSS 的统计专业学生能力倾向聚类分析[J]. 教育教学, 2013(11): 152-153.

[4] 张璇. 大学生一般就业能力及影响因素研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2012.