

分地区行业平均工资水平分析

周子期

浙江师范大学数学科学学院, 浙江 金华

收稿日期: 2024年7月19日; 录用日期: 2024年8月11日; 发布日期: 2024年8月20日

摘要

本文基于2021年全国12个行业的分地区城镇单位就业人员平均工资数据, 使用聚类分析分别对全国主要城市与不同行业间进行分类, 同时利用判别分析验证分类的合理性, 从而研究了在地区 and 行业这两个方面中国的平均工资水平状况, 为将来的就业选择提供建议。

关键词

聚类分析, 判别分析, 平均工资

Analysis of the Average Wage Level of the Industry by Region

Ziqi Zhou

School of Mathematical Sciences, Zhejiang Normal University, Jinhua Zhejiang

Received: Jul. 19th, 2024; accepted: Aug. 11th, 2024; published: Aug. 20th, 2024

Abstract

Based on the average wage data of urban employees in 12 industries in China in 2021, this paper uses cluster analysis to classify major cities and different industries in China, and uses discriminant analysis to verify the rationality of the classification, so as to study the average wage level in China in terms of region and industry, and provide suggestions for future employment choices.

Keywords

Cluster Analysis, Discriminant Analysis, Average Wage



1. 引言

随着我国进入经济高速增长期，我国城镇单位就业人员平均工资水平显著提升，然而由于我国地理环境特殊，地区间发展不均衡现象加剧，导致不同区域间城镇单位就业人员的平均工资水平也呈现出日益明显的差异。工资差异反映了不同行业的发展状况和市场前景，它是求职者一直关心的问题，也是最根本最切合实际的问题。因此通过研究地区和行业间的工资水平差异，对劳动者做出更合理的职业选择有着非常重要的意义。李小春[1]等人基于因子分析、聚类分析和主成分分析对我国 31 个省的职工平均工资进行研究。刘辉[2]等人应用聚类分析，对不同地区、行业的平均工资差距较大问题进行分析。朱轩放[3]基于主成分分析和聚类分析对辽宁省城市职工平均工资进行研究。张子悦采用聚类分析的方法对我国 19 个不同类别行业年均工资水平进行聚类评价。刘茗[4]等人基于系统聚类法研究了在地区和行业这两个方面中国平均工资水平的状况。本文选取 2021 年全国 12 个行业的分地区城镇单位就业人员平均工资数据，通过聚类分析和判别分析，研究工资水平与地区综合实力发展、行业现状及前景的内在联系；同时分别将地区和行业按工资水平进行分类和排序，探究哪些地区、哪些行业工资水平较高，为就业地区与行业选择提供一定参考。

2. 实证分析

2.1. 数据来源

本文数据来源于国家统计局，选取 2021 年我国 31 个省、直辖市、自治区 12 个行业就业人员的平均工资数据作为原始变量，对全国城镇就业人员工资水平进行分析。这 12 个指标分别为 X_1 ：农、林、牧、渔业平均工资(元)； X_2 ：采矿业平均工资(元)； X_3 ：制造业平均工资(元)； X_4 ：电力、燃气及水的生产和供应业平均工资(元)； X_5 ：建筑业平均工资(元)； X_6 ：交通运输、仓储和邮政业平均工资(元)； X_7 ：信息传输、计算机服务和软件业平均工资(元)； X_8 ：批发和零售业平均工资(元)； X_9 ：住宿和餐饮业平均工资(元)； X_{10} ：金融业平均工资(元)； X_{11} ：房地产业平均工资(元)； X_{12} ：教育平均工资(元)。数据如表 1 所示。

Table 1. Average wages of employed persons in urban units by region, 2021
表 1. 2021 年分地区城镇单位就业人员平均工资

