

课题编号：20235-01

# 基于 VAVE 方法的公立医院内部控制 流程设计研究

广州新华学院 广东省生殖科学研究所（广东省生殖医院）

课题组

## 摘 要

本研究聚焦于广东省公立医院的内部控制流程优化，采用 VAVE 方法结合机器学习技术，旨在提升医院管理效率和服务质量。通过对 3008 条医院财务与业务数据的分析，研究运用主成分分析、关联规则学习和聚类分析等技术，识别出在成本控制、功能提升和内部管理方面表现优异的标杆医院。研究结果揭示了影响医院绩效的关键因素，并总结出一系列具有显著业务价值的特征组合，为其他医院提供了可借鉴的管理经验和策略。通过推广这些最佳实践，医院可以在资源有限的情况下，实现服务质量和运营效率的双重提升。本研究不仅填补了当前医院管理研究的空白，还为全国范围内的医院管理实践提供了科学的决策支持和改进措施，推动了医院管理的科学化和精细化。

### 课题研究目的、研究要点

本研究的主要目的是通过数据分析和机器学习技术，识别和推广广东省公立医院中的标杆医院，以优化医院的管理策略和资源配置。具体而言，研究旨在通过对医院财务与业务数据的深入分析，揭示影响医院绩效的关键因素，并为其他医院提供可借鉴的管理经验和策略。研究的要点包括：

1. 多维度绩效评估：通过主成分分析、关联规则学习和聚类分析等技术，全面评估医院在成本控制、功能提升和内部管理等方面的绩效。
2. 标杆医院识别：利用 A 分数和 B 分数筛选出表现优异的医

院，深入分析其成功经验和策略。

3. 成功经验推广：总结标杆医院的特征组合，提供具体的实施建议，帮助其他医院在资源有限的情况下优化运营，提高服务质量。
4. 数据驱动决策支持：通过数据分析和机器学习技术，为医院管理提供科学的决策支持，提升管理效率和服务质量。

# 目录

一、引言	5
(一) 研究背景	5
(二) 研究目标和研究问题	7
(三) 研究方法	8
(四) 研究贡献	9
二、文献综述	11
(一) 价值分析与价值工程 (VAVE)	11
(二) VAVE 与内部控制	13
(三) 关联学习	14
(四) K-Means 聚类算法	15
(五) 自动编码器	16
(六) 综合评述	17
三、数据	19
(一) 指标数据	19
1. 成本类指标 (C)	21
2. 功能类指标 (F)	22
3. 内部控制类指标 (Control)	23
(二) 数据处理	27
四、关联规则学习	28
(一) 关联学习过程	28
(二) 关联学习结果分析	29
(三) 标杆学习策略	30
(四) 经验推广	33
五、聚类分析	36
(一) K-means 聚类	37
(二) 使用自编码器改善聚类效果	40
(三) 维度匹配	42
(四) 聚类的内部属性与标杆挖掘	45
六、总结	51
参考文献	53

## 一、引言

### （一）研究背景

近年来，为加强行政事业单位的内部控制建设，财政部及相关部门自 2015 年起陆续发布了一系列政策文件，旨在规范和指导各单位的内部控制工作。《关于全面推进行政事业单位内部控制建设的指导意见》（财会〔2015〕24 号）为整体内控建设奠定了基础。随后，《行政事业单位内部控制规范（试行）》（财会〔2012〕21 号）和《行政事业单位内部控制规范、内控报告编报工作指南》（财会〔2016〕10 号）细化了内控操作的具体准则和报告编制的细则，进一步保障了内控实施的标准化和一致性。

此外，财政部每年发布的年度内部控制报告编报工作的通知，从 2017 年至 2022 年，包括财会〔2017〕11 号至财会〔2022〕6 号，这些文件通过明细化的报告要求，强化了行政事业单位的内控报告制度，确保内控执行的规范性和及时性。

在医疗领域，《关于加强公立医院运行机制改革的意见》（2020 年）由国家卫生健康委员会、财政部等联合发布，提出了加强公立医院内部管理和控制机制的建议，与《公立医院财务和经济运行管理制度》（2021 年）一起，为医院预算和财务管理提供了详细指引。这些文件强调了完善医院的运行机制与财务制度之间的紧密联系。

同时,《关于完善公立医院药品集中采购的指导意见》(2020年)通过加强药品采购使用的透明度和控制,进一步完善了医院在药品管理领域的内控措施。这些文件相辅相成,共同推动了广东省公立医院在塑造内部控制制度上的发展与实践。

结合这些政策,广东省公立医院在内部控制建设方面已取得一定进展,但仍面临如内部控制意识需进一步提高、实施细节需进一步完善等挑战。通过以上政策文件的指导及执行,进一步促进广东省公立医院在管理效率和风险防控能力方面的提升,实现提高医疗服务质量和运营效率的目标。这些政策文件在功能上相辅相成,形成了一个内控管理的综合体系,为医院高效运行及服务质量的提升提供了保障。

在现代医疗体系中,公立医院的管理和运营效率直接关系到公共卫生资源的优化配置和使用效率。随着医疗需求的不断增长和资源的有限性,如何在有限的资源条件下提升医院的服务质量和运营效率成为了一个亟待解决的问题。广东省作为中国经济最为发达的地区之一,其公立医院的管理模式和运营绩效具有重要的研究价值。本研究以广东省公立医院为样本,旨在通过数据分析和机器学习技术,挖掘出可供学习的标杆医院,以期为其他地区的医院管理提供借鉴。

近年来,数据驱动的决策在医疗行业中逐渐受到重视。通过对医院运营数据的深入分析,可以揭示出影响医院绩效的关键因素,并为管理者提供科学的决策依据。本研究利用了多种数据分

析技术，包括主成分分析、关联规则学习和聚类分析，系统地分析了广东省公立医院的财务与业务数据。通过对 80 个数值型变量的降维和分类，本研究从多维度评估了医院的绩效，识别出在成本控制、功能提升和内部管理等方面表现优异的标杆医院。

在此基础上，本研究进一步探讨了如何通过学习标杆医院的成功经验，优化资源配置和管理策略。通过对标杆医院的深入分析，研究总结出了一系列具有显著业务价值的特征组合，这些组合在不同的医院管理场景中展现出极高的潜力。研究结果不仅为广东省公立医院的管理提供了深入见解，也为全国范围内的医院管理实践提供了可行的改进措施。通过推广和实施这些最佳实践，医院可以在资源有限的情况下，实现服务质量和运营效率的双重提升。

## **(二) 研究目标和研究问题**

本研究的主要目标是通过数据分析和机器学习技术，识别和推广广东省公立医院中的标杆医院，以优化医院的管理策略和资源配置。具体而言，本研究旨在通过对医院财务与业务数据的深入分析，揭示影响医院绩效的关键因素，并为其他医院提供可借鉴的管理经验和策略。通过识别在成本控制、功能提升和内部管理等方面表现优异的医院，本研究希望为全国范围内的医院管理实践提供科学的决策支持和改进措施。

尽管已有研究对医院管理进行了探讨，但仍存在一些研究空

白。首先，现有研究多集中于单一维度的绩效评估，缺乏对医院在成本、功能和内部控制等多方面的综合分析。因此，研究需要探讨如何通过多维度的数据分析方法，全面评估医院的绩效并识别出标杆医院。其次，尽管一些研究识别了表现优异的医院，但缺乏对其成功经验的系统总结和推广。因此，研究需要解决如何系统地总结标杆医院的成功经验，并将其转化为可推广的管理策略和实践。最后，许多医院管理决策仍然依赖于经验和直觉，缺乏基于数据分析的科学支持。因此，研究需要探讨如何利用数据分析和机器学习技术，为医院管理提供科学的决策支持，提升管理效率和服务质量。

通过解决这些研究问题，本研究不仅填补了当前医院管理研究中的空白，还为实际的医院管理实践提供了有价值的指导和支持。研究结果将有助于推动医院管理的科学化和精细化，提升公共卫生资源的使用效率。

### **（三）研究方法**

本研究采用了一系列数据分析和机器学习技术，以系统地识别和分析广东省公立医院的标杆医院。首先，数据收集阶段涵盖了广东省内不同级别和功能的公立医院，共计 3008 条观测值。这些数据包括医院的财务和运营绩效指标，提供了一个全面的视角来评估医院的整体表现。为了确保数据的准确性和一致性，研究对数据进行了预处理，包括数据清洗、标准化和分类变量的 one-hot

编码。这些步骤确保了数据的质量，为后续分析奠定了坚实的基础。

在数据分析阶段，研究采用了主成分分析（PCA）对 80 个数值型变量进行降维，将其归类为功能、成本和控制三大类指标。每类指标又细分为鼓励和抑制子类，以便更细致地分析医院的绩效表现。通过这种降维方法，研究能够简化数据结构，突出关键特征，从而提高分析的效率和准确性。接下来，研究应用了关联规则学习和聚类分析等机器学习技术。关联规则学习采用 Apriori 算法发掘频繁项集，并生成关联规则，这些规则揭示了不同医院管理指标间的内在联系。聚类分析则通过 K-means 聚类进一步识别表现相似的医院集群，结合自编码器优化聚类效果，提高了聚类的准确性和解释性。

在标杆识别阶段，研究通过计算 A 分数和 B 分数，筛选出在成本控制、功能提升和内部管理等方面表现优异的医院。A 分数和 B 分数分别代表医院的成本控制能力和功能与内部控制能力，通过这些指标，研究能够有效地评估和比较不同医院的绩效表现。最后，研究对标杆医院进行了深入分析，总结出其成功经验和策略。这些经验和策略为其他医院提供了可借鉴的管理实践，帮助它们在资源有限的情况下优化运营，提高服务质量。

#### **（四）研究贡献**

本研究在多个方面做出了重要贡献，首先是通过系统的分析

方法识别和推广了广东省公立医院中的标杆医院。通过对医院财务与业务数据的深入分析，研究揭示了影响医院绩效的关键因素，并识别出在成本控制、功能提升和内部管理等方面表现优异的医院。这一识别过程不仅为广东省的医院管理提供了具体的改进方向，也为全国范围内的医院管理实践提供了科学的决策支持和改进措施。这种基于数据驱动的标杆识别方法为医院管理的科学化和精细化提供了新的视角。

其次，本研究通过多维度的绩效评估方法填补了现有研究的空白。传统的医院绩效评估多集中于单一维度，忽视了医院管理的复杂性和多样性。本研究通过主成分分析和聚类分析等技术，将医院绩效分解为功能、成本和控制三大类指标，并进一步细分为鼓励和抑制子类。这种多维度的分析方法不仅提高了评估的准确性和全面性，还为医院管理者提供了更为细致的绩效改进路径。通过这种方法，研究能够更准确地识别出标杆医院的成功经验，并为其他医院提供可操作的改进建议。

此外，本研究在标杆医院成功经验的系统总结和推广方面也做出了重要贡献。通过深入分析标杆医院的管理策略和运营模式，研究总结出了一系列具有显著业务价值的特征组合。这些特征组合在不同的医院管理场景中展现出极高的潜力，值得在整个组织内推广和实施。研究不仅揭示了这些特征组合的内在逻辑，还提供了具体的实施建议，帮助其他医院在资源有限的情况下实现服务质量和运营效率的双重提升。

最后，本研究通过数据分析和机器学习技术，为医院管理提供了科学的决策支持。这种基于数据驱动的决策支持系统不仅提升了医院管理的效率和效果，还为未来的医院管理研究提供了新的方法论基础。研究表明，通过不断学习和改进，医院可以始终保持在行业中的领先地位，并确保其管理策略能够在新的环境下继续发挥作用。这一贡献为推动医院管理的持续改进和创新提供了重要的理论和实践支持。

