

多元统计分析法在公司招聘中的应用

罗丹*, 周柳, 黎勇, 林珊, 王松华#

百色学院数学与统计学院, 广西 百色

收稿日期: 2024年11月23日; 录用日期: 2024年12月19日; 发布日期: 2024年12月25日

摘要

文章以公司招聘中收到的应聘者的简历数据为研究对象, 利用主成分分析、聚类分析等多元统计分析方法, 将原公司招聘中的15个应聘指标重新组合成一组新的互相无关的六个综合指标, 同时在对事物类别、面貌等不清楚的情况下根据实际需要, 利用聚类分析从中可选取几个较少的组合指标尽可能多地反映原来指标的信息并进行有效分类, 这对公司招聘过程中能录取理想的应聘者具有一定实际参考意义。

关键词

企业招聘, 多元统计分析, 主成分分析, K-均值聚类, SPSS

The Application of Multivariate Statistical Analysis in Company Recruitment

Dan Luo*, Liu Zhou, Yong Li, Shan Lin, Songhua Wang#

School of Mathematics and Statistics, Baise University, Baise Guangxi

Received: Nov. 23rd, 2024; accepted: Dec. 19th, 2024; published: Dec. 25th, 2024

Abstract

The article resume data are received by the company A as the research object, using the principal component analysis, clustering analysis and multivariate statistical analysis method, the original 15 on your company's recruitment into a new set of applying index has nothing to do with each other six comprehensive index, at the same time in the category, the appearance such as unclear cases according to actual needs, using the combination of cluster analysis to select a few less index reflecting the original information as much as possible and effective classification. For the company hiring

*第一作者。

#通讯作者。

文章引用: 罗丹, 周柳, 黎勇, 林珊, 王松华. 多元统计分析法在公司招聘中的应用[J]. 统计学与应用, 2024, 13(6): 2461-2471. DOI: 10.12677/sa.2024.136238

process that can be admitted to the ideal candidate has a certain practical reference value.

Keywords

Company Recruitment, Multivariate Analysis, Principal Component Analysis, K-Means Clustering, SPSS

Copyright © 2024 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

公司招聘的质量对一个企业的发展就显得尤为重要。但当前大多数公司在招聘选拔流程的各个环节中没有科学的决策方法，评估工具有限且缺乏正规性，选拔中主观随意性比较强，这些问题导致公司不能够准确的选拔出优秀的员工[1]。如何在应聘者各类信息繁杂且人数众多的简历选出合适的且最优秀的人员已成为公司领导决策层非常关注的问题。

本文使用多元统计分析中的聚类分析法和主成份分析法对公司招聘进行应聘者优秀程度的筛选，其基本思想是，把原变量组合成一组新的独立互不相关的几个综合变量，同时根据实际需要从中可以取出几个较少的综合变量尽可能多地反映原来变量的信息，从而达到将高维信息进行降维处理的效果[2]。在此基础上利用 SPSS 统计软件实现相关数据的计算[3]，为招聘工作的人员选拔提供有效依据。

2. 公司招聘中的多元统计分析

2.1. 数据介绍

假设有 A 公司在本次招聘中收到的 48 份简历，公司根据 15 个判别标准对 48 份简历进行打分，其中 15 个标准与学历、外貌、责任感、自信、精明、诚实、仪表、经验、积极性、适应性、抱负、创新、读写、交流、申请信有关，每个评分标准分值为 0~10 分[4]，具体得分情况取自文献[2] (高惠璇(2005)) [5] 中的数据(如表 1 所示)。

Table 1. Score of resume received by company A
表 1. A 公司收到的简历表得分情况

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
应聘者	学历	经验	责任感	自信	精明	诚实	仪表	外貌	积极性	适应性	抱负	创新	读写	交流	申请信
1	6	7	2	5	8	7	8	8	3	8	9	7	5	7	10
2	9	10	5	8	10	9	9	10	5	9	9	8	8	8	10
3	7	8	3	6	9	8	9	7	4	9	9	8	6	8	10
4	5	6	8	5	6	5	9	2	8	4	5	8	7	6	5
5	6	8	8	8	4	4	9	5	8	5	5	8	8	7	7
6	7	7	7	6	8	7	10	5	9	6	5	8	6	6	6