地区	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
北京	77,376	152,773	169,210	198,210	146,308	134,361	290,038	180,062	63,642	298,200	131,866	201,129
天津	75,125	158,580	109,826	175,282	104,859	114,432	157,725	105,605	41,251	155,286	98,668	149,064
河北	60,706	95,838	80,289	115,727	69,392	97,698	132,218	63,068	45,346	95,401	71,833	86,608
山西	52,618	99,948	74,758	101,101	74,724	101,149	99,130	73,526	39,400	87,734	68,728	80,472
内蒙	71,429	142,301	91,047	120,748	60,880	104,584	113,337	79,896	48,506	97,446	59,438	91,871
辽宁	24,231	92,311	83,631	92,673	69,565	94,245	121,947	71,675	39,671	102,613	71,733	96,709
吉林	53,192	82,233	93,088	100,559	64,525	86,901	93,158	70,055	44,784	84,244	60,278	88,274
黑龙江	40,996	98,793	81,878	92,695	62,166	93,288	93,945	77,983	41,197	74,150	53,953	92,853
上海	86,572	326,402	152,248	246,387	137,248	157,455	303,573	203,844	63,409	397,655	126,648	199,498

续表

江苏	54,841	118,302	103,619	161,610	78,114	110,171	180,782	107,279	56,449	164,177	95,203	149,417
浙江	80,806	99,510	97,735	168,584	75,792	121,968	257,631	120,042	61,543	175,773	106,292	151,272
安徽	55,304	123,744	84,197	130,890	72,577	94,979	111,935	75,833	52,830	110,117	82,975	112,889
福建	69,504	67,457	86,707	134,343	73,221	109,125	143,350	97,427	50,328	131,573	91,617	111,035
江西	54,194	73,052	71,719	95,553	65,957	95,206	104,940	69,612	43,963	106,042	75,983	92,233
山东	66,054	113,104	83,162	122,998	76,278	106,536	116,084	75,628	53,297	97,701	81,054	109,680
河南	52,304	86,525	64,495	100,153	61,606	87,331	91,501	64,891	45,984	125,279	69,174	78,468
湖北	51,621	111,974	84,225	132,010	78,880	103,342	131,663	73,330	44,962	127,544	75,549	103,170
湖南	58,199	68,983	82,484	106,637	62,179	96,132	117,793	70,829	43,349	108,042	77,348	85,103
广东	77,837	187,947	92,303	164,727	83,706	121,907	213,031	104,679	54,516	202,771	100,580	136,671
广西	70,545	75,850	76,986	115,644	71,797	101,606	111,988	76,768	43,101	112,750	85,939	85,427
海南	53,585	136,604	86,122	122,836	66,994	117,052	247,110	99,841	64,328	113,934	83,909	99,027
重庆	73,140	96,743	89,966	110,863	67,918	97,665	155,067	87,988	48,566	129,860	87,315	127,699
四川	75,541	119,846	88,490	123,842	66,936	103,073	147,727	83,512	50,888	113,147	78,956	105,197
贵州	57,459	87,356	92,949	125,216	80,317	101,505	118,727	91,019	49,925	142,097	77,398	95,667
云南	53,253	866,87	87,341	119,044	70,694	107,484	111,243	90,795	45,886	144,635	81,086	108,775
西藏	66,674	127,699	88,355	126,469	71,241	131,777	179,818	96,498	67,630	245,574	95,252	164,731
陕西	63,090	120,683	86,840	118,877	78,757	99,476	192,699	73,015	46,113	112,682	74,182	88,927
甘肃	62,699	111,348	83,374	96,962	62,289	96,585	93,005	65,574	43,082	86,769	53,377	95,211
青海	47,609	152,682	89,234	129,072	91,110	115,229	137,031	82,010	52,375	144,717	60,770	120,726
宁夏	57,826	160,833	84,879	136,010	78,574	96,108	130,972	64,886	48,210	112,052	70,276	106,735
新疆	56,131	160,056	88,422	123,101	81,501	114,087	119,526	87,274	50,496	132,160	62,481	94,580

2.2. 聚类分析

聚类分析[5]是一种无监督的学习算法。聚类分析将数据集中的对象分为若干个类，使得同一类内的对象彼此相似，而不同类的对象之间差异较大，其核心原理在于使类内的相似性或距离最小化，类间距离最大化。聚类分析可以对变量进行分类(通常采用 R 型聚类方法)，也可以对样品进行分类(通常采用 Q 型聚类分析)，本文采用的是系统聚类法。

