

基于主成分分析与系统聚类分析的数字化企业经济绩效研究

姜兴明

贵州大学管理学院, 贵州 贵阳

收稿日期: 2024年8月15日; 录用日期: 2024年9月9日; 发布日期: 2024年11月28日

摘要

随着数字化转型的深入, 企业经济绩效的评估与提升成为重要议题。本研究旨在探讨数字化企业经济绩效的评估方法, 通过主成分分析(PCA)和系统聚类分析, 构建了一个综合性的研究框架。选取50家中国上市数字化企业为样本, 基于其2022年的财务数据, 运用PCA降维提取关键绩效指标, 并通过系统聚类分析对企业进行分类和比较。研究发现, 四个主成分累积解释了83.664%的方差, 有效地代表了原始数据集中的主要信息。系统聚类分析揭示了企业间经济绩效的显著差异, 其中上海钢联电子商务股份有限公司表现最为突出, 而中信尼雅葡萄酒股份有限公司则需要进一步分析其绩效较低的原因。本研究为数字化企业经济绩效的评估和提升提供了新的视角和方法, 同时指出未来研究可以扩展到非财务指标和全球范围的企业, 以增强研究的普遍性和适用性。

关键词

数字化企业, 经济绩效, 主成分分析, 系统聚类分析, 财务指标

A Study on Digital Enterprise Economic Performance Based on PCA and Hierarchical Clustering Analysis

Xingming Jiang

School of Management, Guizhou University, Guiyang Guizhou

Received: Aug. 15th, 2024; accepted: Sep. 9th, 2024; published: Nov. 28th, 2024

Abstract

With the deepening of digital transformation, the evaluation and improvement of enterprise economic performance have become important issues. This study aims to explore the evaluation method of

digital enterprise economic performance and constructs a comprehensive research framework through principal component analysis (PCA) and hierarchical clustering analysis. Fifty Chinese listed digital enterprises were selected as samples, and key performance indicators were extracted by PCA dimensionality reduction based on their financial data in 2022. The enterprises were classified and compared by hierarchical clustering analysis. The study found that four principal components explained 83.664% of the variance, effectively representing the main information in the original dataset. Hierarchical clustering analysis revealed significant differences in economic performance among enterprises, with Shanghai Ganglian E-commerce Co., Ltd. performing most prominently, while CITIC Ninya Wine Co., Ltd. needs further analysis of the reasons for its low performance. This study provides new perspectives and methods for the evaluation and improvement of digital enterprise economic performance, and points out that future research can be expanded to non-financial indicators and global enterprises to enhance the universality and applicability of the research.

Keywords

Digital Enterprise, Economic Performance, Principal Component Analysis, Hierarchical Clustering Analysis, Financial Indicators

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

随着信息技术的飞速发展和全球经济一体化的不断推进,数字化转型已经成为企业提升竞争力、实现可持续发展的关键途径[1]。在这一背景下,对数字化企业经济绩效的研究显得尤为重要,它不仅有助于企业了解自身的经营状况,更为企业制定战略决策提供了科学依据[2]。然而,数字化企业经济绩效的多维度特性和复杂性使得对其进行准确评估和深入分析成为一大挑战[3]。

主成分分析(PCA)和系统聚类分析作为两种强大的数据处理技术,为解决上述问题提供了可能。主成分分析通过降维技术,能够从多个相关变量中提取出少数几个代表性的主成分,从而简化数据结构,揭示数据的内在结构和规律[4]。系统聚类分析则是一种基于系统思想的聚类方法,它能够根据企业间的经济绩效差异,将企业划分为不同的类别,以发现企业间的相似性和差异性[5]。

本文旨在通过结合主成分分析和系统聚类分析两种方法,构建一个综合性的研究框架,对数字化企业的经济绩效进行深入研究。首先,本文将介绍数字化企业经济绩效的相关概念及其重要性,然后详细阐述主成分分析和系统聚类分析的理论基础和实施步骤。在此基础上,本文将选取具有代表性的数字化企业作为研究对象,收集并处理相关数据,运用主成分分析提取关键绩效指标,并通过系统聚类分析对企业进行分类和比较。最后,本文将通过案例分析和实证研究,验证所提出的方法和结论,并总结研究成果及其对企业数字化转型的启示。

通过本研究,我们期望能够为数字化企业经济绩效的评估和提升提供新的视角和方法,帮助企业更好地理解数字化转型对经济绩效的影响,从而为相关决策提供科学依据。同时,本研究也试图为数字化经济领域的学术研究贡献新的理论和实证分析,推动该领域的理论发展和实践应用。

2. 文献综述

数字化企业经济绩效的研究是近年来管理科学和信息系统领域的热点之一[6]。随着数字化转型在全球范围内的深入发展,学者们开始关注企业如何通过数字化手段提升其经济绩效。陈德球等[7]在其开创

性的工作中指出, 数字化企业能够通过重新定义产品、服务和市场来获得竞争优势。此后, 大量研究开始探讨数字化转型与企业绩效之间的关系。

在评估企业经济绩效方面, 主成分分析(PCA)作为一种有效的多变量统计技术, 已被广泛应用于降维和数据简化[8]。吴海建[9]认为 PCA 通过提取数据中的主要成分来揭示潜在的模式和关系。这能够帮助研究者集中关注对绩效评估最为关键的指标。然而, PCA 的应用并非没有争议, 李尔国等[10]学者指出, PCA 可能忽视数据中的非线性关系和复杂的相互作用。