续表

7	9	9	8	8	8	8	8	8	10	8	10	8	9	8	10
8	9	9	9	8	9	9	8	8	10	9	10	9	9	9	10
9	9	9	7	8	6	8	8	5	9	8	9	8	8	8	10
10	4	7	10	2	10	10	7	10	3	10	10	10	9	3	10
11	4	7	10	0	10	8	3	9	5	9	10	8	10	2	5
12	4	7	10	4	10	10	7	8	2	8	8	10	10	3	7
13	6	9	8	10	5	4	9	4	4	4	5	4	7	6	8
14	8	9	8	9	6	3	8	2	5	2	6	6	7	5	6
15	4	8	8	7	5	4	10	2	7	5	3	6	6	4	6
16	6	9	6	7	8	9	8	9	8	8	7	6	8	6	10
17	8	7	7	7	9	5	8	6	6	7	8	6	6	7	8
18	6	8	8	4	8	8	6	4	3	3	6	7	2	6	4
19	6	7	8	4	7	8	5	4	4	2	6	8	3	5	4
20	4	8	7	8	8	9	10	5	2	6	7	9	8	8	9
21	3	8	6	8	8	8	10	5	3	6	7	8	8	5	8
22	9	8	7	8	9	10	10	10	3	10	8	10	8	10	8
23	7	10	7	9	9	9	10	10	3	9	9	10	9	10	8
24	9	8	7	10	8	10	10	10	2	9	7	9	9	10	8
25	6	9	7	7	4	5	9	3	2	4	4	4	4	5	4
26	7	8	7	8	5	4	8	2	3	4	5	6	5	5	6
27	2	10	7	9	8	9	10	5	3	5	6	7	6	4	5
28	6	3	5	3	5	3	5	0	0	3	3	0	0	5	0
29	4	3	4	3	3	0	0	0	0	4	4	0	0	5	0
30	4	6	5	6	9	4	10	3	1	3	3	2	2	7	3
31	5	5	4	7	8	4	10	3	2	5	5	3	4	8	3
32	3	3	5	7	7	9	10	3	2	5	3	7	5	5	2
33	2	3	5	7	7	9	10	3	2	2	3	6	4	5	2
34	3	4	6	4	3	3	8	1	1	3	3	3	2	5	2
35	6	7	4	3	3	0	9	0	1	0	2	3	1	5	3
36	9	8	5	5	6	6	8	2	2	2	4	5	6	6	3
37	4	9	6	4	10	8	8	9	1	3	9	7	5	3	2
38	4	9	6	6	9	9	7	9	1	2	10	8	5	5	2
39	10	6	9	10	9	10	10	10	10	10	8	10	10	10	10

续表

40	10	6	9	10	9	10	10	10	10	10	10	10	10	10	10
41	10	7	8	0	2	1	2	0	10	2	0	3	0	0	10
42	10	3	8	0	1	1	0	0	10	0	0	0	0	0	10
43	3	4	9	8	2	4	5	3	6	2	1	3	3	3	8
44	7	7	7	6	9	8	8	6	8	8	10	8	8	6	5
45	9	6	10	9	7	7	10	2	1	5	5	7	8	4	5
46	9	8	10	10	7	9	10	3	1	5	7	9	9	4	4
47	0	7	10	3	5	0	10	0	0	2	2	0	0	0	0
48	0	6	10	1	5	0	10	0	0	2	2	0	0	0	0

2.2. 公司招聘中的主成分分析

主成分分析是考虑各指标之间的相互关系，利用降维的方法将多个指标转化为少数几个互不相关的指标，从而使进一步研究变得简单的一种统计方法。其中这几个互不相关的指标称为主成分，每个主成分均是原始变量的线性组合，且各个主成分之间互不相关，从而使得主成分比原始变量具有某些更优越的性能[6]。