## 二、文献综述

### （一）价值分析与价值工程（VAVE）

价值分析与价值工程（VAVE）运用系统化的方法，通过最大化功能和最小化成本来提升产品或服务的整体价值。它的核心理念是将功能和成本分别归类为鼓励类和抑制类，目标是实现资源的最优配置。这种方法被广泛应用于多个领域，如电子商务和软件开发。

张彩江和徐咏梅（2002）对我国价值工程理论与实践的发展进行了回顾，分析了影响其应用效果的深层次原因。他们指出，尽管价值工程在中国的应用逐渐增多，但理论与实践的融合仍需不断深化。这项研究为后续研究提供了方向，强调了在不同领域中进一步探索价值工程的必要性。张彩江和马庆国（2005）探讨了价值工程应用系统的复杂性机理。他们的研究为复杂项目中的

价值最大化提供了理论支持，揭示了在复杂系统中应用价值工程时可能遇到的挑战和解决方案。王乃静和时文庆（2008）将价值工程应用于合理质量分析，强调了合理质量与最优成本之间的关系。这项研究推进了质量管理理论的实践应用发展，展示了价值工程作为决策支持工具在平衡产品质量与制造成本中的有效性。孟秀丽（2016）的研究聚焦于制造与运营管理领域，特别是客户订单分离点定位中的冲突。他们提出了基于价值工程的方法进行冲突消解，展示了价值工程在优化生产与供需匹配过程中的有效性，通过提升运营效率来增加经济效益。刘红华和曹连众（2020）从价值工程视角分析了大型体育赛事赞助的功能及其实现路径。他们的研究为赞助商提升投入回报和提高品牌价值提供了策略指导，突显了价值工程在体育经济中的应用潜力。

Akkermans 和 Gordijn（2003）在电子商务需求分析中强调了价值最大化的重要性，而 Chulani 等人（1999）通过贝叶斯分析优化软件开发成本，体现了 VAVE 在优化过程效率方面的应用。Bertoli 等人（2009）的系统建模工具虽然主要用于性能评估，但其资源优化理念与 VAVE 不谋而合。在工程领域，VAVE 也展示了其广泛的适用性。Colombo 等人（2003）的研究通过声发射信号分析评估混凝土梁的损伤，提示了如何通过精确分析提升工程设计的结构价值。而 Cornell（1968）的地震风险分析强调了风险管理在工程设计中的必要性，与 VAVE 理念中通过优化决策提高

价值的原则高度契合。Katcher 等人（1993）则在软件工程中展示了通过优化调度策略来提高系统性能和价值的可能性。Hazelrigg（1998）和 Roy（2003）分别从决策框架和成本管理的角度，进一步强化了在设计和项目管理过程中提升价值的重要性。

此外，VAVE 在教育和服务工程领域也展现了独特的价值导向潜力。Matusovich 等人（2010）研究了工程学科的教育动机价值，而 Sakao 和 Shimomura（2007）通过结合服务和产品的服务工程学科来增加价值，展示了 VAVE 在现代工程设计中广泛应用的前景。这些跨学科的研究共同展示了 VAVE 方法在实现资源最优配置和价值最大化方面的巨大潜力，强调了其在工程设计、风险管理和成本控制中的重要性。

## **（二）VAVE 与内部控制**

Bodnar（1975）强调了通过系统化的控制流程来提高组织的可靠性和效率，这与 VAVE 的目标一致，即通过优化流程和资源配置来最大化价值。可靠的内部控制能够减少浪费和错误，从而降低成本并提高功能价值。类似地，Borrajo et al.（2005）探讨了中小企业的自主内部控制系统，强调灵活性和适应性的重要性，这与 VAVE 方法中对灵活性和创新能力的重视相符。通过自主控制系统，中小企业能够更有效地管理资源和风险，实现更高的价值。此外，Masli et al.（2010）研究了内部控制监控技术的潜在好处，指出技术可以提高控制的有效性和效率，这与 VAVE 通过技术

和创新来优化成本和功能的理念相符。Stringer 和 Carey (2002) 的研究则强调了重新设计控制流程以适应变化环境的重要性，与 VAVE 关注通过重新设计和创新来提高产品和服务价值的目标一致。通过优化内部控制流程，组织可以更好地适应市场变化，提升竞争力。Tuan (2020) 的研究表明，内部控制对企业绩效有显著影响，尤其是在越南建筑企业中。有效的内部控制能够通过优化控制和管理流程来提高组织的整体价值和绩效。这些研究共同表明，优化内部控制流程不仅支持 VAVE 的目标，还能通过减少浪费、提高效率 and 增强灵活性来实现组织价值的最大化。

### **(三) 关联学习**

关联学习作为应对大数据时代挑战和机遇的强大工具，其研究领域不断扩展，特别是在提高算法的可扩展性和处理大规模高维数据集的效率方面 (Erlandsson et al., 2016)。关联规则学习已超越传统的市场分析和推荐系统，开始应用于医疗诊断、金融风险管理及智能交通系统等多个领域 (Li & Sheu, 2021)。例如，Doub et al. (2016) 的研究通过分析用户行为来区分用户偏好，为餐饮行业提供市场营销洞见，而在社交媒体中，Erlandsson et al. (2016) 应用此技术识别有影响力的用户，优化营销和信息传播策略。

Fürnkranz 和 Kliegr (2015) 以及 Kumbhare 和 Chobe (2014) 分别提供了关联规则学习的基础概述和算法综述，为数据科学家

选择合适的算法提供了指导。而 Kavšek et al. (2003) 则通过改进传统的 Apriori 算法，在子群发现领域成功应用关联规则学习，揭示了数据集中的关键模式。Li 和 Sheu (2021) 提出的启发式方法则有效解决了传统算法在处理大规模数据集时的性能限制。紧随其后，Santolucito et al. (2017) 将关联规则学习应用于编程语言研究领域，展示了其在软件工程中的潜力。

此外，Versichele et al. (2014) 的案例研究通过 Bluetooth 追踪数据，利用关联规则学习挖掘游客对旅游景点的访问模式，为城市管理提供了实用的建议。这些广泛的应用展示了关联规则学习从商业营销到社交网络、智能技术配置及旅游管理等领域的潜力，证明了其作为数据分析工具的重要价值。

#### **(四) K-Means 聚类算法**

近年来，K-means 聚类算法在多个领域特别是在财务和数据分析中取得了显著应用。作为一种无监督学习算法，K-means 通过将数据样本划分为 K 个簇，最大化同类样本的相似性并最小化不同簇间的相似性，从而有效地识别数据模式和异常。Argyrou 和 Andreev (2011) 将其应用于会计数据库的内部控制分析，显著增强了内部控制的有效性；Jokipii et al. (2011) 结合聚类和可视化技术，使得管理者得以直观地优化内部控制系统。进一步地，Lämsiluoto et al. (2016) 运用 K-means 揭示了不同组织内的控制策略有效性。

K-means 在预算管理和支出预测中同样显示出巨大潜力。Long 和 Zhu (2023) 的研究展示了其在大学预算支出智能预警中的应用，而 Novaliendry et al. (2015) 通过优化算法，提高了其在 Padang 预算分析中的效率和准确性。Nawrin et al. (2017) 将 K-means 与内部有效性指数结合，以增强交通管理系统中的聚类结果的有效性。此外，Zhu et al. (2022) 提出了一种改进的 K-means 算法，进一步提升了聚类的性能和稳定性。

在异常检测领域，K-means 应用也不可忽视。Sahoo 和 Sahoo (2021) 使用该算法进行会计欺诈检测，在识别财务数据异常方面展示了其有效性。类似地，Wakoli et al. (2014) 在医疗索赔欺诈检测中应用 K-means，提高了系统的安全性。这些研究表明，K-means 在财务、预算管理以及欺诈检测中的广泛应用前景。随着算法的不断优化，K-means 在更复杂数据集和应用场景中 will 发挥更大作用。

## **(五) 自动编码器**

近年来，自动编码器作为一种无监督学习模型，因其在异常检测和数据降维中的出色表现，获得了广泛应用。通过将输入数据压缩至低维潜在空间后再重构至原始空间，自动编码器能够学习数据的内在结构，从而在检测异常模式和数据重建方面表现优异。Patel et al. (2022) 利用该模型检测财务报表中的欺诈行为，显著提高了财务审计的准确性和效率。Rubio et al. (2020)

同样采用自动编码器对支付模式进行分类，体现出其在金融交易数据分析中的潜力。此外，Schreyer 等人（2018，2019，2020a，2020b）在会计数据异常检测中，通过使用深度自动编码器和对抗性自动编码器，提升了大规模数据中异常交易的识别能力，并展示了其在财务报表审计中的应用潜力。

在金融领域，自动编码器也被用于信用卡欺诈检测。Zou et al. (2019) 的研究表明，自动编码器神经网络能够通过学习正常交易的模式，有效识别信用卡交易中的异常行为。Zupan et al. (2020) 结合变分自动编码器和长短期记忆网络 (LSTM)，用于会计日记账的重建，展示了自动编码器在处理复杂时间序列数据中的应用潜力。此外，Schultz 和 Tropmann-Frick (2020) 的研究比较了自动编码器和外部会计师在检测财务报表审计中异常日记账条目方面的有效性，结果表明自动编码器能够在一定程度上替代传统审计方法，提供更为自动化和高效的异常检测手段。

总体来看，这些研究展示了自动编码器在金融和会计领域中广泛的应用前景，尤其是在异常检测和数据重建方面的优势。随着技术的不断进步，自动编码器有望在更多领域中发挥重要作用。

## **(六) 综合评述**

价值分析与价值工程 (VAVE) 通过优化功能和成本配置来提高产品和服务的整体价值。此方法体系强调功能与成本的分类，并通过系统化的方法实现资源的最佳配置。在这一框架下，机器

学习和内部控制流程作为现代技术与管理方法的结合，为 VAVE 的实施和价值最大化提供了重要支持。尤其是在电子商务、软件工程和现代服务产品设计中，基于数据驱动的决策成为优化流程和提升价值的关键所在。

机器学习在优化内部控制方面发挥了至关重要的作用，使企业得以通过数据分析预测和识别潜在问题，从而提高管理效率。内部控制系统通过应用机器学习工具，可以实现更高的灵活性和适应性，满足 VAVE 对创新和快速响应市场变化的需求。例如，通过使用先进的监控技术和自动化算法，企业能更加迅速地检测和反应误差及非期望的运营情况。这种实时反馈机制提升了内部控制的有效性，进而促使企业通过持续优化资源配置和流程设计，实现价值最大化。

在这一背景下，关联学习算法和工具如 K-Means 聚类、自动编码器等正成为支持 VAVE 的重要技术手段。K-Means 聚类被广泛应用于识别和分类内部控制中的异常模式以及优化资源配置，从而间接支持了 VAVE 的实施。此外，自动编码器通过学习数据的潜在结构，有效助力于财务与运营异常的检测，提高了企业对于潜在风险的反应能力。这些数据驱动的技术不仅能够提升现有流程的效率，还帮助企业在动态市场环境中保持竞争力，最大化其产品和服务的价值。