系统聚类分析作为一种分类技术, 能够帮助研究者识别企业间的相似性和差异性, 从而对企业进行有效的分类。刘秋彤等[11]指出系统聚类方法在企业绩效研究中的应用, 使得研究者能够根据企业的经济绩效特征将其分为不同的群体, 以进行更深入的比较和分析。尽管如此, 系统聚类分析在处理大规模数据集时可能会遇到挑战, 需要研究者具备一定的专业知识来选择合适的聚类算法和参数设置。

结合 PCA 和系统聚类分析的研究方法在数字化企业经济绩效评估中的应用还相对较少。一些先行研究者何远霞等[12]尝试将这两种方法结合起来, 以期获得更准确的分析结果。然而, 这些研究大多集中在特定的行业或地区, 缺乏对不同类型和规模企业的全面考察。

综上所述, 尽管已有研究为我们提供了宝贵的理论和实证基础, 但在数字化企业经济绩效评估领域, 特别是在结合 PCA 和系统聚类分析方法的应用上, 仍有较大的研究空间。本文将填补这一空白, 通过实证研究探讨这两种方法在评估数字化企业经济绩效方面的有效性和可行性。

3. 多元统计分析方法

3.1. 定义

主成分分析(Principal Component Analysis, PCA)是一种统计技术, 用于数据降维和模式识别。通过 PCA, 我们可以将一组可能存在相关性的变量转换为一组线性不相关的变量, 这些新变量被称为主成分[13]。这些主成分按照方差递减的顺序排列, 第一个主成分具有最大的方差, 每个后续的主成分都有尽可能最大的剩余方差。具体来说, 主成分分析的目标是找到一个新的坐标系统, 使得在这个坐标系统中, 数据的方差最大化, 同时保持数据的内在结构。

系统聚类分析(Systematic Clustering Analysis), 也称为层次聚类分析(Hierarchical Clustering Analysis), 是一种常用的数据分类技术, 用于将数据集中的对象根据它们的特征分成不同的组或“簇”[14]。系统聚类分析的核心思想是根据对象间的相似性或距离来逐层构建一个嵌套的簇结构, 从而形成一个树状图(称为树状图或聚类图)。

3.2. 主成分分析基本原理

假设有 m 个样本, 每个样本有 n 个指标, 那么可以将这 m 个样本在 n 维空间中表示出来。如果 n 个指标之间存在相关性, 那么这些指标的变化不是独立的, 而是相互关联的, 因此可以使用主成分分析将这些指标转换为一组新的、相互独立的指标, 即主成分[15]。主成分是原始指标的线性组合, 第一个主成分包含了原始数据中最多的方差, 第二个主成分包含了次多的方差, 以此类推。具体确定各个主成分的方法如下:

设 Z_i 表示第 i 个主成分, $i=1, 2, \dots, p$, 可设

$$\begin{cases} Z_1 = c_{11}X_1 + c_{12}X_2 + \dots + c_{1p}X_p \\ Z_2 = c_{21}X_1 + c_{22}X_2 + \dots + c_{2p}X_p \\ \vdots \\ Z_p = c_{p1}X_1 + c_{p2}X_2 + \dots + c_{pp}X_p \end{cases} \quad (1)$$

其中对每一个 i ，均有 $2_c^{i1} + 2_c^{i2} + \dots + 2_c^{ip} = 1$ 。

主成分分析基本步骤

(1) 标准化数据

首先需要对原始数据进行标准化处理，使得所有指标具有相同的尺度和权重。假设有 m 个样本，每个样本有 n 个指标，那么可以将原始数据表示为一个 $m \times n$ 的矩阵 X ，其中 X_{ij} 表示第 i 个样本在第 j 个指标上的取值。标准化的公式如下：

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (2)$$

其中， \bar{x}_j 和 s_j 分别表示第 j 个指标的平均值和标准差。

(2) 计算协方差矩阵

通过矩阵运算，计算出原始数据的协方差矩阵 S 。协方差矩阵反映了不同指标之间的相关性程度，对角线上的元素是每个指标的方差，非对角线上的元素是每对指标之间的协方差。协方差矩阵的公式如下：

$$S = \frac{1}{m-1} \sum_{i=1}^m (z_i - \bar{z})(z_i - \bar{z})^T \quad (3)$$

其中， z_i 是第 i 个样本的标准化后的数据向量， \bar{z} 是所有样本的标准化数据的平均值。

(3) 计算特征值和特征向量

通过特征值分解，求出协方差矩阵的特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 和对应的特征向量 V_1, V_2, \dots, V_n 。特征向量是一个 n 维列向量，表示协方差矩阵的某个特征值所对应的特征向量。特征值表示特征向量的重要程度，越大的特征值表示该特征向量代表的因素对数据影响越大。特征值和特征向量的计算公式如下：

$$Sv_i = \lambda_i v_i, \quad i = 1, 2, \dots, n \quad (4)$$

其中， S 是协方差矩阵， v_i 是第 i 个特征向量， λ_i 是第 i 个特征值。

(4) 选择主成分

根据特征值的大小，对特征向量进行排序，选取前 k 个特征向量作为主成分，其中 k 是预设的降维后的维数。通常，可以选取特征值较大的前若干个主成分，使其累计贡献率达到一定的阈值，例如 70%、80% 或 90% 等。