2.2.1. R 型系统聚类

现在，让我们对表 1 的数据进行聚类分析。首先，我们猜测选取的这九项指标可能存在一定的线性关系，于是想到尝试将相关性强的指标归并到一起，首先采用 R 型聚类，将行业作为变量进行系统聚类得到如图 1 的聚类结果。

根据图 1，可以将 12 个行业分为四类。

第一类中包含两个高收入水平行业。首先是金融业，这一行业长久以来都被视为职业发展的黄金领域，随着中国金融市场高水平开放持续推进，其发展前景愈发广阔。由于金融行业的特殊性和复杂性，对高端人才的需求尤为突出，尤其是在基金、证券等高端市场中，对人才的专业素养和综合能力有着极高的要求，因此提供的薪酬水平也相应较高。同时，信息传输、计算机服务和软件业也在信息化时代的

大潮中崭露头角。随着“人工智能”、“网络工程”、“软件开发”等前沿技术的迅猛发展，这些行业正逐渐成为推动社会进步和经济发展的重要力量。在这一过程中，对人才的需求量也呈现出爆炸式增长。因此，这些行业也提供了相对较高的薪酬水平，以吸引和激励更多优秀人才投身于其中。

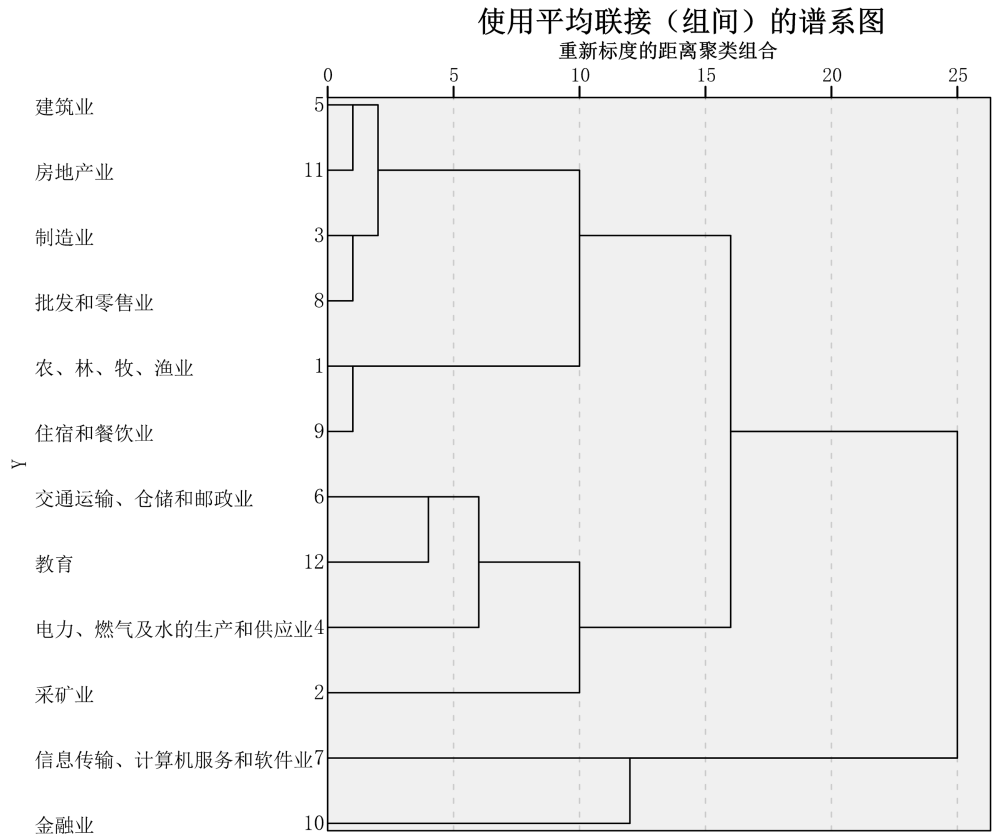


Figure 1. Cluster chart of average wages by industry
图 1. 行业平均工资聚类图

第二类中的行业多数属于传统的较为稳定且薪酬较高的行业，随着全球化和技术革新的浪潮，这些行业也面临着转型与升级的重要机遇。电力、燃气及水的生产和供应业，作为国民经济的重要支柱，目前正处于从传统向现代的转变之中，需要从传统的高耗能、高污染的模式向清洁、高效、绿色能源模式转变，从发展前景来看，这些行业无疑具有巨大的潜力和机遇。教育行业作为另一个传统热门行业，其重要性不言而喻。随着社会的不断进步和科技的飞速发展，各行各业对人才的需求日益增加，对教育的要求也越来越高。因此，教育行业需要投入更多的资源和力量，培养更多的人才。交通运输、仓储和邮政业得益于网络时代的发展，快递、外卖等业务兴起，工薪待遇也较为不错。

第三类行业普遍具有劳动密集型特征，导致对基础劳动力的高度依赖，这在一定程度上限制了整体工资水平的提升。同时，这些行业面临着激烈的市场竞争，企业为保持竞争力而严格控制成本，包括人力成本等。与金融业和信息传输、计算机服务和软件业等高薪行业相比，第三类别行业在技术革新和产业升级方面的步伐相对较慢，难以支撑起更高的薪资水平。

由于信息化时代的发展，农林牧渔业、住宿和餐饮业等传统领域，由于长期以来依赖较为传统的经营模式和较低的自动化、信息化水平，难以迅速适应市场的新变化，面临收入水平相对较低的困境，这也进一步削弱了它们对年轻一代劳动力的吸引力，从而导致传统行业人才流失加剧，创新能力受限。