### 三、数据

#### (一) 指标数据

本研究通过国家卫健委、财政部官方网站获取 2008-2022 年中国卫生健康统计年鉴，搜集整理广东省纳入卫生财务年报填报系统的 614 家公立医院 2018-2022 年财务年报相关数据。数据涵盖了不同级别和功能的医院，共计 3008 条观测值。作为一种全面的分析方法，我们深入整合财务和运营业绩指标，以促进 VAVE（价值分析与价值工程）和内部控制的深入理解。这些指标按其在医院管理中的角色分为：成本类（C）、功能类（F）和内部控制类（Control），从而提供用于评价医院绩效的多维度视角。

表 1. 指标说明

指标名称	说明	角色	鼓励/抑制	观测值
hospital	医院			3008
year	年份			3008
city	所在城市			3008
type	医院定位			3008
organization	医院类型			3008
budget	预算级别			3008
level	医院等级			3008
start_funds	年初基建投资结余资金	C	-	3008
start_funds_public	财政性投资	C	-	3008
Retire	年平均离退休人数	C	-	3008
expenses	费用总计	C	-	3008
medince_exp	药品费	C	-	3008
material_exp	卫生材料	C	-	3008
admin_exp	单位管理费用	C	-	3008
average_publicpensions	离退休人均财政补助	C	-	3008
surplus	本期盈余	Control	-	3008
receivable	应收在院病人医疗款	Control	-	3008

bad_debt	坏账准备-应收账款坏账准备	Control	-	3008
debt_assets	资产负债率	Control	-	3008
public_revenue	财政拨款收入占总收入比	Control	-	3008
FA_per_bed	每床位占用固定资产	Control	-	3008
bed_turnovers	病床周转天数	Control	-	3008
average_days_discharges	出院者平均住院天数	Control	-	3008
material_turnover	卫生材料周转天数	Control	-	3008
public_exp	财政拨款收入占总费用比例	Control	-	3008
drug_incrate	药品收入占医疗收入比重	Control	-	3008
wage_exprate	人员费用占费用总额的比重	Control	-	3008
admin_exprate	管理费用占费用总额的比重	Control	-	3008
medexp_medinc	百元医疗收入的医疗费用（不含药品）	Control	-	3008
material_medinc	百元医疗收入消耗的卫生材料（不含药品）	Control	-	3008
surplus_rate	医疗盈利率	Control	-	3008
liability	负债总计	F	-	3008
inventory	存货	F	-	3008
invest	本年建设项目批复总投资	C	+	3008
new_invest	本年新开工建设项目批复总投资	C	+	3008
funds_used	基本建设资金结转	C	+	3008
funds_provided_public	本年基建投资到位资金	C	+	3008
funds_self	单位自筹资金	C	+	3008
operat_exp	业务活动费用	C	+	3008
depreciation	固定资产累计折旧	C	+	3008
discharge_days	出院者占用总床日	Control	+	3008
ocf	日常活动产生的现金流量净额	Control	+	3008
bed_used_rate	病床使用率	Control	+	3008
medincome_FA	百元固定资产医疗收入（不含药品收入）	Control	+	3008
current_ratio	流动比率	Control	+	3008
quick_ratio	速动比率	Control	+	3008
cash_ratio	现金比率	Control	+	3008
outpatient_incrate	门诊收入占医疗收入的比重	Control	+	3008
bed_incrate	住院收入占医疗收入的比重	Control	+	3008
medser_incrate	医疗服务收入（不含药品、耗材收入）占医疗收入比例	Control	+	3008
material_incrate	卫生材料收入占医疗收入比重	Control	+	3008
netasset_growth	净资产增长率	Control	+	3008

FA_retention	固定资产增值保值率	Control	+	3008
project	本年建设项目个数	F	+	3008
new_project	本年新开工	F	+	3008
area	项目总建筑面积	F	+	3008
new_area	本年新开工工程建筑面积	F	+	3008
finish_area	本年竣工工程建筑面积	F	+	3008
staff	年末在职职工人数	F	+	3008
plan_beds	编制床位	F	+	3008
average_beds	平均开放床位	F	+	3008
end_beds	年末实际开放床位	F	+	3008
consultants	诊疗人次	F	+	3008
actual_open_days	实际开放总床日数	F	+	3008
actual_used_days	实际占用总床日数	F	+	3008
discharges	出院人数	F	+	3008
assets	资产总计	F	+	3008
fixedassets	固定资产净值	F	+	3008
revenue	收入总计	F	+	3008
public_income	财政拨款收入	F	+	3008
self_income	事业收入	F	+	3008
medical_income	医疗收入	F	+	3008
medicine_income	药品收入	F	+	3008
material_income	卫生材料收入	F	+	3008
increase_ppe	固定资产增长	F	+	3008
intangible_assets	无形资产净值	F	+	3008
FA_growth	固定资产增长率	F	+	3008
netFA_ratio	固定资产净值率	F	+	3008
average_publicwage	在职职工人均财政基本支出补助	F	+	3008

## 1. 成本类指标 (C)

成本类指标的设立旨在详细分析医院的支出结构和资源投入效率。年初基建投资结余资金 (start\_funds) 及其财政性投资部分 (start\_funds\_public) 直接反映了医院在年度初期可用的投资资源, 是衡量其基础设施发展潜力的一项重要指标。随着基础设施改进需求日益增长, 这部分资金的合理配置对医院的长期发展具有重

大意义。

费用总计 (**expenses**) 代表医院在医疗服务、人员和基础设施维护方面的总支出。子类指标如药品费 (**medince\_exp**) 和卫生材料费 (**material\_exp**) 则提供了更具体的成本分摊情况。这类指标帮助我们深入探究上升的医疗支出如何平衡服务质量和成本效益。此外，单位管理费用 (**admin\_exp**) 为管理层提供了有效控制行政开支的反馈机制。

在退休人员方面，年平均离退休人数 (**Retire**) 和离退休人均财政补助 (**average\_publicpensions**) 的数据为分析现有福利体系和未来的退退休负担提供了基础。这些数据对管理决策具有长远影响，尤其是在支出优化和福利制度改革方面。

## 2. 功能类指标 (F)

功能类指标用于评估医院的服务能力和规模扩展，其目的是通过深入理解这些指标，可以找出提高医院产出效率的途径。“负债总计 (**liability**)”和“存货 (**inventory**)”是医院财务健康状态的重要衡量标准，影响着其短期流动性及应对财务风险的能力。因此，合理规划负债和库存能够优化资金利用并维护财务稳健。

医院的固定资产增长 (**increase\_ppe**) 反映了其在基础设施投资和设备更新方面的投入频率。新开工建设项目 (**new\_project**) 和项目总建筑面积 (**area**) 则衡量医院的扩展规模及新服务能力的

开发潜力。“出院人数(discharges)”以及“诊疗人次数(consultants)”作为直接的服务产出指标，能够为医院业务能力和服务水平提供量化评估。

### 3. 内部控制类指标 (Control)

内部控制类指标侧重于评估医院资源管理和财务控制的效率，这些指标在维持医院健康运营的过程中扮演了重要角色。债务与资产比率 (debt\_assets)、坏账准备 (bad\_debt) 等指标用于监控医院在财务管理和风险控制方面的措施有效性。高效的内部控制可以减少坏账损失，并保证负债结构的合理性，以此提高医院的财务安全性。

财政拨款收入占总收入比 (public\_revenue) 以及药品收入占医疗收入比重 (drug\_incrate) 等指标，帮助我们了解医院收支结构的合理布局。这些数据不仅能支持医院在公共财政资源有限的情况下优化医疗服务的供给策略，还推动了对药品销售依赖度的重新评估。

此外，每床位占用固定资产 (FA\_per\_bed) 和病床周转天数 (bed\_turnovers) 等指标则直接关系到医院的运营效率和资源利用率。出院者平均住院天数 (average\_days\_discharges) 和卫生材料周转天数 (material\_turnover) 可以更详细地揭示医院的服务流程和效率水平。

三类指标内部分为两个子类：鼓励类和抑制类。**鼓励类指标的提升**一般是医院希望实现的目标，因为这些指标的积极变化意味着医院在资源利用、收入提升、业务扩展等方面取得了成功。它们是医院良好运营的基石，有助于推动医院的整体发展和服务质量的提高。**抑制类指标的控制和减少**在于避免不必要的财务负担和运营低效。通过对这些指标的抑制，医院可以减轻财务压力，避免资源浪费并提高管理效率。例如，通过优化运营流程减少不必要的住院天数，直接改善患者周转和医院效率。

鼓励类指标包括：

1. **收入类指标**：如医疗收入（`medical_income`）、药品收入（`medicine_income`）、卫生材料收入（`material_income`）等，这些指标的增长通常是受鼓励的，因为它们直接反映了医院的盈利能力和市场需求的满足程度。增加的收入能够为医院提供更多资金用于再投资和服务改善。

2. **资源利用类指标**：如每床位占用固定资产（`FA_per_bed`）、病床使用率（`bed_used_rate`）、出院人数（`discharges`）以及诊疗人次数（`consultants`）。这些指标表明医院有效利用其现有资源和设施来最大化产出，显示出较高的运营效率和良好的管理能力。

**3. 增长类指标：** 固定资产增长率（FA\_growth）、净资产增长率（netasset\_growth）等指标的增加是积极信号，表示医院在规模和资产价值上能持续增长，从而增强其竞争力和服务能力。

抑制类指标包括：

**1. 费用与成本类指标：** 如费用总计（expenses）、单位管理费用（admin\_exp）、药品费（medince\_exp）等。尽管某些费用增加可能是不可避免的，但总体上这些成本应当被仔细管理，以确保医院的支出合理且能与收入增长相协调。

**2. 负债类指标：** 如负债总计（liability）、资产负债率（debt\_assets）。这些指标的增加常常是需要抑制的，除非医院有明确的偿债计划和高度安全的收入来源。保持合理的负债水平是维持财务健康和稳定的重要策略。

**3. 低效类指标：** 如疾病住院周期长（average\_days\_discharges）和卫生材料周转天数长（material\_turnover）可能意味着医院的运营效率和资源使用效率有待提高。这些指标反映出随着住院时间和材料库存时间的延长，资源被不合理占用，并可能导致机会成本的增加。

从 VAVE 方法来讲，为了最大化价值，应当最大化功能鼓励类指标和成本鼓励类指标，最小化功能抑制类指标和成本抑制类

指标。因此，价值分析的核心不是增加所有功能，而是尽可能增加客户受欢迎的功能。同时，价值分析也不是削减所有成本，而是削减不增加目标功能的成本。把功能和成本区分为鼓励类和抑制类，本身是 VAVE 方法的核心理念所决定。

区分内部控制指标的理由在于，内部控制的目标本质上并不总是与极值优化相一致。内部控制通常旨在确保组织的资源使用效率、合规性和风险管理，但这些目标可能与最大化组织价值的目标存在一定程度的矛盾。首先，内部控制的目标往往侧重于减少风险和防止错误或舞弊。例如，严格的内部控制措施可以防止财务报表中的错误，确保法律和监管合规性，以及保护组织资产免受盗窃或损失。然而，这些措施在某些情况下可能导致流程的僵化和资源的过度消耗。例如，为了确保完全的合规性，组织可能需要投入大量资源进行审计和监控，这些资源本可以用于直接创造价值的活动。其次，内部控制的过度追求可能导致效率的降低。内部控制通常涉及多个层次的审批和监控，这在减少错误和违规的同时，也可能延缓决策过程，增加运营成本。例如，在一个强控制环境下，即使是小的决策也可能需要多层审批，这不仅延缓了项目的推进，还可能导致员工的积极性下降，因为他们可能感到缺乏自主权。这种低效率和高成本可能直接影响到企业的利润率，从而限制了企业的价值创造能力。