Table 2. Eigenroots and variance contribution table
表 2. 特征根和方差贡献表

成份	初始特征值 α			提取平方和载入		
	合计	方差的%	累积%	合计	方差的%	累积%
原始	1	66.393	54.186	54.186	66.393	54.186
	2	18.28	14.919	69.105	18.28	14.919
	3	10.618	8.665	77.77	10.618	8.665
	4	6.765	5.521	83.292		
	5	3.986	3.253	86.545		
	6	3.628	2.961	89.506		
	7	2.895	2.362	91.869		
	8	2.832	2.311	94.18		
	9	1.96	1.599	95.779		
	10	1.635	1.334	97.114		
	11	1.106	0.903	98.017		
	12	0.851	0.694	98.711		
	13	0.725	0.591	99.303		
	14	0.521	0.425	99.728		
	15	0.334	0.272	100		

续表

重新标度	1	66.393	54.186	54.186	7.455	49.699	49.699
	2	18.28	14.919	69.105	2.04	13.601	63.3
	3	10.618	8.665	77.77	1.432	9.549	72.849
	4	6.765	5.521	83.292			
	5	3.986	3.253	86.545			
	6	3.628	2.961	89.506			
	7	2.895	2.362	91.869			
	8	2.832	2.311	94.18			
	9	1.96	1.599	95.779			
	10	1.635	1.334	97.114			
	11	1.106	0.903	98.017			
	12	0.851	0.694	98.711			
	13	0.725	0.591	99.303			
	14	0.521	0.425	99.728			
	15	0.334	0.272	100			

注：在主成分分析提取方法，协方差矩阵的分析中，原始特征值在全体原始解和重标刻度解中均相同。

一般来说，提取主成分的累计贡献率在 80%左右是较为合理的，主成分影响力度大小的指标是特征值，在多数情况下应当将累计贡献率和特征值结合起来以综合适合的数量[7]。

因为各个变量(指标)度量的单位都是分，故在协方差的基础上进行计算与分析，得出如表 2 所示的特征根和方差的贡献表和初始特征值 α 就是数据协方差矩阵的特征值。前三个特征值累积占了总方差的 77.77%，后边的特征值的贡献越来越低。

选取前三个主成份利用 SPSS 作碎石图，如图 1 所示，发现前三个主成分的碎石图是最陡的，这与表 1 的筛选结果是一致的，故用前三个主成分表示这十五个主成分是合理的。

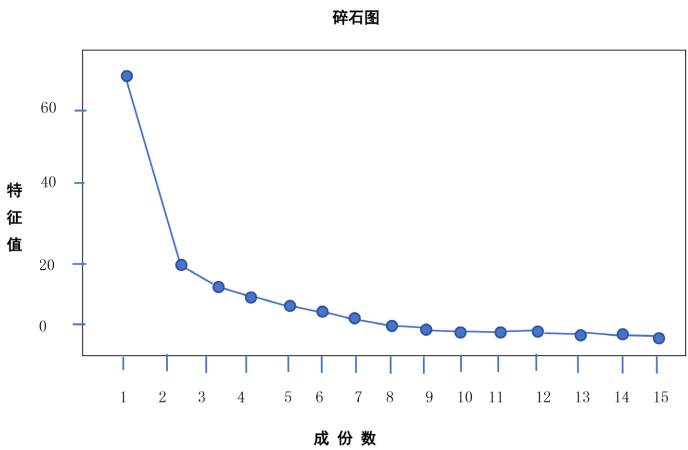


Figure 1. Gravel plot of the first three principal components
图 1. 前三个主成份的碎石图

由于主成份是原始变量的线性组合，根据表 2 的特征根和方差贡献值的计算出如表 3 所示的成份矩阵 α 。

Table 3. Component matrix α
表 3. 成份矩阵 α

	原始成份			重新标度成份		
	1	2	3	1	2	3
学历	1.215	1.588	0.641	0.454	0.594	0.24
经验	1.078	-0.12	0.141	0.548	-0.061	0.072
责任感	0.242	0.431	-0.431	0.122	0.217	-0.217
自信	1.657	-0.381	2.02	0.591	-0.136	0.72
精明	1.85	-1.059	-0.632	0.766	-0.438	-0.261
诚实	2.748	-0.835	-0.399	0.867	-0.263	-0.126
仪表	0.982	-1.272	1.463	0.387	-0.502	0.577
外貌	3.094	-0.391	-0.918	0.9	-0.114	-0.267
积极性	1.334	2.715	0.064	0.403	0.821	0.019
适应性	2.577	0.055	-0.367	0.874	0.019	-0.125
抱负	2.544	-0.519	-0.787	0.867	-0.177	-0.268
创新	2.764	-0.315	-0.161	0.911	-0.104	-0.053
写作	2.914	-0.102	0.136	0.915	-0.032	0.043
交流	1.843	-0.18	1.256	0.694	-0.068	0.473
申请信	2.237	2.01	0.041	0.678	0.609	0.012