2.2.2. Q 型系统聚类

接下来，采用 Q 型聚类，将省市作为变量进行系统聚类，分析比较所得到的结果。根据图 2，可以将 31 个省市分为四类。

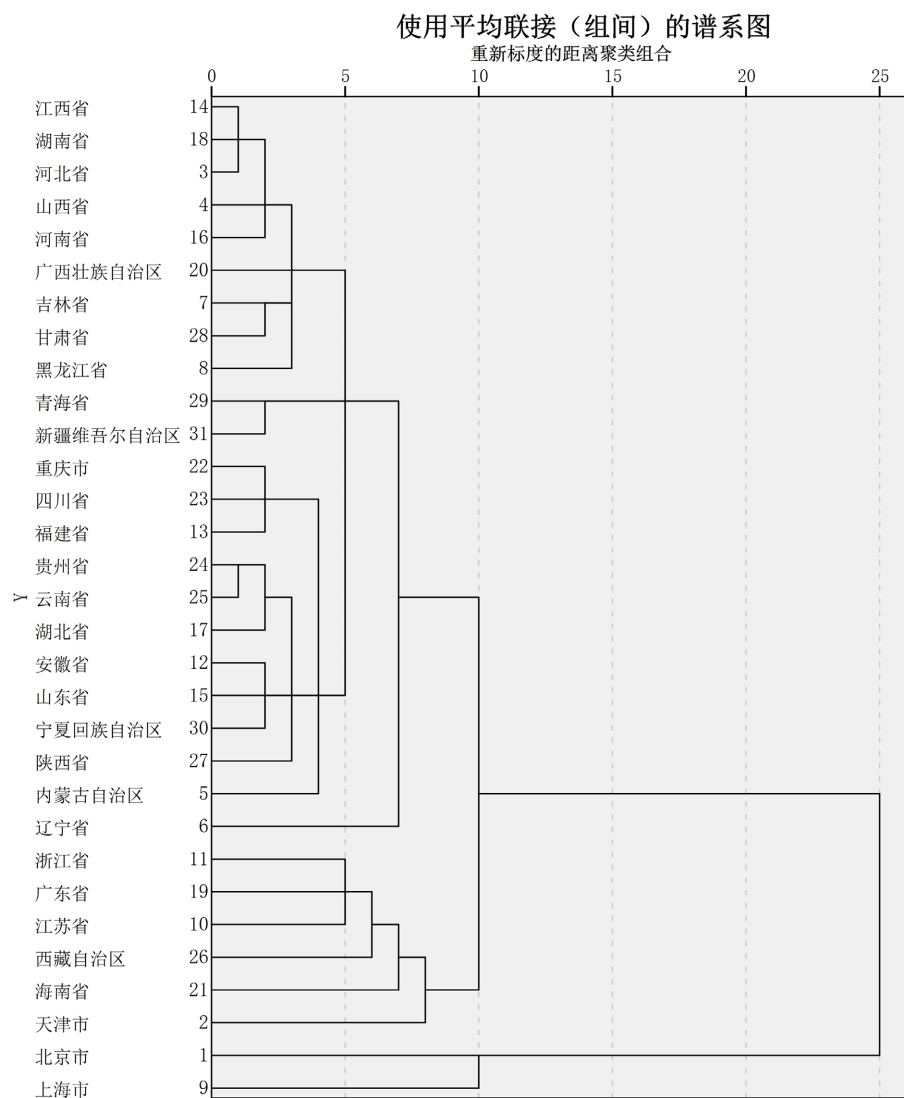


Figure 2. Genealogy diagram of the connection method between provincial average wage groups

图 2. 省份平均工资组间联接法谱系图

北京属于第一类高收入城市，年平均工资始终稳居全国前列。作为国家的政治中心、文化中心、国际交流中心以及科技创新中心，北京不仅吸引了大量的国内外投资，还汇聚了众多优秀人才，共同推动了北京经济的繁荣与发展。作为科技创新的核心城市，北京拥有众多的高等院校、科研机构和科技园区。这些机构和企业吸引了大量的人才进行研发和创新，为北京市带来了丰厚的收入。众多贸易公司、跨国企业和外国驻华机构在北京设立办事处，与国际市场进行紧密的贸易往来，为北京市带来了大量的外贸收入。北京的平均工资水平较高，特别是金融、信息技术等高收入行业，吸引了大量的人才聚集。这些人才的高收入也带动了整个城市的消费水平，进一步推动了经济的发展。除此之外，北京市的最低工资标准也高于其他地区，这就意味着北京低收入群体的平均工资较其他地区也有一定的优势。