此外，内部控制的重点在于稳定和可持续性，而不是极值优

化。内部控制系统旨在建立一个稳定的运营环境，通过减少意外事件的发生来保护组织的长期利益。这种稳定性是通过风险的最小化和流程的标准化来实现的，然而，这与极值优化——即在最小资源投入的情况下获取最大回报——并不完全一致。过于强调内部控制可能会牺牲灵活性和创新能力，而这些往往是实现价值最大化的关键因素。

基于这些原因，本文把内部控制指标和功能性指标和成本性指标分离开，作为独立的指标。内部控制指标也分为鼓励类和抑制类。与功能和成本指标不同，内部控制鼓励类的指标不是最大化目标，而是目标会相对较高。内部控制抑制类指标则是目标设定相对较低。

## **（二）数据处理**

首先，做数据的预处理。对于因子分析中涉及的分类型变量，包括医院等级、地理区域等类别型数据，本研究进行了 one-hot 编码。通过这种编码方式，将每个分类变量的每一个类别都被转化为新的二进制特征，这样能够捕捉到所有类别细粒度的信息，以识别变量之间的潜在关联和结构。

然后，本研究对数值型特征做标准化处理。原始数据以广东省的医院为个体，2018-2022 年度为时间维度。标准化需要以个体分组进行，以避免不同个体之间数值范围的差异带来的影响。为

此，对于每个数值型特征，本研究在每个个体内计算其均值和标准差。这样就得到完整的原始数据。

现在，有 80 个数值型变量。分别将这 80 个数值型变量进行归类。如上所述，分为功能类指标、成本类指标和控制类指标。每类指标又区分为鼓励类和抑制类两个子类。所以，一个有六大类指标。首先，本研究将 80 类数值型变量通过主成分分析 (PCA) 降维为六大类指标。然后，使用六大类指标进行进一步学习。

## 四、关联规则学习

### (一) 关联学习过程

在关联规则学习之前，需要对连续型数据进行分散化处理，即对所有数值型指标转化为 High、Medium 和 Low 三个顺序元素的因子指标。

本研究在关联规则学习中采用 Apriori 算法。Apriori 算法是一种经典的关联规则学习算法，广泛应用于数据挖掘和市场篮子分析等领域。它的主要目的是在大型数据集中发现频繁项集，并基于这些项集生成关联规则。通过关联规则，数据分析人员可以识别出不同变量或事件之间的潜在关系，从而为业务决策提供依据。Apriori 算法通过迭代方式逐步扩大项集的大小，并利用频繁项集的特性来减少候选项集的数量，从而提高计算效率。

频繁项集是指在数据集中频繁出现的项目组合。通过设置支

持度阈值，算法只会保留那些在至少一定比例样本中出现的项集。这种过滤帮助减少候选项集的数量，使得计算更加高效。

在生成频繁项集之后，利用这些项集生成关联规则。关联规则由前件（**antecedents**）和后件（**consequents**）组成，描述了前件发生时后件发生的可能性。生成规则的核心指标是置信度（**confidence**），它衡量了在前件发生的情况下，后件发生的概率。通过设置置信度阈值，算法可以过滤掉置信度较低的规则，只保留那些具有较高可靠性的规则。

## （二）关联学习结果分析

规则学习一共筛选出 170 个关联规则。其中，本研究感兴趣的关联规则如表 1 所示。

表 2. 重要的关联规则

从关联规则学习结果中，我们可以观察到多个重要的规则。首先，规则显示，当成本抑制性特征处于高水平时，功能鼓励性特征也处于高水平的可能性是 100%。提升度接近 3 表明这两者之间存在强烈的正相关关系。另外，当成本鼓励性特征为高时，功能鼓励性特征为高的可能性为 81.3%。虽然置信度稍低，但提升度为 2.437，仍然表示两者之间有显著的正相关关系。在某些

规则中，我们看到在成本鼓励性特征较高且控制鼓励性特征较低的情况下，功能鼓励性特征较高的可能性为 80.4%。尽管支持度较低，但提升度表明这种组合对功能鼓励性特征的提升有一定的影响。此外，当成本抑制性特征较高且控制鼓励性特征较低时，医院更有可能同时表现出高成本鼓励性和高功能鼓励性特征。置信度高达 92.2%，提升度为 3.403，显示这种特征组合对医院的成本和功能表现有很强的积极影响。在另一个规则中，当成本鼓励性特征较低且控制鼓励性特征较高时，功能抑制性特征较低的概率为 70.9%。提升度为 2.126，表明这种组合有助于降低功能抑制性特征。

### **（三）标杆学习策略**

从关联规则的结果中提取并制定标杆学习策略，是一个系统的过程，它不仅需要深刻理解数据中的潜在关系，还需要将这些关系转化为具体的、可操作的行动计划。这一过程可以通过一系列有条不紊的步骤来实现，最终帮助组织识别和推广最佳实践，以提升整体绩效。

首先，识别高置信度和高提升度的规则是这一过程的起点。关联规则中具有高置信度和高提升度的规则通常表示某些前件与后件之间有强烈的关联性，即当前件发生时，后件几乎总是发生。置信度高表明规则在数据中具有较高的可靠性，而提升度高则表明前件对后件的实际影响大于偶然性，这意味着这些规则能够有

效揭示出导致卓越绩效的关键因素。例如，某个规则的前件是“成本抑制性特征为高”，而后件是“功能鼓励性特征为高”，且该规则的置信度为 1.0，这就意味着每当成本抑制性特征表现优异时，功能鼓励性特征也总是表现良好。提升度的高值进一步加强了这一判断，表明这种关联关系远超过随机分布的可能性。这类规则往往揭示了哪些特定的行为或管理策略在实现卓越绩效中扮演了重要角色。

在识别这些高价值的规则后，接下来的关键是分析与最佳实践相关的规则。通常，标杆学习的目标是识别并推广那些在行业或领域中被认为是卓越表现的特征或行为。因此，当关联规则的后件与这些卓越表现相关时，例如“功能鼓励性特征为高”或“成本抑制性特征为高”，这些规则的前件就成为了标杆学习的重点目标。这一过程涉及深入理解规则中的前件，这些前件可能包括具体的管理措施、资源配置方式或流程优化策略，它们共同促成了后件中的最佳表现。

在进一步的分析中，识别出最优特征组合是制定标杆策略的核心。例如，通过分析一系列规则，管理者可能会发现，某些特征组合如“成本抑制性特征高”与“控制鼓励性特征适中”共同出现时，往往伴随着卓越的功能表现。此时，标杆学习策略的制定就应围绕如何在组织中实现和推广这一特征组合展开。这不仅仅是对这些特征的简单模仿，而是深入研究这些特征背后的逻辑，理解为何它们能够带来优异表现。通过这种分析，管理者可以确定

哪些具体的措施最有可能在本组织中复制成功，并据此制定详细的实施计划。

聚焦关键指标也是从关联规则中提炼标杆策略的重要步骤。在多个规则中反复出现的关键指标，往往是对整体绩效有重要影响的因素。例如，某些特定的控制管理措施可能在多条高置信度规则中频繁出现，表明这些措施在不同环境下都具有显著的效果。因此，管理者可以优先考虑这些措施，将其作为标杆学习的核心内容，并在组织内部进行推广。通过聚焦这些关键指标，管理者可以确保资源和精力集中在最能提升绩效的领域，从而最大化标杆学习的效果。

制定行动计划是标杆学习的落地环节。在明确了哪些特征组合和关键指标最有助于实现卓越表现之后，下一步就是制定具体的行动计划。这些计划应该包括资源的合理分配，确保资源集中用于已识别的关键特征组合上。同时，管理者需要设计相应的培训和能力建设计划，以提升团队在这些关键领域的技能和知识，确保标杆策略能够被有效实施。此外，过程优化也是关键的一环，管理者应考虑如何在现有流程中引入和推广标杆学习内容，可能需要调整现有的流程或开发新的流程以适应这些最佳实践。

在行动计划制定完成后，将标杆策略在组织内进行推广和分享是必要的步骤。这可以通过内部培训、案例分享、成功故事宣传等方式来实现。内部培训可以帮助团队成员理解和掌握新策略，案例分享则能够展示这些策略在实际应用中的成功实例，而成功

故事的宣传可以激励员工积极参与标杆学习的实施。通过这种方式，标杆策略可以在组织内得到广泛的接受和应用，从而提高整体绩效。

标杆学习不仅是一个一次性的过程，而是一个持续的改进过程。随着时间的推移，市场环境、技术进步和竞争格局都会发生变化，标杆策略也需要随之调整。因此，定期使用关联规则分析工具来更新和迭代标杆策略是非常重要的。通过不断学习和改进，组织可以始终保持在行业中的领先地位，并确保其标杆策略能够在新的环境下继续发挥作用。

#### **（四）经验推广**

从关联规则的分析中，我们可以提炼出几组具有显著业务价值的特征组合，这些组合在不同的医院管理场景中展现出极高的潜力，值得在整个组织内推广和实施。

首先，最值得关注的是高成本抑制性特征与高功能鼓励性特征之间的强关联。这一规则揭示了一个重要的管理理念：当医院能够有效地控制成本，特别是在减少不必要的支出和提高资源利用效率方面，往往能够同时提升其功能表现。这不仅仅是因为节约了资金，更因为通过优化资源分配，医院能够更高效地运行，并在医疗服务质量和患者体验上表现优异。因此，这种特征组合的推广应集中在加强成本管理，如通过先进的成本控制技术、精细化管理以及流程优化，确保资源以最有效的方式得到使用。管

理者可以通过推行节约型医院建设，强化内部审核机制，优化采购流程等措施，来实现这一目标。

与此同时，高成本鼓励性特征与高功能鼓励性特征之间的关联也表明，当医院在资源投入上表现积极，尤其是在扩展设施和技术投资方面，其功能表现往往也更加出色。这提示我们，适度的投入，特别是在那些能够直接改善医疗服务和患者体验的领域，能带来显著的绩效提升。这一策略的推广可以通过制定明确的投资计划，将资源优先分配给那些能够带来最大功能提升的项目，例如新设备的采购、员工培训的加强、信息技术的升级等。管理层应通过科学的预算规划，确保每一笔投入都能最大限度地支持医院的核心目标，即提高医疗服务的质量和效率。

此外，另一个值得深入探讨的特征组合是低控制鼓励性与高成本抑制性特征在实现高功能鼓励性特征和高成本鼓励性特征时的作用。这个组合表明，在适度减少控制压力的情况下，医院能够更灵活地管理资源，从而在成本和功能表现上都取得优异成绩。这反映出，过度的管理控制可能会限制医院的运营效率和创新能力，而通过适度的去中心化管理和简化流程，医院可以在保持控制的同时，释放更多的创新潜力和资源自由度，从而推动整体绩效的提升。推广这一策略，管理者可以考虑减少繁琐的审批程序，赋予基层更多的决策权，同时确保在关键环节保持必要的控制。这种灵活而高效的管理模式，有助于提升医院的响应速度和整体效率，尤其是在快速变化的医疗环境中。

低成本鼓励性与高控制鼓励性特征组合在实现低功能抑制性特征时的作用也值得关注。这条规则表明，即使在资源有限的情况下，通过加强控制，医院仍能保持良好的功能表现。这为资源紧张的医院提供了一种可行的管理策略，即通过严格的预算管理、质量控制和流程标准化，确保每一项活动都能得到有效执行，从而避免浪费和低效。管理者可以通过加强内部控制，特别是在资金使用、采购管理和人员配置等方面，确保资源得到最优配置。这种策略在预算紧张或资源分配受限的环境中尤为重要，能够帮助医院在挑战中保持稳健运行。