注：提取方法是主成份，表中表示已提取了 3 个成分。

从表 3 中可以看出，每一列就是一个主成份作为原始变量的线性组合的系数，这些系数成为主成份载荷，其值(绝对值)越大，表示该主成份对于这个原始变量的代表性就越大。故第一主成份对“写作”、“创新”、“诚实”、“适应性”、“抱负”、“外貌”的解释很充分；第二主成份对“积极性”、“申请信”的解释很充分；第三主成份对“自信”的解释很充分。

3. 公司招聘中的聚类分析——K-均值聚类

在招聘方案设计中，一般要应聘者分为不同的种类，但无法判断具体分成几类是合适，为了解决这个问题，首先把数据进行聚类分析，并进行快速调试对比[8]。尽管随着聚类数的不同，得出的结论不尽相同，但是分析方式是相似的，故仅选取“聚类数为 2”情形进行一个完整的分析。

3.1. 聚类数为 2 的情形

表 4 是初始聚类中心表，由于没有指定初始聚类中心，故列出了由系统指定的聚类中心，与原数据比较，可见它们分别是 48 号和 40 号。

Table 4. Initial cluster center table
表 4. 初始聚类中心表

	聚类			聚类	
	1	2		1	2
学历	0	10	积极性	0	10
经验	6	6	适应性	2	10
责任感	10	9	抱负	2	10
自信	1	10	创新	0	10
精明	5	9	读写	0	10
诚实	0	10	交流	0	10
仪表	10	10	申请信	0	10
外貌	0	10			

根据表 4 得出如表 5 所示的迭代历史,由表 5 可知,第一次迭代后,2 个类的中心点分别变化了 11.737 和 10.099 一共进行了 3 次迭代,达到聚类结果的设置,聚类分析结束,并得出最终聚类结果如表 6 所示。

Table 5. Iteration history
表 5. 迭代历史

迭代	聚类中心内的更改	
	1	2
1	11.737	10.099
2	0.853	0.708
3	0	0

Table 6. Cluster center table of final clustering results
表 6. 最终聚类结果的类中心表

	聚类			聚类	
	1	2		1	2
学历	5	7	积极性	3	5
经验	6	8	适应性	3	7
责任感	7	7	抱负	3	8
自信	5	7	创新	4	8
精明	5	8	读写	3	8
诚实	4	8	交流	4	7
仪表	7	9	申请信	4	8
外貌	2	7			

表 7 是方差分析表,各项数据的含义分别是:组间均方、组间自由度、组内均方、组内自由度,其

中 Sig 值表示显著度，其置信区间通常是 95%。若 Sig 值大于 0.05，则说明接受原假设，即不同类别之间没有显著性差异。反之，则说明拒绝接受原假设，即不同类别之间有显著性差异。

Table 7. Analysis of variance table—aggregation number of 2
表 7. 方差分析表——聚合数为 2

	聚类		误差		F	Sig.
	均方	df	均方	df		
学历	30.561	1	6.64	46	4.603	0.037
经验	51.857	1	2.822	46	18.376	0
责任感	3.857	1	3.952	46	0.976	0.328
自信	66.667	1	6.594	46	10.111	0.003
精明	121.04	1	3.336	46	36.287	0
诚实	225.074	1	5.375	46	41.876	0
仪表	24.107	1	6.039	46	3.992	0.052
外貌	356.984	1	4.326	46	82.521	0
积极性	33.23	1	10.462	46	3.176	0.081
适应性	233.889	1	3.792	46	61.683	0
抱负	261.35	1	3.122	46	83.702	0
创新	222.354	1	4.579	46	48.557	0
读写	289.096	1	4.07	46	71.032	0
交流	60.86	1	5.89	46	10.332	0.002
申请信	164.827	1	7.545	46	21.845	0