上海属于第二类高收入城市。作为中国的金融中心，上海的金融市场繁荣活跃，吸引了全球投资者的目光。除了金融业外，上海在服务业、高新技术产业等领域也取得了显著成就。上海地处长江入海口，海陆位置优越，使得上海成为连接内陆与海洋的重要枢纽，为其经济发展提供了便利条件。同时，上海也是中国对外开放的重要窗口之一，其开放包容的城市形象吸引了大量外资企业的入驻。因此，由于金融业的巨大贡献、多元化的经济结构、独特的地理位置等因素共同作用，使得上海成为了一个经济繁荣、收入水平高的地区。

江苏、浙江、天津、广东、西藏、海南作为第三类，平均收入水平仅次于北京、上海，这些地区普遍拥有较高的经济发展水平，江苏、浙江、广东作为中国东部沿海的经济大省，凭借其优越的地理位置、丰富的自然资源和良好的工业基础，实现了经济的快速增长。天津作为直辖市，也拥有雄厚的工业实力和多元化的经济结构。而西藏和海南近年来在国家政策的扶持下，经济增长速度显著加快，特别是在旅游业和特色产业发展方面取得了显著成效。这些地区展现出强劲的经济活力和良好的发展态势。

其他省份为第四类，其经济发展水平相对较低，经济增速可能也相对较慢，这直接影响到居民的收入水平。同时，就业市场供需失衡，高技能岗位稀缺，导致平均工资受到限制。此外，政策与制度环境的不完善，也影响了居民的实际可支配收入。因此，要提高这些省份的平均工资水平，需要综合考虑并协同推进产业升级、就业创造、政策优化、教育提升、资源利用和社会稳定等多方面的措施。

2.3. 判别分析

判别分析[6]是在分类确定的条件下，旨在根据某一研究对象的各种特征值判别其类型归属问题。常用的判别分析方法有距离判别、Fisher 判别、贝叶斯判别，本文采用的是 Fisher 判别法。接下来根据前面聚类得到的结果，使用判别分析验证分类的合理性。