同时，低控制鼓励性与高功能鼓励性特征的组合也提供了有益的洞见。即使在功能鼓励性特征较低的情况下，通过加强控制，医院仍能有效管理成本。这提示我们，即使在某些功能性投入不足的情况下，通过加强控制措施，医院仍然可以保持成本的低水平运营。这一策略的推广应重点放在加强内部审计、严格预算控制以及强化合规性管理上，以确保医院在资源有限的情况下，仍能高效地运行并控制成本。

综合这些特征组合的分析，标杆策略的推广应注重以下几个方面：首先，在资源管理上，医院应通过加强成本控制，合理规划资源投入，确保每一项支出都能为提升医疗服务质量和患者体验做出贡献。其次，在管理控制上，应灵活运用管理工具，平衡控制与自由度之间的关系，避免过度控制带来的运营僵化。同时，通过适度简化流程，赋予基层更多的决策权，医院可以在保持高

效控制的同时，增强其创新能力和市场响应速度。此外，在资源有限的情况下，加强内部控制和预算管理，确保每一项资源都能得到最优配置，是保持医院稳健运营的关键。

## 五、聚类分析

在关联规则学习之后进行聚类分析，能够为数据分析和决策提供更深入、更全面的洞见。这两种技术虽然有不同的侧重点，但它们在结合使用时能够互补，为复杂的业务问题提供更有力的支持。

关联规则学习主要用于发现变量之间的关联性，揭示在某些条件下哪些特征或行为会同时出现。然而，关联规则并不直接考虑数据的整体结构或相似性。聚类分析则通过将数据划分为若干个组（簇），帮助我们理解数据中不同样本之间的相似性和差异性。通过聚类分析，可以将具有相似特征的样本归类，从而在不同的关联规则组合中找出更具代表性的群体。这有助于我们理解哪些关联规则在特定的样本群体中更为显著。

其次，关联规则揭示了变量之间的关联，但这些规则在不同的样本群体中可能有不同的表现。聚类分析能够将数据分成多个具有相似特征的群体，然后对每个群体分别应用关联规则分析。这种方法可以帮助识别某些规则是否在特定群体中更有效，或者是否需要为不同的群体制定不同的策略。例如，某些规则可能在

特定的高成本低控制的医院群体中表现出更强的关联性，而在其他群体中表现较弱。通过聚类分析，可以对这些规则进行细化和优化，使其应用更具针对性。

### （一）K-means 聚类

K-means 聚类是一种广泛使用的无监督机器学习算法，主要用于将数据集中的样本分成  $K$  个不同的簇（clusters）。它的核心思想是通过迭代的方式，将样本根据其特征的相似性归类到不同的簇中，使得同一簇内的样本彼此之间的相似性最大化，而不同簇之间的样本相似性最小化。使用 K-means 聚类可以得到初步的聚类分析结果。

使用 K-means 聚类首先对数据进行进一步处理。因为本研究的目标是价值最大化（功能和成本的鼓励类指标最大化以及抑制类成本最小化）和内部控制的选择关系。所以，设置两个分数，A 分数是内部控制一致性（内部控制鼓励类越大越好，内部控制抑制内越小越好）。B 分数是体现价值最大化的分数。因此，A 分数和 B 分数可以作为样本形成聚类的两个依据。

定义每个个体的 A 分数=每个个体内部控制鼓励类变量-每个个体内部控制抑制类变量。定义每个个体 B 分数=每个个体功能鼓励类变量-每个个体功能和所有成本变量（对成本鼓励类和成本抑制类设置 20%和 80%的权重）。聚类个数  $K$  是超参数。一般计

算每个  $K$  值下的轮廓系数。轮廓系数（Silhouette Coefficient）是一种衡量聚类质量的指标，它结合了簇内的紧密性和簇间的分离度。高轮廓系数表示样本点在簇内紧密且与其他簇分离良好，意味着聚类效果更好。轮廓系数取值范围为 $[-1, 1]$ 。越接近 1 表示聚类效果越好。如果轮廓系数较低，说明样本点可能被错误地归入了某个簇，或者  $K$  值设置过大或过小。

表 3 是不同  $K$  值下的轮廓系数。

表 3. 不同  $K$  值下的轮廓系数

$K$	轮廓系数
2	0.348
3	0.338
4	0.324
5	0.321
6	0.328

在实际分析的时候，大多数情况会选择 2-6 类。轮廓系数一般在 0.5 以上表现为比较好的聚类分离。此时轮廓系数较低主要可能是两方面的原因。一是指定了聚类的标准，这里是 A 分数和 B 分数。如果数据对指定的聚类标准不匹配，那么聚类的效果可能不好。第二是数据本身的分布可能不满足 K-Means 聚类的假设。

以  $K=6$  为例，聚类可视化结果图如图 1 所示。

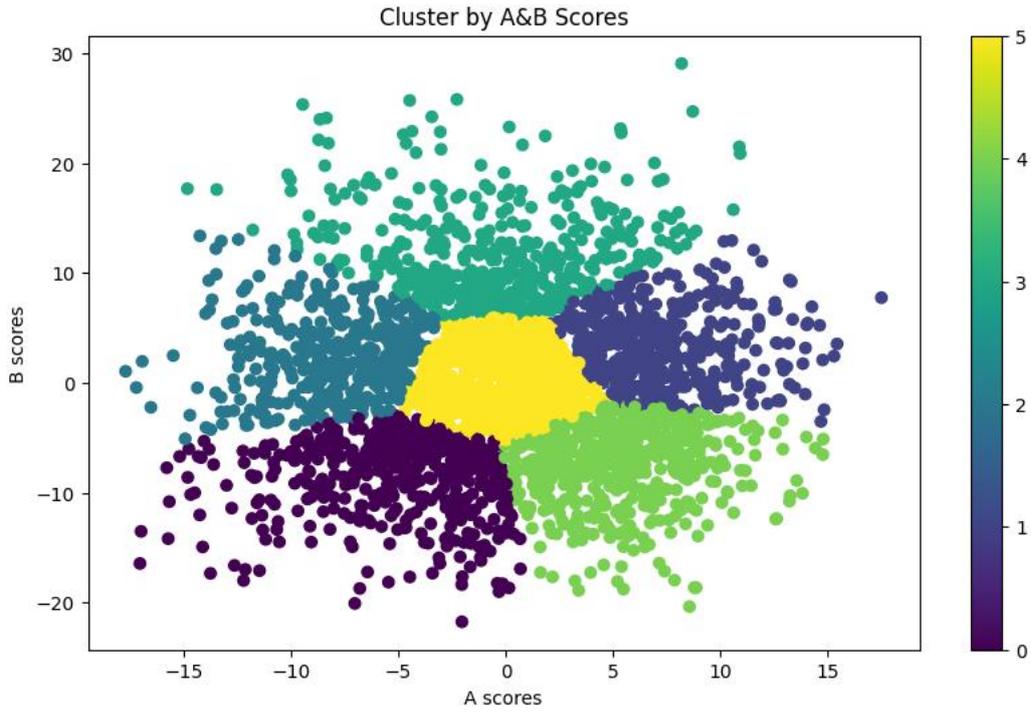


图 1. K-Means 聚类结果可视化

图 2 是不同聚类的轮廓系数。

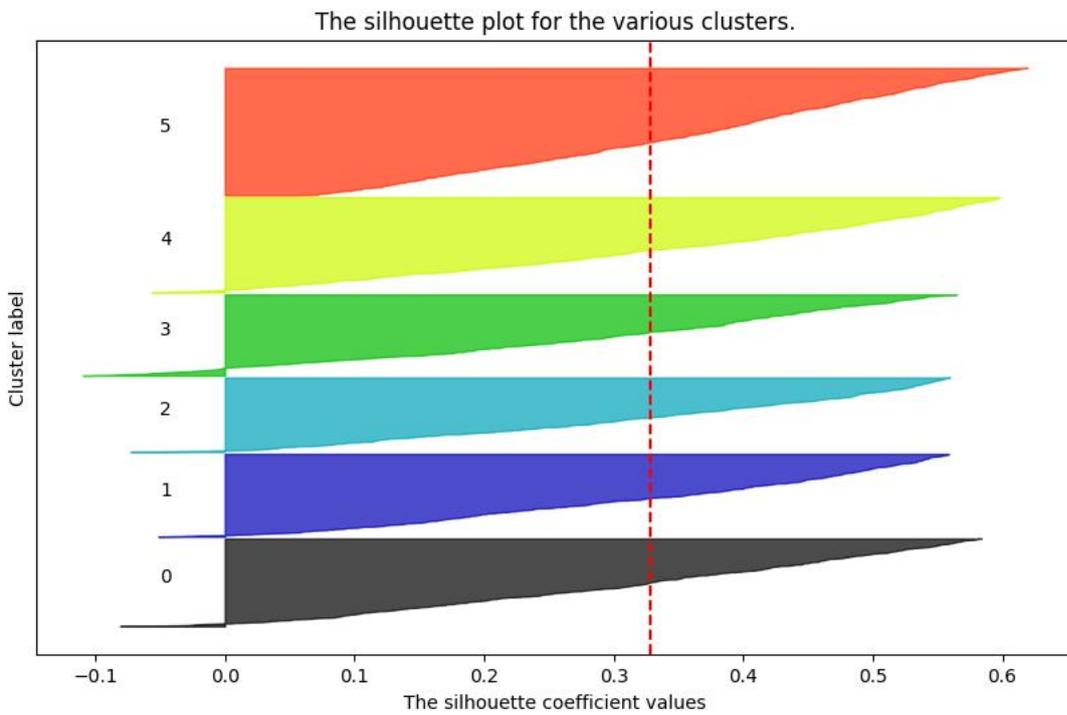


图 2. K-Means 聚类下不同聚类的轮廓系数

从轮廓系数图中看，每个簇（不规则条形图）被红色虚线截得的长度为其每个簇的轮廓系数。在样本平均水平（0.328）上，簇 5 的聚类效果最好。簇 0-3 聚类效果较差。因为整体分类效果不好，需要对当前的 K-Means 模型做进一步优化。

## （二）使用自编码器改善聚类效果

自编码器（Autoencoder）是一种用于无监督学习的神经网络，主要用于数据降维、特征学习和去噪等任务。自编码器由两个部分组成：编码器（Encoder）和解码器（Decoder）。编码器将输入数据映射到一个低维的潜在空间（即编码表示），这个过程可以看作是将数据压缩或提取特征。解码器则尝试从这个低维表示中重建出原始的高维输入数据。通过训练，自编码器的目标是最小化输入数据与重建数据之间的误差，确保低维表示能够最大程度地保留原始数据的重要特征。这种降维过程保留了数据中的关键信息，同时降低了数据的复杂性。

之前的 K-Means 聚类的输入数据是通过 PCA 降维的六类指标数据。通过自编码器降维，可以比 PCA 方法降维保留更多的信息。然后再进行 K-Means 分类，可以提高聚类的准确度。在自编码器降维的过程中，为了增强模型的学习能力，添加了四层隐藏层，并在每层隐藏层后添加了 Dropout 层。通过多个隐藏层和 Dropout 层，自编码器可以学习复杂的非线性特征，并通过 L2 正则化减少过拟合的风险。使用解码层将压缩后的数据还原为与输入数据相

同的维度，以便训练时可以进行自监督学习。自编码器模型使用 Adam 优化器进行编译，并以均方误差（MSE）作为损失函数来衡量重建误差。模型在训练过程中通过输入与输出的差异来调整权重，使得编码层能够有效地捕捉数据的主要特征。