(5) 计算主成分得分

根据选定的主成分，计算每个样本的主成分得分，即每个样本在每个主成分上的投影值。主成分得分是描述样本在低维度空间中位置的重要指标，可以用于后续聚类、分类或回归等分析。计算公式如下：

$$y_{ij} = T_v^i z_j \quad (5)$$

其中， y_{ij} 是第 j 个样本在第 i 个主成分上的得分， v_i 是第 i 个主成分的特征向量， z_j 是第 j 个样本的标准化数据向量。

3.3. 系统聚类分析基本步骤

(1) 初始化

开始时，每个数据点被视为一个单独的簇。对于包含 n 个数据点的数据集，初始时会有 n 个簇。

(2) 计算距离矩阵

确定一个距离度量(如欧几里得距离、曼哈顿距离或余弦相似性等)来计算所有可能的簇对之间的距离。这个距离矩阵将用于后续的簇合并决策。

(3) 寻找最近的簇对

在当前的簇划分基础上，找到距离最近的两个簇，以便将它们合并成一个新的簇。合并的依据可以采用不同的策略，如最小距离法(single-linkage)、最大距离法(complete-linkage)、平均距离法(average-linkage)等。

(4) 合并簇

根据步骤 3 中确定的最近的簇对，将它们合并成一个新的簇，并更新距离矩阵以反映新簇与其他簇之间的距离。

(5) 重复合并过程

重复步骤 3 和步骤 4，每次都合并距离最近的簇对，直到所有的数据点都被合并到一个单一的簇中，或者达到了某个预定的簇数量。

(6) 构建树状图

系统聚类的结果是一棵树状图(dendrogram)，它展示了数据点如何被逐步合并成簇。树状图的每个节点代表一个簇，而节点之间的连线表示簇之间的合并。

(7) 确定最终的簇划分

根据树状图和研究目的，选择一个合适的高度来划分簇。这个高度决定了簇的数量和大小。在实际应用中，研究者可能会基于领域知识或使用一些准则(如树状图的截断)来确定最终的簇划分。

4. 数据准备及分析

4.1. 样本数据来源

在这篇文章中，我们挑选了同花顺数据中心的数字化理念部分的上市企业作为研究对象，经过筛选出数据不足的企业以及做出权衡处理，确定了 50 家企业。数据来源于各个企业的年度报告和同花顺数据中心。

本文按照各个企业的年度报告和同花顺数据中心所公布的数据，选取 2022 年的数据作为观测样本进行分析，所选指标包括应付职工薪酬、营业总成本、固定资产总计、营业总收入、营业总收入同比增长率、总资产周转率、资产负债率、净资产收益率，运用 SPSS 软件中的数据处理方法对该数据进行处理。

4.2. 主成分分析

主成分分析(PCA)是一种数据降维技术，它通过识别数据中的模式来减少数据集的维度，同时保留最重要的变量。这种方法通过提取数据的核心特征，有效地减少了数据的复杂性，同时尽可能地保留了原始数据集中的信息。本文对我国 50 个数字化企业的相关指标进行了主成分分析，分析结果见表 1：

Table 1. Correlation coefficient matrix

表 1. 相关系数矩阵

	应付职工薪酬	固定资产总计	营业总成本	净资产收益率	营业总收入同比增长率	营业总收入	资产负债率	总资产周转率
应付职工薪酬	1	0.516	0.212	0.454	0.145	0.283	0.397	0.223
固定资产总计	0.516	1	0.242	0.241	0.062	0.304	0.187	-0.072
营业总成本	0.212	0.242	1	0.273	-0.020	0.972	0.166	0.145
净资产收益率	0.454	0.241	0.273	1	0.472	0.384	0.181	0.338

续表

营业总收入同比增长率	0.145	0.062	-0.020	0.472	1	0.039	0.090	0.253
营业总收入	0.283	0.304	0.972	0.384	0.039	1	0.169	0.167
资产负债率	0.397	0.187	0.166	0.181	0.090	0.169	1	0.432
总资产周转率	0.223	-0.072	0.145	0.338	0.253	0.167	0.432	1

营业总成本与营业总收入同比增长率的相关系数为-0.020，这表明两者之间几乎没有线性关系。这可能意味着公司的营业成本变化对收入增长率的影响不大，或者这种影响在数据中不明显。净资产收益率与营业总收入的相关系数为 0.384，这是一个中等程度的正相关，表明公司的净资产收益率与营业总收入之间存在一定的正向关系。这可能意味着公司的盈利能力与其营业总收入有一定的关联。资产负债率与总资产周转率的相关系数为 0.432，这是一个中等程度的正相关，表明公司的负债水平与管理效率(通过总资产周转率反映)之间存在一定的正向关系。这可能意味着公司的负债管理对其运营效率有一定的正面影响。