从表 7 中可以看出，类别一与类别二的“责任感”因素没有显著性差异；类别一与类别二的“仪表”因素没有有显著性差异；类别一与类别二的“积极性”因素没有显著性差异。于是可以看出，把 48 个样本聚合成 2 类，这 2 类出现了较多的因素没有显著性差异，分类的效果明显不够理想。

3.2. 聚类数为 3, 4, 5, 6, 7, 8 的情形

与聚类数为 2 时，分析基本类似，只需要比较它们的方差分析表中 Sig 的值即可确定合适的聚类数，如下所示的表 8 分别表示聚类数为 3, 4, 5, 6, 7, 8 时，各指标的 Sig 值。

Table 8. Sig value table
表 8. Sig 值表

	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7	K = 8
学历	0.037	0.055	0	0	0	0	0
经验	0	0	0	0.009	0	0.001	0.002
责任感	0.328	0.408	0.123	0.795	0.06	0.072	0.075
自信	0.003	0.018	0	0	0	0	0
精明	0	0	0	0	0	0	0
诚实	0	0	0	0	0	0	0

续表

仪表	0.052	0	0	0	0	0	0
外貌	0	0	0	0	0	0	0
积极性	0.081	0	0	0	0	0	0
适应性	0	0	0	0	0	0	0
抱负	0	0	0	0	0	0	0
创新	0	0	0	0	0	0	0
读写	0	0	0	0	0	0	0
交流	0.002	0	0	0	0	0	0
申请信	0	0	0	0	0	0	0

通过观察，当聚合数等于 2 时，有三个 sig 值是大于 0.05 的，说明类别一和类别二对责任感、仪表、积极性的描述没有显著性差异，可见，样本聚合成两类还不够合理；当聚合数等于 3 时，有两个 sig 值是大于 0.05 的，说明类别一、类别二和类别三对学历和责任感的描述没有显著性差异，可见，样本聚合成三类依然不够合理；当聚合数等于 4 时，只有一个 sig 值是大于 0.05 的，说明类别一、类别二、类别三和类别四对责任感的描述没有显著性差异，此时 sig 值为 0.123，可见样本聚合成四类还是不够合理；当聚合数等于 5 时，只有一个 sig 值大于 0.05 的，说明类别一、类别二、类别三、类别四和类别五对责任感的描述没有显著性差异，此时 sig = 0.795,可见样本聚合成五类还是不够合理；当聚合数等于 6 时，只有一个 sig 值是大于 0.05 的，说明类别一、类别二、类别三、类别四、类别五和类别六对责任感的描述没有显著性差异，此时 sig = 0.06；当聚合数等于 7 时，只有一个 sig 值是大于 0.05 的，说明类别一、类别二、类别三、类别四、类别五、类别六和类别七对责任感的描述没有显著性差异，此时 sig = 0.072；当聚合数等于 8 时，只有一个 sig 值是大于 0.05 的，说明类别一、类别二、类别三、类别四、类别五、类别六、类别七和类别八对责任感的描述没有显著性差异，此时 sig = 0.075。

由此可得，当聚合数超过 4 时，不同类别进对一个指标的描述存在没有显著性差异，且对应的 sig 值先是减少，再是增加，在聚合数为 6 的时候，达到最小值，故将样本聚合成 6 类是成功的，于是各应聘者属于哪个类别，在表 9 中得到描述。

Table 9. Cluster member table
表 9. 聚类成员表

案例号	应聘者	聚类	距离	案例号	应聘者	聚类	距离
1	1	2	8.361	25	25	3	4.984
2	2	2	3.821	26	26	1	4.968
3	3	2	5.856	27	27	1	6.091
4	4	1	6.369	28	28	5	5.359
5	5	1	6.905	29	29	5	8.449
6	6	1	6.351	30	30	3	5.817
7	7	2	4.606	31	31	3	5.488
8	8	2	4.85	32	32	3	6.01