训练完成后，使用编码器部分将数据降维为 2 个维度。然后，将降维后的数据输入到 K-means 算法中进行聚类分析。K-means 算法在降维后的低维空间中分配簇标签，并计算每个簇的中心点。在降维后的二维空间中进行 K-means 聚类，可以更容易地分离不同的簇，同时减小维度对聚类算法带来的计算复杂性和噪音影响。使用轮廓系数（Silhouette Score）来评估聚类的效果。轮廓系数用于衡量聚类的质量，分数越高表示聚类效果越好，簇内样本的相似性高且簇间的分离度大。

现在有两个超参数，聚类数和维度。根据实际需求，设置聚类数和维度分别为 {2,3,4,5,6}。轮廓系数计算如表 4 所示。

表 4. 自编码器优化下 K-Means 轮廓系数的计算结果

聚类数	维度	轮廓系数
2	2	0.600
2	3	0.682
2	4	0.454
2	5	0.479
2	6	0.683
3	2	0.627
3	3	0.627
3	4	0.585
3	5	0.424
3	6	0.427
4	2	0.604

4	3	0.449
4	4	0.421
4	5	0.587
4	6	0.418
5	2	0.633
5	3	0.602
5	4	0.580
5	5	0.577
5	6	0.405
6	2	0.413
6	3	0.415
6	4	0.403
6	5	0.393
6	6	0.406

如果要树立和分析不同的标杆，聚类数应设置较大为好。因此，聚类数  $K=5$ 。在维度越多，保留的信息越多，考虑到分类的准确性，维度设置为 4。在实际的工作中，聚类数和维度的超参数选择，需要根据实际的数据集特征和分析目标设置。

### (三) 维度匹配

通过自编码器降维的维度意义是不明确的，因此需要匹配具体的指标。本研究的做法是将之前 PCA 生成的六类指标和本次聚类的结果计算相关系数矩阵，观察聚类结果和六类指标的线性关系。

相关系数热力图如图 3 所示。

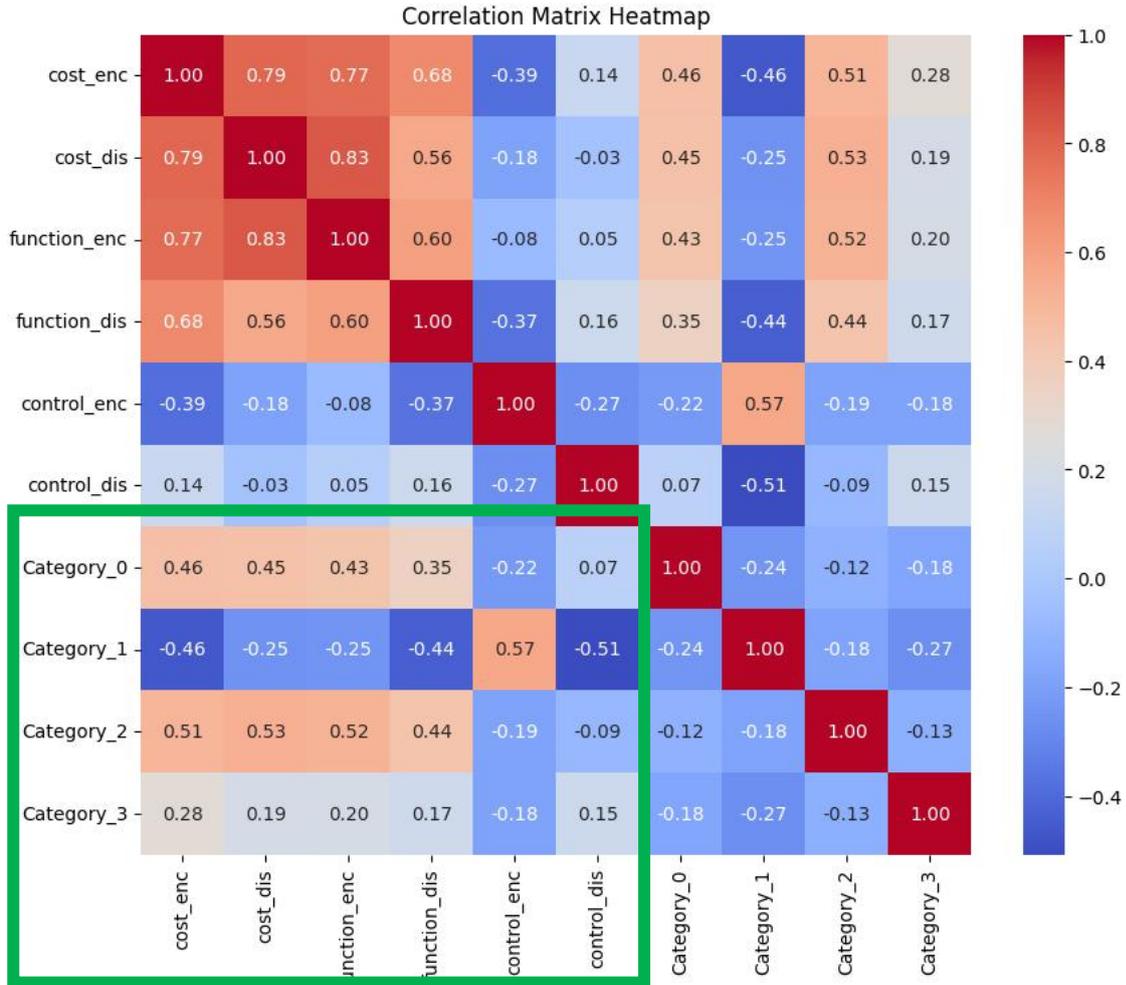
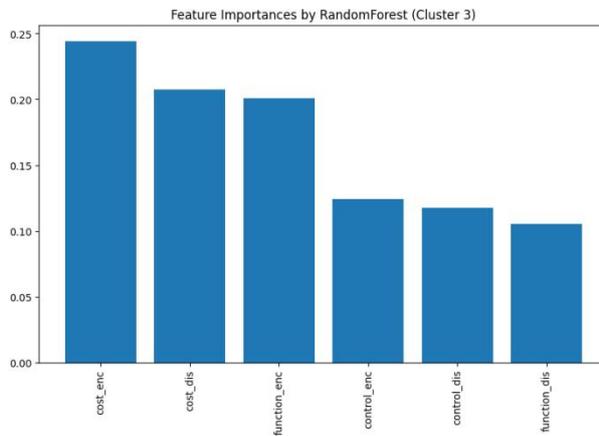
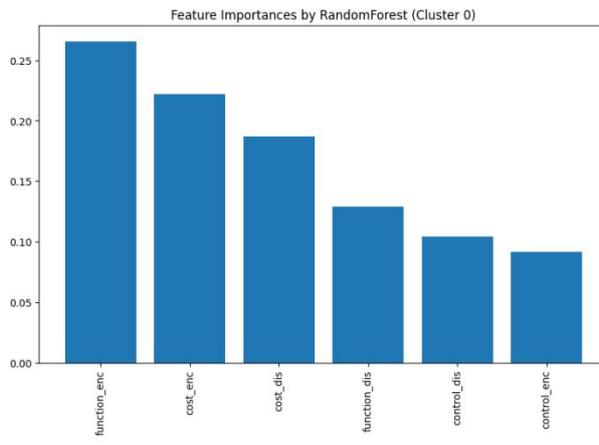
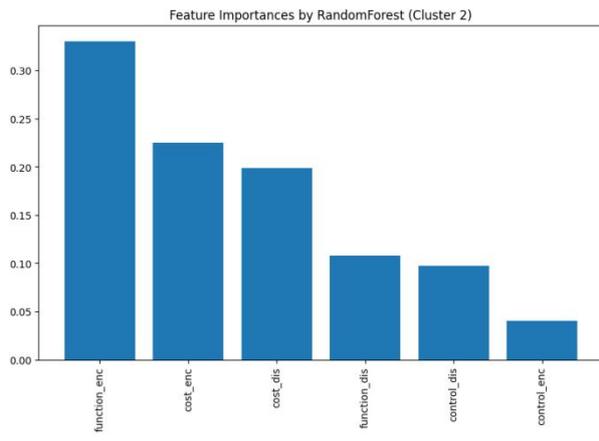
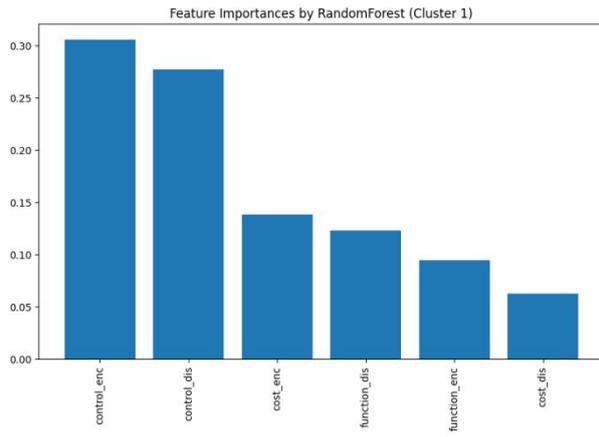


图 3. 相关系数热力图

可以看到，聚类 0、聚类 2 和价值测度指标相关度比较高，聚类 2 的相关程度更高。聚类 1 和内部控制测度指标相关度比较高。因此，聚类 2 表示价值测度的样本。聚类 1 表示内部控制相关的样本。

用随机森林算法进一步验证每个聚类中特征的贡献程度。图 4 是随机森林算法寻找重要特征的结果。依然是内部控制特征对聚类 1 贡献最多。价值测度特征对聚类 2 贡献最多。



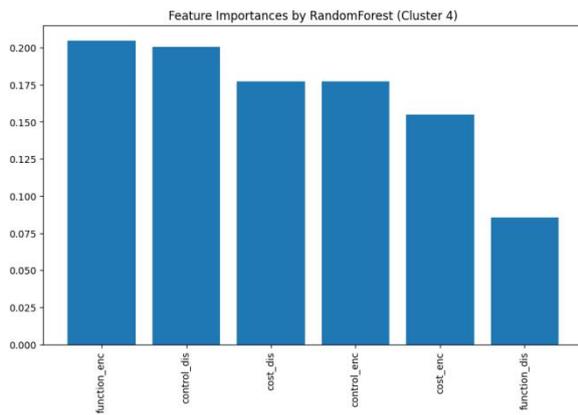
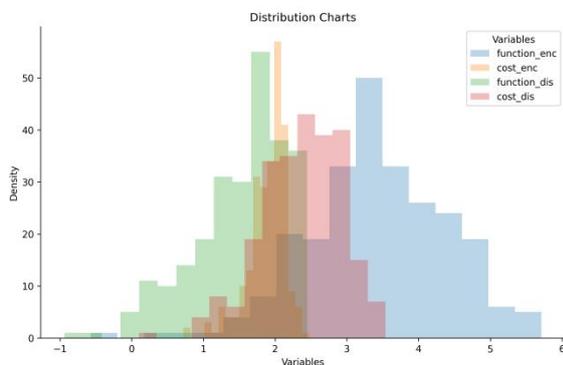


图 4. 用随机森林算法寻找聚类重要特征

#### (四) 聚类的内部属性与标杆挖掘

聚类 2 是价值测度簇。从图 5 可以看出，聚类 2 对功能类指标的区分度很好，因为功能鼓励类指标的分布趋向较大的数。但是，聚类 2 的成本类指标区分度不好。成本的鼓励类和抑制类的分布重叠在一起。因此，需要进一步对成本类指标过滤。比如，按照 20%和 80%对鼓励类成本和抑制类成本进行加权，计算加权成本指数（cost）。