Table 2. KMO and Bartlett tests
表 2. KMO 和巴特利特检验

KMO 和 Bartlett 的检验		
Bartlett 球形度检验	KMO 值	0.583
	近似卡方	218.823
	df	28
	p 值	0.000

主成分分析探索定量数据可以浓缩为几个方面(主成分)，通常用于权重计算；

第一：分析 KMO 值，KMO 值为 0.58，这表明数据比较适合进行因子分析，但并不是非常理想。在实际应用中，如果 KMO 值在 0.6 以上，通常被认为是可接受的，可以进行主成分分析(PCA)。然而，KMO 值只是决定是否进行 PCA 的多个因素之一。如果其他条件(如 Bartlett 的球形度检验)满足要求，并且认为进行 PCA 对于研究目标是有意义的，那么即使 KMO 值略低于 0.6，也可以继续进行分析。在这种情况下，即使 KMO 值为 0.58，但 Bartlett 的球形度检验的 p 值小于 0.05，这表明数据的协方差矩阵与单位矩阵显著不同，适合进行主成分分析。

第二：如果 Bartlett 检验对应 p 值小于 0.05 也说明适合进行主成分分析；

第三：如果仅两个分析项，则 KMO 无论如何均为 0.5。

由表 2 可知，KMO 值和 p 值均满足要求，因此，该数据做主成分分析是有价值的。

MSA 指标 = A/(A + B)，A 为该项与其余项相关系数平方和，B 为该项与其余项偏相关系数平方和；MSA 指标的意义为某项与其余各项的相关关系情况，该值介于[0, 1]之间；如果 MSA 指标值过低(比如小于 0.2)，意味着其可能对信息浓缩帮助较小，可以考虑对其进行删除，以提高 KMO 值表现等；如果 MSA 指标值过高(比如大于 0.8)，意味着该项与其余项的信息重叠性过高，可以考虑对其进行删除后再次分析。由表 3 可知，MSA 值均满足要求。

表 4 展示了主成分提取的详细情况以及它们所包含的信息量。根据表 4 的分析结果，可以得知，通过主成分分析(PCA)共识别并提取了 4 个主要的成分。每个主成分的特征根值都超过了 1，这表明它们都具有显著的统计意义。此外，这 4 个主成分的方差解释率也得到了具体的计算和描述，此 4 个主成分的

方差解释率分别是 36.476%，19.286%，14.605%，13.296%，累积方差解释率为 83.664%。另外，本次分析共提取出 4 个主成分，它们对应的加权后方差解释率它们对应的加权后方差解释率即权重依次为： $36.476/83.664 = 43.60\%$ ； $19.286/83.664 = 23.05\%$ ； $14.605/83.664 = 17.46\%$ ； $13.296/83.664 = 15.89\%$ 。

Table 3. MSA measurement scale
表 3. MSA 测量表

项	MSA 指标
整体	0.583
应付职工薪酬	0.694
固定资产总计	0.623
营业总成本	0.512
净资产收益率	0.601
营业总收入同比增长率	0.625
营业总收入	0.531
资产负债率	0.634
总资产周转率	0.633

Table 4. Eigenvalue and variance contribution rate table
表 4. 特征值与方差贡献率表

编号	特征根			主成分提取		
	特征根	方差解释率%	累积%	特征根	方差解释率%	累积%
1	2.918	36.476	36.476	2.918	36.476	36.476
2	1.543	19.286	55.762	1.543	19.286	55.762
3	1.168	14.605	70.368	1.168	14.605	70.368
4	1.064	13.296	83.664	1.064	13.296	83.664
5	0.530	6.622	90.286	-	-	-
6	0.427	5.335	95.621	-	-	-
7	0.331	4.136	99.757	-	-	-
8	0.019	0.243	100.000	-	-	-

运用 SPSS 画出碎石图，如图 1 所示。

图 1 给出了不同成分数下的特征值，在碎石图中，通常会寻找一个“肘点”，在这个点之前，方差贡献的下降速度较快，而在“肘点”之后，下降速度明显变缓。根据数据，我们可以看到在成分数量 3 到 4 之间，方差贡献的下降速度有所减缓，这可能暗示着在这个点之前提取的主成分已经捕捉了大部分的变异性。

综上分析可知，本次主成分分析应该提取四个主成分。

表 5 详细呈现了主成分分析(PCA)在提取研究项信息方面的成效，以及各主成分与研究项之间的对应关系。根据表 5 的数据，可以观察到所有研究项的共同度值都超过了 0.4 的阈值，这表明了研究项与主成分之间存在显著的正相关性，从而确保了主成分能够有效地捕捉到研究项的关键信息。在确认了主

成分能够充分提取研究项信息的基础上，接下来需要进一步分析主成分与研究项之间的具体对应情况。当载荷系数的绝对值达到或超过 0.4 时，即可认为该研究项与相应的主成分存在显著的对应关系。通过这种分析，可以更深入地理解主成分分析如何帮助我们揭示数据背后的结构和模式。