续表

9	9	2	5.483	33	33	3	5.914
10	10	6	3.59	34	34	5	4.552
11	11	6	4.888	35	35	5	6.511
12	12	6	3.543	36	36	1	7.053
13	13	1	6.397	37	37	1	9.186
14	14	1	6.168	38	38	1	9.064
15	15	3	6.61	39	39	2	6.707
16	16	2	5.276	40	40	2	6.753
17	17	1	6.378	41	41	4	2.915
18	18	1	6.415	42	42	4	2.915
19	19	1	6.578	43	43	3	9.357
20	20	1	6.888	44	44	2	6.539
21	21	1	5.717	45	45	1	6.772
22	22	2	5.105	46	46	1	7.591
23	23	2	5.363	47	47	5	7.145
24	24	2	6.207	48	48	5	7.169

4. 综合分析

综上所述，最终把公司招聘中的指标确定为 6 个类。第一类：(36, 45, 46, 14, 17, 6, 26, 5, 13, 18, 19, 4, 20, 37, 38, 21, 27)，这些应聘者的水平比较接近，除了在仪表方面得分都很高以外，在其他方面没有特别突出的表现；第二类：(39, 40, 2, 7, 8, 9, 22, 24, 3, 23, 44, 1, 16)，这些应聘者各方面的得分都很高，说明各方面都很优秀；第三类：(25, 31, 15, 30, 32, 43, 33)，这些应聘者各方面得表现都很平庸；第四类：(41, 42, 28)，这些应聘者在学历和责任感程度方面都得分很高，其他方面得分很低；第五类：(35, 29, 34, 47, 48)，这些应聘者的特点是在责任感和仪表方面表现突出，在其他方面得分变得很差，得分均在零附近；第六类：(12, 10, 11)，这些应聘者在某些方面表现极为突出。

5. 结论

本文对某公司招聘数据进行分析，以便为公司选择出理想的应聘者，采用了两种多元分析方法(主成分分析，聚类分析)，并以统计软件 SPSS 作计算分析，完成了统计分析过程，解决了以往人力资源主管在招聘公司员工时的盲目性问题[9]。对于一般情况，只需要用快速聚类方法，把 N 个应聘人员聚类为 M 类即可，再通过主成分分析法计算主成分，将主成分蕴涵的信息给与适当的解释，SPSS 软件的运用，简化了复杂的计算，提高了招聘过程中的可操作性。

通过统计分析可以知道，在主成分分析下，可以把 15 维的样本空间降到 3 维，考虑选择前三个主成分，用这三个主成分代替所有主成分；在聚类分析中的 K-均值法中，把 48 个应聘者聚成 6 类。最后可利用这两种方法得出的结果可以轻松的做进一步数据分析。公司可以根据自己的需求挑选符合要求的应聘者。

利用多元分析分析公司招聘，从理论到实践上看可以得到比较可行的分类，可是任何分析方法都有

其一定的局限性,当各考核项目之间不存在线性关系时,主成份分析方法便不适用了。虽然主成分分析、聚类分析可以对考核项目和应聘者进行较合理的分析,但仍存在不足,还需要不断地对应聘者在实际工作中认真观察。

基金项目

广西教育规划 2021 年度课题资助(项目编号: 2021A024)、广西自然科学基金资助(项目任务书编号: 2024GXNSFAA010478)、教育部产学研合作协同育人项目(项目编号: 220505876240149)、广西高校中青年教师科研基础能力提升项目(项目编号: 2024KY0757)。

参考文献

- [1] 张志军. 员工招聘与选拔实务[M]. 北京: 中国物资出版社, 2010.
- [2] 邓维斌, 唐兴艳, 胡大权, 周玉敏. SPSS19(中文版)统计分析实用教程[M]. 北京: 电子工业出版社, 2013.
- [3] 余建英, 等. 数据统计分析与 SPSS 应用[M]. 北京: 人民邮电出版社, 2003.
- [4] 侯静. 企业招聘的内部影响因素问题研究[J]. 商业现代化, 2008(28): 29-220.
- [5] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005.
- [6] 张建同, 孙昌言. 应用统计学[M]. 北京: 清华大学出版社, 2010.
- [7] 胡良平. 现代统计学与 SAS 应用[M]. 北京: 军事医学科学出版社, 2000.
- [8] 于秀林, 任雪松. 多元统计分析[M]. 北京: 中国统计出版社, 2011.
- [9] 冯梅. 聚类分析在公务员招聘中的应用及 SPSS 实现[J]. 数学的实践与认识, 2006, 36(10): 46-52.