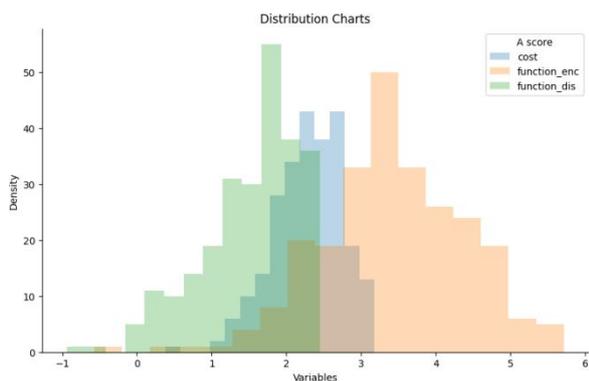


图 5. 聚类 2 不同指标的分布情况

成本指数主要分布在 1-3 之间。在成本指数小于 1.5 时，有 17 个样本，见表 5。因为聚类 2 本身对功能鼓励类和功能抑制类做了较好的区分，所以在截取成本指数较小的样本可以看成从聚类 2 中挖掘的标杆医院（趋向于价值最大化的医院）。

表 5. 聚类 2 中挖掘学习标杆

根据表 5 的分析，这 17 个样本的控制鼓励类和抑制类指标处于较低的水平，而成本鼓励类和成本抑制类指标则位于样本的较高水平。这个现象可能与这些医院的控制鼓励类指标水平较低密切相关。特别是当控制的激励效果不足时，医院在资源调配和成本管理上可能会面临较大的挑战，这会导致成本相关指标处于较高水平。

从医院特征来看，大多数样本为综合医院，且主要集中在大湾区的九个城市。然而，这些样本中并没有太多的三级医院，这表明在综合医院中，高级别医院可能在样本中占比较少。相反，部分小规模专科医院虽然规模不大，但在功能鼓励类指标上表现得非常出色。这种现象说明，一些小规模的专科医院在其专业领域内具备较强的竞争力和较高的服务质量，甚至超过了一些综合医院。它们在某些专业领域内的优异表现，为其他医院，尤其是那些规模较大但在特定领域表现平平的医院，提供了学习和借鉴的机会。

这进一步提示我们，医院的规模和综合性并不一定是其在所有功能领域表现优秀的必要条件。相反，专注于某一领域的专科医院可能更能够集中资源和力量，提升特定领域的医疗服务质量。因此，在未来的医院管理和改进措施中，可以考虑从这些小规模专科医院的成功经验中借鉴，特别是在特定功能领域的提升上，寻找具有针对性和可操作性的改进方案。同时，综合医院可以通过加强内部控制激励机制，优化资源配置和成本管理，提升整体服务水平。

从聚类 1 看（图 6），聚类对控制鼓励类和控制抑制类的分离比较好，两者的分布存在很明显的差异。在这个样本中，控制目标处于相对合理的区间。此时可以通过前文计算的 A 分数和 B 分数过滤出学习标杆。

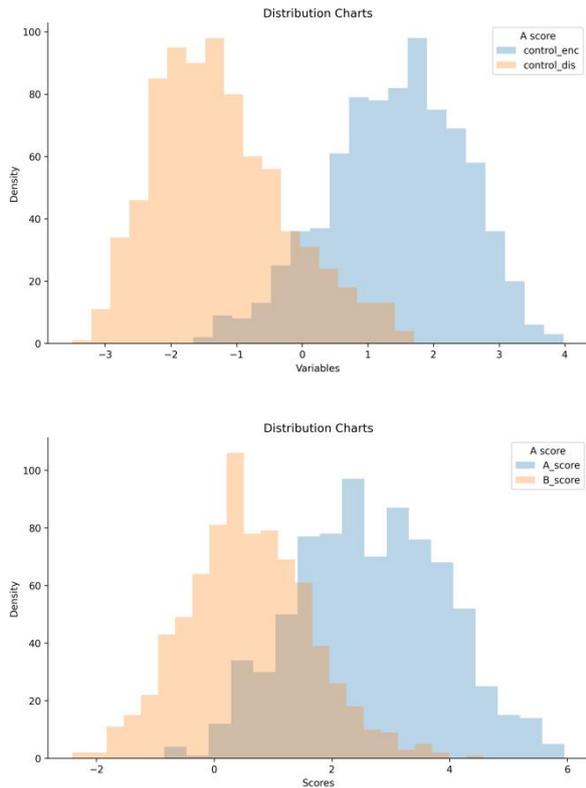


图 6. 聚类 1 不同指标的分布情况

按照 A 分数（取值在 4-6 之间）和 B 分数（取值在 3 以上）从聚类 1 中可以找到 15 个样本作为学习标杆。

表 6. 聚类 1 中挖掘学习标杆

从医院类型来看，标杆医院中多数是区县医院（type 为 0），而不是城市医院（type 为 1）。这表明在大多数情况下，即使是较小规模或资源有限的区县医院，也能通过有效的管理和策略成为

标杆。这些医院在成本鼓励类、功能鼓励类、和控制鼓励类指标上表现出色，意味着它们能够在有限的资源下优化运营，提升效率，并实现较高的服务质量。这为其他区县医院提供了一个重要的启示：在资源配置不如城市医院的情况下，通过精细化管理和资源的合理利用，同样可以取得卓越的绩效。

地理位置的分析显示，标杆医院大部分集中在大湾区九市（big\_bay 为 1）。大湾区作为中国经济最发达的区域之一，其医疗资源相对丰富，医疗需求较高。这些医院能够在激烈的市场竞争中脱颖而出，成为标杆，说明其在应对高压环境和复杂医疗需求方面具有强大的能力。值得注意的是，大湾区外的医院也有若干入选标杆，这些医院同样展示了强大的竞争力，尤其是在资源配置相对不足的地区，这更加凸显了这些医院的管理优势和创新能力。

大多数标杆医院为综合医院，这与其在多功能服务提供上的优势有关。综合医院通常涵盖了更广泛的医疗服务范围，这要求其在多个方面都具备较高的管理和服务水平。从数据中可以看出，这些综合医院在功能鼓励类和控制鼓励类指标上通常表现出色，这意味着它们能够在确保基本功能的同时，优化资源配置，提升服务质量。此外，部分专科医院虽然在综合性上不如大型综合医院，但它们在特定领域的专业性使得其在功能鼓励类指标上表现不俗，这显示出专科医院在其专业领域内的深耕和卓越表现。

标杆医院中，部分医院依赖于省级或地级市的预算支持，这

些医院往往在管理上有着更为规范的流程和更充分的资源保障。预算的充足使得这些医院能够在成本管理和资源配置上做得更好，从而在成本鼓励类和控制鼓励类指标上有着优异的表现。而那些没有预算支持的医院能够入选标杆，说明其在资源利用效率上有着卓越的表现，能够在有限的资源下通过高效管理达到较高的绩效。

医院等级的分析显示，三级医院虽然有更高的资源配置和更广泛的服务覆盖，但在标杆医院中所占比例并不大。这说明即使是非三级医院，只要在管理上做到精细化，在成本、功能、和控制方面的激励措施到位，同样可以达到标杆水平。这一发现对于许多非三级医院而言具有重要的参考意义，强调了医院管理的重要性，而不仅仅是依赖于医院的规模和等级。

从成本鼓励类、功能鼓励类和控制鼓励类指标的角度来看，标杆医院在这些方面通常有较高的得分，特别是在控制鼓励类指标上，几乎所有标杆医院都达到了较高水平。这表明这些医院在内部控制、管理激励和成本优化上有着非常严格的管理措施和执行能力。而在成本抑制类、功能抑制类和控制抑制类指标上，这些标杆医院的得分普遍较低，显示了它们在削减不必要开支和降低运营成本方面的有效性。这种表现进一步突显了这些医院在平衡成本与功能之间的能力，能够在保持高水平服务质量的同时，降低运营成本，优化资源配置。

具体来看，医院 N42 虽然是区县医院，但其控制鼓励类指标

表现优异，并且在大湾区内，表现出色。这说明其在面对高压环境时，依然能够通过有效的控制措施，维持较高的管理效率。而医院 N261 虽然不在大湾区内，但其成本和功能鼓励类指标均处于高水平，尤其是在控制鼓励类方面表现突出，说明其在非发达地区依然能够通过优化管理达到优秀水平，这为类似医院提供了有力的参考。

## 六、总结

本研究在广东省公立医院的财务与业务数据集基础上，研究涵盖了各种级别和功能的医院，总计 3008 条观测值。研究中的指标被分为成本类、功能类和内部控制类，这样的分类旨在从多维度评估医院的绩效。数据预处理包括对分类变量进行 one-hot 编码和对数值型特征进行标准化处理，确保了数据质量和分析的准确性。同时，80 个数值型变量通过主成分分析降维，划分为功能、成本和控制三大类，每类再细分为鼓励和抑制子类。

应用机器学习技术如关联规则学习和聚类分析来深入分析数据。关联规则学习采用 Apriori 算法发掘频繁项集，并生成关联规则，这些规则揭示了不同医院管理指标间的内在联系。聚类分析则通过 K-means 聚类进一步识别表现相似的医院集群，结合自编码器优化聚类效果，提高了聚类的准确性和解释性。

根据关联规则和聚类结果，识别并推广最佳实践，特别关注成本控制与服务能力提升之间的关系。制定具体的行动计划，包

括资源分配、培训和能力建设，以确保标杆策略的有效实施和持续改进。通过这种方法，研究不仅提供了公立医院管理的深入见解，还推动了实际的改进措施，优化了公共卫生资源的配置和使用效率。