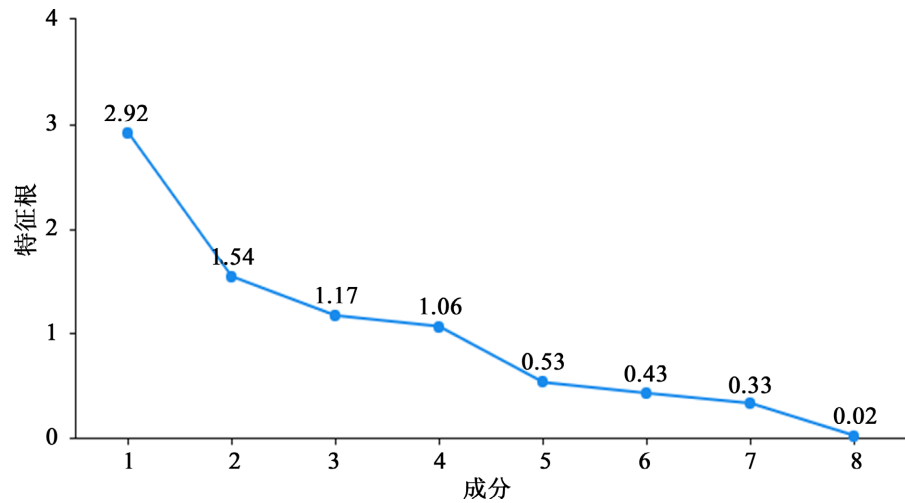


Figure 1. Gravel diagram
图 1. 碎石图

Table 5. Load factor table
表 5. 载荷系数表格

名称	载荷系数				共同度 (公因子方差)
	主成分 1	主成分 2	主成分 3	主成分 4	
应付职工薪酬	0.684	0.197	-0.508	-0.081	0.771
固定资产总计	0.527	-0.145	-0.705	0.118	0.809
营业总成本	0.710	-0.615	0.309	-0.014	0.978
净资产收益率	0.696	0.319	0.090	0.405	0.758
营业总收入同比增长率	0.333	0.582	0.198	0.557	0.799
营业总收入	0.776	-0.558	0.267	0.052	0.987
资产负债率	0.508	0.329	-0.077	-0.653	0.798
总资产周转率	0.462	0.495	0.441	-0.374	0.792

Table 6. Linear combination coefficient matrix
表 6. 线性组合系数矩阵

名称	成分			
	成分 1	成分 2	成分 3	成分 4
应付职工薪酬	0.400	0.158	-0.470	-0.079
固定资产总计	0.309	-0.117	-0.652	0.114
营业总成本	0.416	-0.495	0.286	-0.014
净资产收益率	0.407	0.257	0.083	0.393

续表

营业总收入同比增长率	0.195	0.468	0.183	0.540
营业总收入	0.454	-0.449	0.247	0.050
资产负债率	0.298	0.265	-0.071	-0.633
总资产周转率	0.270	0.399	0.408	-0.363

由表 5 和表 6 可知：

成分得分 1 = 0.400 * 应付职工薪酬 + 0.309 * 固定资产总计 + 0.416 * 营业总成本 + 0.407 * 净资产收益率 + 0.195 * 营业总收入同比增长率 + 0.454 * 营业总收入 + 0.298 * 资产负债率 + 0.270 * 总资产周转率

成分得分 2 = 0.158 * 应付职工薪酬 - 0.117 * 固定资产总计 - 0.495 * 营业总成本 + 0.257 * 净资产收益率 + 0.468 * 营业总收入同比增长率 - 0.449 * 营业总收入 + 0.265 * 资产负债率 + 0.399 * 总资产周转率

成分得分 3 = -0.470 * 应付职工薪酬 - 0.652 * 固定资产总计 + 0.286 * 营业总成本 + 0.083 * 净资产收益率 + 0.183 * 营业总收入同比增长率 + 0.247 * 营业总收入 - 0.071 * 资产负债率 + 0.408 * 总资产周转率

成分得分 4 = -0.079 * 应付职工薪酬 + 0.114 * 固定资产总计 - 0.014 * 营业总成本 + 0.393 * 净资产收益率 + 0.540 * 营业总收入同比增长率 + 0.050 * 营业总收入 - 0.633 * 资产负债率 - 0.363 * 总资产周转率

以及综合得分是方差解释率与成分得分乘积后累加计算得到。针对当前数据的计算公式为：

$$(36.476 * \text{成分得分 1} + 19.286 * \text{成分得分 2} + 14.605 * \text{成分得分 3} + 13.296 * \text{成分得分 4}) / 83.664$$

最终为：0.436 * 成分得分 1 + 0.231 * 成分得分 2 + 0.175 * 成分得分 3 + 0.159 * 成分得分 4。各公司得分结果如表 7 所示。

Table 7. Principal component scores and rankings
表 7. 主成分得分及排名情况表

公司	综合得分	第 4 主成分得分	第 3 主成分得分	第 2 主成分得分	第 1 主成分得分	排名
上海钢联电子商务股份有限公司	1.76	-2.43	3.07	1.43	2.94	1
北京国联视讯信息技术股份有限公司	1.59	0.03	2.98	1.94	1.41	2
三人行传媒集团股份有限公司	1.58	1.34	2.48	1.17	1.52	3
浙江海亮股份有限公司	1.07	-0.78	0.55	0.06	2.48	4
中科软科技股份有限公司	0.91	-0.27	0.48	-0.05	2.02	5
云鼎科技股份有限公司	0.80	2.59	0.31	3.27	-0.96	6
深圳市桑达实业股份有限公司	0.76	-1.03	-0.92	0.27	2.35	7
杭州海康威视数字技术股份有限公司	0.76	0.55	-1.03	-1.08	2.53	8
中国移动有限公司	0.65	0.67	-0.78	-1.42	2.31	9
深圳传音控股股份有限公司	0.62	-0.42	-1.07	-0.14	2.08	10
城发环境股份有限公司	0.57	-0.10	-0.65	-0.39	1.80	11