2.3.1. 判别函数

在判别分析之前，我们需要建立标准化的判断函数来计算每个观测的得分，并根据观测得分对观测进行分类。

Table 2. Eigenvalue
表 2. 特征值

函数	特征值	方差百分比	累积百分比	典型相关性
1	25.643a	78.62201	78.62201	0.981054
2	5.170a	15.85021	94.47222	0.915377
3	1.803a	5.527778	100	0.802015

表 2 反映了判别函数的特征根、解释方法的比例和典型相关系数。三个判别函数解释了全部方差。因此判别时，使用三个判别函数进行预判。

运用 12 个行业就业人员平均工资对各省份进行判别分析，建立了 3 个判别函数。表 3 显示的是标准化的判别函数，表示为

$$\begin{aligned} y_1 &= 0.038489X_1 + 0.508829X_2 + 0.449119X_3 - 0.31373X_4 + 0.024739X_5 - 0.55167X_6 \\ &\quad + 0.095556X_7 + 0.626669X_8 - 0.39603X_9 + 0.867744X_{10} + 0.210612X_{11} - 0.00068X_{12}, \\ y_2 &= 0.326125X_1 - 0.32086X_2 + 0.750166X_3 - 1.43609X_4 + 1.477894X_5 - 0.56218X_6 \\ &\quad - 0.16857X_7 - 0.01X_8 + 1.290252X_9 - 0.25639X_{10} + 0.28746X_{11} - 0.62486X_{12} \end{aligned}$$

$$y_3 = -0.02043X_1 - 0.88038X_2 + 0.199166X_3 + 0.335282X_4 + 0.000745X_5 + 0.539294X_6 + 0.44818X_7 - 0.87725X_8 - 0.12163X_9 - 0.29408X_{10} - 0.01794X_{11} + 0.859253X_{12}。$$

我们可以根据这个判断函数来计算每个观测的得分，并根据观测得分对观测进行分类。

Table 3. Standardized canonical discriminant function coefficients
表 3. 标准化典则判别函数系数

	1	2	3
农、林、牧、渔业	0.038489	0.326125	-0.02043
采矿业	0.508829	15.85021	94.47222
制造业	0.449119	0.750166	0.199166
电力、燃气及水的生产和供应业	-0.31373	-1.43609	0.335282
建筑业	0.024739	1.477894	0.000745
交通运输、仓储和邮政业	-0.55167	-0.56218	0.539294
信息传输、计算机服务和软件业	0.095556	-0.16857	0.44818
批发和零售业	0.626669	-0.01	-0.87725
住宿和餐饮业	-0.39603	1.290252	-0.12163
金融业	0.867744	-0.25639	-0.29408
房地产业	0.210612	0.28746	-0.01794
教育	-0.00068	-0.62486	0.859253

2.3.2. 分类结果

分类结果(图 3)显示，省份分为第一类的共 1 个，经过判别分析后，省份分为第一类的为 1 个，判别正确率为 100.00%；分为第二类的共有 5 个，经判别分类后，分为第二类的共有 5 个，判别正确率为 100.00%；实际分为第三类的共 24 个，经过判别分析后，省份分为第三类的为 24 个，判别正确率为 100.00%；实际分为第四类的共 1 个，经过判别分析后，省份分为第四类的为 1 个，判别正确率为 100.00%；4 级判别的正确率为 100.00%，因此判别函数的总体判别正确率为 100.00%。

分类结果 ^a						
Average Linkage (Between Groups)		预测组成员信息				总计
		1	2	3	4	
计数	1	1	0	0	0	1
	2	0	5	0	0	5
	3	0	0	24	0	24
	4	0	0	0	1	1
原始	1	100.0	0.0	0.0	0.0	100.0
	2	0.0	100.0	0.0	0.0	100.0
	3	0.0	0.0	100.0	0.0	100.0
	4	0.0	0.0	0.0	100.0	100.0