续表

中国软件与技术服务股份有限公司	0.52	-0.89	-0.30	-1.46	2.40	12
神州数码集团股份有限公司	0.49	-2.24	-0.35	2.83	0.58	13
深圳华强实业股份有限公司	0.48	-0.51	-0.20	0.92	0.88	14
国电南瑞科技股份有限公司	0.42	0.66	-1.02	-0.45	1.38	15
一心堂药业集团股份有限公司	0.42	0.10	-1.79	1.48	0.86	16
珠海华发实业股份有限公司	0.39	-0.54	-0.41	-0.45	1.49	17
深信服科技股份有限公司	0.34	0.20	-0.08	-1.19	1.38	18
三七互娱网络科技集团股份有限公司	0.28	0.81	-1.72	1.12	0.44	19
软通动力信息技术(集团)股份有限公司	0.25	0.40	-0.96	1.04	0.27	20
广东申菱环境系统股份有限公司	0.18	0.37	-0.84	0.64	0.27	21
杭州格林达电子材料股份有限公司	0.17	1.30	1.65	-1.52	0.07	22
上海家化联合股份有限公司	0.15	0.02	-0.76	-1.47	1.42	23
广州赛意信息科技股份有限公司	0.13	0.85	-0.41	0.55	-0.13	24
慧博云通科技股份有限公司	0.04	0.87	0.52	1.89	-1.44	25
北京东方通科技股份有限公司	0.04	0.81	1.61	-2.06	0.23	26
中兴通讯股份有限公司	0.01	-0.61	-0.74	1.73	-0.37	27
荣联科技集团股份有限公司	0.00	-0.60	0.57	-0.15	0.08	28
浙江大华技术股份有限公司	-0.01	-0.13	-1.01	-0.28	0.58	29
曙光信息产业股份有限公司	-0.05	0.30	-1.19	1.34	-0.46	30
拓尔思信息技术股份有限公司	-0.11	0.37	0.19	-2.40	0.80	31
杭州壹网壹创科技股份有限公司	-0.12	1.72	-0.50	0.90	-1.17	32
新疆青松建材化工(集团)股份有限公司	-0.12	0.07	-0.59	-0.64	0.28	33
远光软件股份有限公司	-0.15	1.18	-0.61	0.31	-0.69	34
天阳宏业科技股份有限公司	-0.26	0.36	0.21	0.60	-1.13	35
北京海天瑞声科技股份有限公司	-0.33	1.78	0.34	0.01	-1.54	36
普元信息技术股份有限公司	-0.35	0.51	1.24	-1.13	-0.89	37
国投智能(厦门)信息股份有限公司	-0.47	0.29	-0.64	-0.52	-0.66	38
北京映翰通网络技术股份有限公司	-0.54	0.77	1.24	-1.03	-1.48	39
易点天下网络科技股份有限公司	-0.65	-0.56	0.69	-0.57	-1.27	40
汇纳科技股份有限公司	-0.67	0.62	0.53	-1.39	-1.24	41
北京东方国信科技股份有限公司	-0.68	0.39	-0.78	-0.91	-0.91	42
新疆准东石油技术股份有限公司	-0.75	-1.18	0.22	0.80	-1.80	43
万达信息股份有限公司	-0.90	-2.11	0.18	-0.88	-0.90	44
杭州当虹科技股份有限公司	-0.98	0.08	1.16	-1.63	-1.87	45
深圳市银之杰科技股份有限公司	-1.06	-0.87	0.44	0.63	-2.63	46
北京中科金财科技股份有限公司	-1.24	-0.48	0.30	-0.20	-2.68	47

续表

山东汉鑫科技股份有限公司	-1.57	-1.24	0.03	0.15	-3.25	48
北京飞利信科技股份有限公司	-1.71	-1.62	-0.99	-0.59	-2.62	49
中信尼雅葡萄酒股份有限公司	-2.65	-1.40	-0.68	-1.04	-4.75	50

4.3. 系统聚类分析

利用 SPSS 软件，对上述 8 个指标变量进行系统聚类，可得结果如下所示：

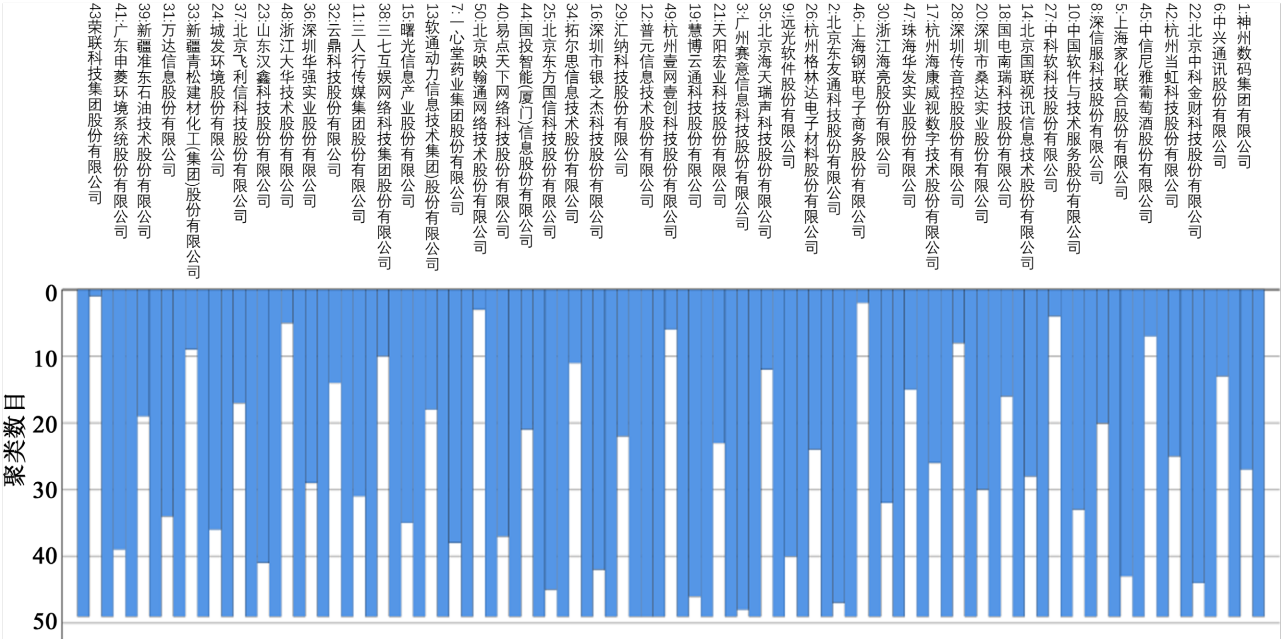


Figure 2. Ice wall chart
图 2. 冰挂图

从图 2 可以直观地看出，2022 年中国数字化企业的分类情况，群集数由上至下依次递增。

图 3 反映了 50 个数字化企业的分类情况，展示了分类状态和各类之间的距离。以普元信息技术股份有限公司、汇纳科技股份有限公司为例，它们距离最近且相似性极高，将它们合并成一类，按照此步骤进行下去，最后所有的数字化企业聚成一类，此时类间距离已经很大，聚类过程完成。

5. 结论与展望

5.1. 结论

通过主成分分析(PCA)，我们从 8 个原始指标中提取了 4 个主成分，这些主成分累积解释了 83.664% 的方差。这一结果表明，所提取的主成分能够有效地代表原始数据集中的主要信息。根据碎石图和特征值分析，我们确定提取四个主成分是合适的，因为它们在碎石图中形成了一个明显的“肘点”，并且在累积方差贡献率上达到了一个较高的水平。

系统聚类分析利用 SPSS 软件对数字化企业的经济绩效进行了分类。通过构建树状图和冰挂图，我们观察到企业之间的相似性和差异性。聚类结果显示，尽管存在一些紧密相连的企业，但整体上，企业间的经济绩效差异较大。特别是在聚类过程的后期，类间距离显著增加，表明不同类别的企业在经济绩效上