a. 正确地 对 100.0% 个原始已分组个案进行了分类。

Figure 3. Categorize the results
图 3. 分类结果

从图 4 中可以看出,通过判别函数预测,四组地区分类全部被判对,整体准确率 100%。因此,系统聚类与费歇尔判别具有较好的一致性。

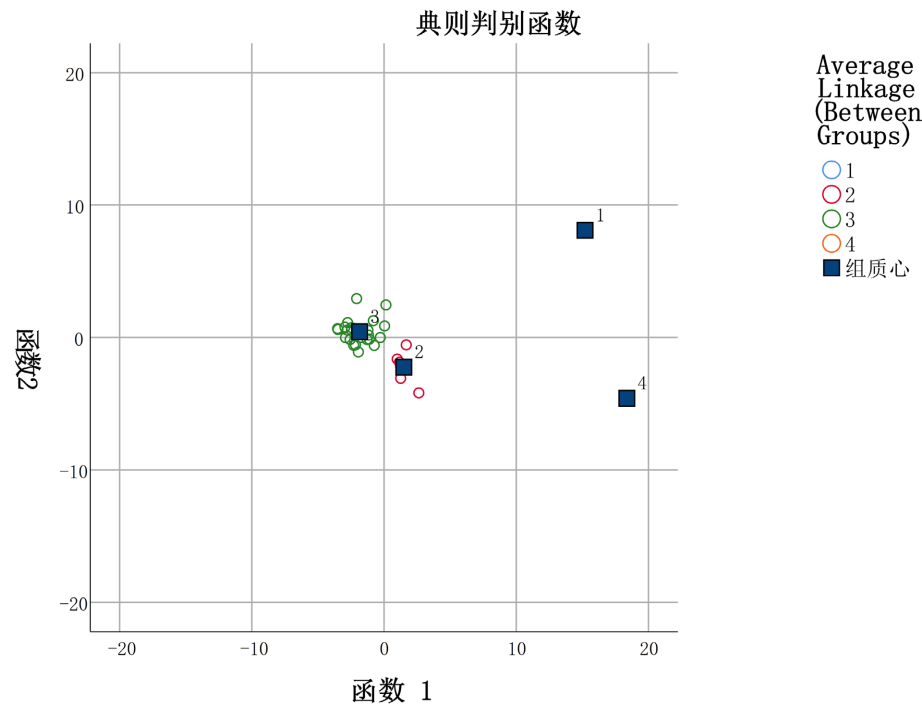


Figure 4. Classification result graph
图 4. 分类结果图

3. 结论与建议

本文采用聚类分析法和判别分析法,重点研究了中国不同地区与行业间工资水平的差别。从分类结果来看,基于平均工资水平的地区划分与地理上的东部、中部和西部有着很高的一致性,不同地区因资源禀赋、产业结构、地理位置等因素的差异,导致经济发展水平不平衡,表明了地理环境因素对经济发展的重要影响;同时,政府的政策导向和资源配置也会影响地区的工资水平。在就业选择层面,若以工资收入作为核心考虑因素,北京、上海无疑是求职者的首选就业城市,江苏、浙江、广东等因相对较高的收入水平成为求职者选择的热门城市。当然在做出职业选择时,也要考虑到自身专业背景、兴趣方向和行业前景,基于此,求职者在综合考虑个人条件的基础上,可优先考虑金融业、信息与计算机相关行业以及水电气行业等薪酬相对较高且前景看好的行业。

参考文献

- [1] 李小春, 夏小文. 各地区职工平均工资的多元统计分析[J]. 科协论坛(下半月), 2010(11): 127-129.
- [2] 刘辉, 柳炳祥. 一种分地区、行业平均工资的聚类分析方法[J]. 科技创新导报, 2013(3): 237-238.
- [3] 朱轩放. 主成分分析和聚类在辽宁省城市职工工资分析中的应用[D]: [硕士学位论文]. 长春: 吉林大学, 2016.
- [4] 张子悦. 基于聚类分析视角浅析我国年均工资水平行业差异[J]. 西部皮革, 2018, 40(17): 97-98.
- [5] 刘茗, 赵华, 徐泽水. 基于系统聚类法的中国不同地区和行业工资水平分析[J]. 数学的时间与认识, 2023, 53(2): 277-283.
- [6] 何晓群. 多元统计分析[M]. 第 5 版. 北京: 中国人民大学出版社, 2019.