具有明显的区别。

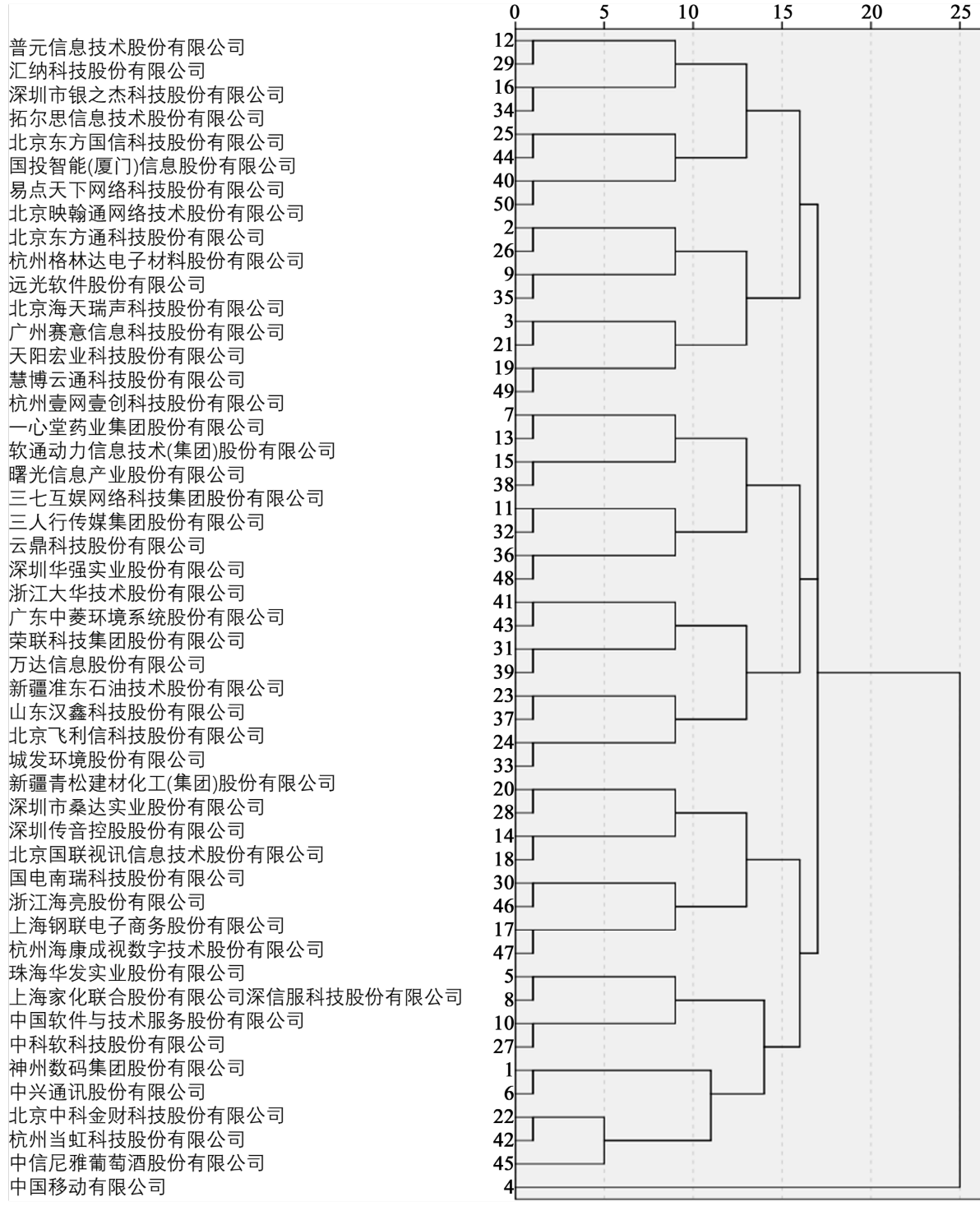


Figure 3. Pedigree diagram

图 3. 谱系图

根据综合得分，上海钢联电子商务股份有限公司在所有研究的企业中排名第一，显示出其经济绩效最为突出。而中信尼雅葡萄酒股份有限公司则排名最低，可能需要进一步分析其经济绩效较低的原因。

从载荷系数表格中可以看出, 营业总成本和净资产收益率在主成分 1 上有较高的载荷系数, 表明这两个指标对企业经济绩效的贡献较大。此外, 固定资产总计在主成分 3 上有显著的载荷系数, 这可能意味着固定资产的规模对企业的经济绩效有重要影响。

5.2. 研究启示与展望

系统聚类分析不仅帮助我们识别了数字化企业之间的相似性和差异性, 而且为进一步的比较分析提供了基础。通过聚类, 我们可以将企业分为不同的群体, 针对每个群体制定更为精准的策略和措施。

尽管本研究通过结合 PCA 和系统聚类分析为数字化企业经济绩效的评估提供了新的视角, 但也存在一些局限性。例如, 本研究仅考虑了有限数量的财务指标, 未来研究可以扩展到非财务指标, 如客户满意度、品牌影响力等。其次, 本研究的 KMO 值较低, 意味着数据中的变量之间存在较多的相关性, 这可能会对主成分分析(PCA)或因子分析的结果产生一些影响。此外, 本研究的样本仅限于中国上市的数字化企业, 未来可以扩大到全球范围内的企业, 以增强研究的普遍性和适用性。

参考文献

- [1] 余江, 白宇彤. 数字化转型战略对企业数字创新绩效影响研究[J]. 科研管理, 2024, 45(4): 1-11.
- [2] 姚晨. 数字化时代如何将众创空间创业资源转化为新创企业绩效——数字平台能力的中介[J]. 企业经济, 2024, 43(3): 80-89.
- [3] 池雅琼, 刘峰, 齐佳音. 数字化转型背景下企业数据保护成熟度模型构建[J]. 情报杂志, 2021, 40(9): 133-140.
- [4] 阮永芬, 李鹏辉, 王必顺. 基于主成分与聚类分析滇池湖相粉土物理力学参数[J]. 昆明理工大学学报(自然科学版), 2023, 48(3): 41-52.
- [5] 施应玲, 余欣玥. 基于 LMDI 和系统聚类的电力行业碳排放影响因素分析[J]. 生态经济, 2024, 40(2): 22-29.
- [6] 孙忠娟, 卢燃. 企业数字化转型的研究述评与展望[J]. 首都经济贸易大学学报, 2023, 25(6): 93-108.
- [7] 陈德球, 张雯宇. 企业数字化转型与产品市场竞争地位[J]. 武汉大学学报(哲学社会科学版), 2024, 77(2): 118-131.
- [8] 杨钊, 刘文瑜, 黄杰. 基于主成分分析和聚类分析综合评价不同品种藜麦在黄土高原地区种植的适宜性[J]. 江苏农业科学, 2023, 51(24): 21-32.
- [9] 吴海建. 主成分分析的基本思想及应用实例[J]. 河南省情与统计, 2003(4): 30-31.
- [10] 李尔国, 俞金寿. 一种基于输入训练神经网络的非线性 PCA 故障诊断方法[J]. 控制与决策, 2003(2): 229-232.
- [11] 刘秋彤, 张应应. 聚类分析和因子分析在绩效考评中的应用[J]. 统计学与应用, 2022, 11(1): 135-149.
- [12] 何远霞, 王兰, 焦登丹. 基于主成分-聚类分析法的 31 个省市经济发展水平的综合评价[J]. 中国管理信息化, 2023, 26(14): 177-179.
- [13] 杨永学, 陈毛华, 刘坤. 基于 ICP-MS/ICP-OES 及主成分分析法研究不同产地芡实 15 种元素的差异[J]. 中国食品添加剂, 2024, 35(3): 254-262.
- [14] 陈壮, 姜红, 倪婷婷. 基于系统聚类-神经网络的塑料饮料瓶差分拉曼光谱检验[J]. 塑料工业, 2023, 51(10): 148-159.
- [15] 林海明, 张文霖. 主成分分析与因子分析的异同和 SPSS 软件——兼与刘玉玫、卢纹岱等同志商榷[J]. 统计研究, 2005(3): 65-69.