## 参考文献

- Akkermans, J. M., & Gordijn, J. (2003). Value-based requirements engineering: Exploring innovative e-commerce ideas. *Requirements Engineering*, 8(2), 114–134. <https://doi.org/10.1007/s00766-003-0169-x>
- Argyrou, A., & Andreev, A. (2011). A semi-supervised tool for clustering accounting databases with applications to internal controls. *Expert Systems with Applications*, 38(9), 11176–11181. <https://www.sciencedirect.com/science/article/pii/S0957417411003873>
- Bertoli, M., Casale, G., & Serazzi, G. (2009). JMT: Performance engineering tools for system modeling. *ACM SIGMETRICS Performance Evaluation Review*, 36(4), 10–15. <https://doi.org/10.1145/1530873.1530877>
- Bodnar, G. (1975). Reliability modeling of internal control systems. *The Accounting Review*, 50(4), 747–757. <https://www.jstor.org/stable/245239>
- Borrajo, M. L., Corchado, J. M., Yáñez, J. C., Fdez-Riverola, F., & Díaz, F. (2005). Autonomous internal control system for small to medium firms. In H. Muñoz-Ávila & F. Ricci (Eds.), *Case-Based Reasoning Research and Development* (Vol. 3620, pp. 106–121). Springer Berlin Heidelberg. [https://doi.org/10.1007/11536406\\_11](https://doi.org/10.1007/11536406_11)
- Chulani, S., Boehm, B., & Steece, B. (1999). Bayesian analysis of empirical software engineering cost models. *IEEE Transactions on Software Engineering*, 25(4), 573–583. <https://ieeexplore.ieee.org/abstract/document/799958/>
- Colombo, Ing. S., Main, I. G., & Forde, M. C. (2003). Assessing damage of reinforced concrete beam using “*b* -value” analysis of acoustic emission signals. *Journal of Materials in Civil Engineering*, 15(3), 280–286. [https://doi.org/10.1061/\(ASCE\)0899-1561\(2003\)15:3\(280\)](https://doi.org/10.1061/(ASCE)0899-1561(2003)15:3(280))
- Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58(5), 1583–1606. <https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/58/5/1583/116673>
- Doub, A. E., Small, M. L., Levin, A., LeVangie, K., & Brick, T. R. (2016). Identifying users of traditional and internet-based resources for meal ideas: An association rule learning approach. *Appetite*, 103, 128–136. <https://www.sciencedirect.com/science/article/pii/S0195666316301271>
- Erlandsson, F., Bródka, P., Borg, A., & Johnson, H. (2016). Finding influential users in social media using association rule learning. *Entropy*, 18(5), 164. <https://www.mdpi.com/1099-4300/18/5/164>
- Fürnkranz, J., & Kliegr, T. (2015). A brief overview of rule learning. In N. Bassiliades, G. Gottlob, F. Sadri, A. Paschke, & D. Roman (Eds.), *Rule Technologies: Foundations, Tools, and Applications* (Vol. 9202, pp. 54–69). Springer International Publishing. [https://doi.org/10.1007/978-3-319-21542-6\\_4](https://doi.org/10.1007/978-3-319-21542-6_4)
- Hazelrigg, G. A. (1998). *A framework for decision-based engineering design*. <https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/120/4/653/432397>
- Jokipii, A., Lansiluoto, A., & Eklund, T. (2011). A clustering and visualisation approach to the analysis of internal control structures. *International Journal of Accounting, Auditing and Performance Evaluation*, 7(3), 151. <https://doi.org/10.1504/IJAAPE.2011.040832>
- Katcher, D. I., Arakawa, H., & Strosnider, J. K. (1993). Engineering and analysis of fixed priority schedulers. *IEEE Transactions on Software Engineering*, 19(9), 920–934. <https://ieeexplore.ieee.org/abstract/document/241774/>
- Kavšek, B., Lavrač, N., & Jovanoski, V. (2003). APRIORI-SD: Adapting association rule learning to

- subgroup discovery. In M. R. Berthold, H.-J. Lenz, E. Bradley, R. Kruse, & C. Borgelt (Eds.), *Advances in Intelligent Data Analysis V* (Vol. 2810, pp. 230–241). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-45231-7\\_22](https://doi.org/10.1007/978-3-540-45231-7_22)
- Kumbhare, T. A., & Chobe, S. V. (2014). An overview of association rule mining algorithms. *International Journal of Computer Science and Information Technologies*, 5(1), 927–930. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d4058d9f3f66c53ddea776c974fd740afd994b4>
- Länsiluoto, A., Jokipii, A., & Eklund, T. (2016). Internal control effectiveness—a clustering approach. *Managerial Auditing Journal*, 31(1), 5–34. [https://www.emerald.com/insight/content/doi/10.1108/MAJ-08-2013-0910/full/html?casa\\_token=8IRfdKqr-IcAAAAA:7AysAAN01bR0GjgCPaI5JuaxXpDMRQml-d-0G7a1toX7Qwey2y\\_gJMjz812ZFqq9BrpvGy4AivDw8OzJ6n1zOgia8kv2\\_Tohuc1E7wIELOQnUfvj8kdn](https://www.emerald.com/insight/content/doi/10.1108/MAJ-08-2013-0910/full/html?casa_token=8IRfdKqr-IcAAAAA:7AysAAN01bR0GjgCPaI5JuaxXpDMRQml-d-0G7a1toX7Qwey2y_gJMjz812ZFqq9BrpvGy4AivDw8OzJ6n1zOgia8kv2_Tohuc1E7wIELOQnUfvj8kdn)
- Li, H., & Sheu, P. C.-Y. (2021). A scalable association rule learning heuristic for large datasets. *Journal of Big Data*, 8(1), 86. <https://doi.org/10.1186/s40537-021-00473-3>
- Long, Q., & Zhu, W. (2023). Research on intelligent early warning of university budget expenditure based on K-means clustering algorithm. *2023 13th International Conference on Information Technology in Medicine and Education (ITME)*, 604–608. <https://ieeexplore.ieee.org/abstract/document/10505592/>
- Masli, A., Peters, G. F., Richardson, V. J., & Sanchez, J. M. (2010). Examining the potential benefits of internal control monitoring technology. *The Accounting Review*, 85(3), 1001–1034. <https://publications.aaahq.org/accounting-review/article-abstract/85/3/1001/3244>
- Matusovich, H. M., Streveler, R. A., & Miller, R. L. (2010). Why do students choose engineering? A qualitative, longitudinal investigation of students' motivational values. *Journal of Engineering Education*, 99(4), 289–303. <https://doi.org/10.1002/j.2168-9830.2010.tb01064.x>
- Nawrin, S., Rahman, M. R., & Akhter, S. (2017). Exploring k-means with internal validity indexes for data clustering in traffic management system. *International Journal of Advanced Computer Science and Applications*, 8(3). [https://www.academia.edu/download/76371971/Paper\\_37-Exploring\\_K-Means\\_with\\_Internal\\_Validity\\_Indexes.pdf](https://www.academia.edu/download/76371971/Paper_37-Exploring_K-Means_with_Internal_Validity_Indexes.pdf)
- Novaliendry, D., Hendriyani, Y., Yang, C.-H., & Hamimi, H. (2015). The optimized K-means clustering algorithms to analyzed the budget revenue expenditure in padang. *Proceeding of the Electrical Engineering Computer Science and Informatics*, 2(1), 61–66. <http://download.garuda.kemdikbud.go.id/article.php?article=556105&val=9548&title=The%20Optimized%20K-Means%20Clustering%20Algorithms%20to%20Analyzed%20the%20Budget%20Revenue%20Expenditure%20in%20Padang>
- Patel, H. A., Parikh, A., & Patel, B. (2022). Autoencoder-based model for detecting accounting statement fraud. In A. Khanna, D. Gupta, S. Bhattacharyya, A. E. Hassanien, S. Anand, & A. Jaiswal (Eds.), *International Conference on Innovative Computing and Communications* (Vol. 1388, pp. 11–21). Springer Singapore. [https://doi.org/10.1007/978-981-16-2597-8\\_2](https://doi.org/10.1007/978-981-16-2597-8_2)
- Roy, R. (2003). *Cost engineering: Why, what and how?* <https://dspace.lib.cranfield.ac.uk/items/f9e98087-7575-4029-b4f1-d1dcb04dcf6f>
- Rubio, J., Barucca, P., Gage, G., Arroyo, J., & Morales-Resendiz, R. (2020). Classifying payment patterns with artificial neural networks: An autoencoder approach. *Latin American Journal of Central Banking*, 1(1–4), 100013. <https://www.sciencedirect.com/science/article/pii/S2666143820300132>
- Sahoo, G., & Sahoo, S. S. (2021). Accounting fraud detection using K-means clustering technique. In D.

- Swain, P. K. Pattnaik, & T. Athawale (Eds.), *Machine Learning and Information Processing* (Vol. 1311, pp. 171–180). Springer Singapore. [https://doi.org/10.1007/978-981-33-4859-2\\_17](https://doi.org/10.1007/978-981-33-4859-2_17)
- Sakao, T., & Shimomura, Y. (2007). Service engineering: A novel engineering discipline for producers to increase value combining service and product. *Journal of Cleaner Production*, 15(6), 590–604. <https://www.sciencedirect.com/science/article/pii/S095965260600182X>
- Santolucito, M., Zhai, E., Dhodapkar, R., Shim, A., & Piskac, R. (2017). Synthesizing configuration file specifications with association rule learning. *Proceedings of the ACM on Programming Languages*, 1(OOPSLA), 1–20. <https://doi.org/10.1145/3133888>
- Schreyer, M., Sattarov, T., Borth, D., Dengel, A., & Reimer, B. (2018). *Detection of anomalies in large scale accounting data using deep autoencoder networks* (No. arXiv:1709.05254). arXiv. <http://arxiv.org/abs/1709.05254>
- Schreyer, M., Sattarov, T., Gierbl, A., Reimer, B., & Borth, D. (2020a). *Learning sampling in financial statement audits using vector quantised autoencoder neural networks* (No. arXiv:2008.02528). arXiv. <http://arxiv.org/abs/2008.02528>
- Schreyer, M., Sattarov, T., Gierbl, A., Reimer, B., & Borth, D. (2020b). Learning sampling in financial statement audits using vector quantised variational autoencoder neural networks. *Proceedings of the First ACM International Conference on AI in Finance*, 1–8. <https://doi.org/10.1145/3383455.3422546>
- Schreyer, M., Sattarov, T., Schulze, C., Reimer, B., & Borth, D. (2019). *Detection of accounting anomalies in the latent space using adversarial autoencoder neural networks* (No. arXiv:1908.00734). arXiv. <http://arxiv.org/abs/1908.00734>
- Schreyer, M., Schulze, C., & Borth, D. (2020). *Leaking sensitive financial accounting data in plain sight using deep autoencoder neural networks* (No. arXiv:2012.07110). arXiv. <http://arxiv.org/abs/2012.07110>
- Schultz, M., & Tropmann-Frick, M. (2020). *Autoencoder neural networks versus external auditors: Detecting unusual journal entries in financial statement audits*. <https://aisel.aisnet.org/hicss-53/os/risks/3/>
- Stringer, C., & Carey, P. (2002). Internal control re-design: An exploratory study of australian organisations. *Accounting, Accountability & Performance*, 8(2), 61–86. <https://search.informit.org/doi/abs/10.3316/INFORMIT.072066053849748>
- Tuan, T. (2020). The effect of internal control on the performance of vietnamese construction enterprises. *Accounting*, 6(5), 781–786. <http://growingscience.com/beta/ac/4044-the-effect-of-internal-control-on-the-performance-of-vietnamese-construction-enterprises.html>
- Versichele, M., De Groote, L., Bouuaert, M. C., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on bluetooth tracking data: A case study of ghent, belgium. *Tourism Management*, 44, 67–81. <https://www.sciencedirect.com/science/article/pii/S0261517714000417>
- Wakoli, L. W., Orto, A., & Mageto, S. (2014). *Application of the K-means clustering algorithm in medical claims fraud/abuse detection*. <http://41.89.229.23/handle/123456789/1002>
- Zhu, G., Li, X., Zhang, S., Xu, X., & Zhang, B. (2022). An improved method for k-means clustering based on internal validity indexes and inter-cluster variance. *International Journal of Computational Science and Engineering*, 25(3), 253. <https://doi.org/10.1504/IJCSE.2022.123112>
- Zou, J., Zhang, J., & Jiang, P. (2019). *Credit card fraud detection using autoencoder neural network* (No.

- arXiv:1908.11553). arXiv. <http://arxiv.org/abs/1908.11553>
- Zupan, M., Letinic, S., & Budimir, V. (2020). Accounting journal reconstruction with variational autoencoders and long short-term memory architecture. *SEBD*, 88–99. <https://ceur-ws.org/Vol-2646/05-paper.pdf>
- 黄龙森, 房俊, 周云亮, 等, 2024. 基于变分自编码器的近似聚合查询优化方法[J]. 浙江大学学报(工学版), 58(05): 931-940.
- 贾森, 刘宽, 徐萌, 等, 2024. 基于空谱多路自编码器的高光谱图像异常检测[J]. 遥感学报, 28(01): 55-68.
- 李飒, 2018. 基于关联规则的学习行为关联度分析方法研究[J/OL]. 微电子学与计算机, 35(06): 65-68. DOI:[10.19304/j.cnki.issn1000-7180.2018.06.014](https://doi.org/10.19304/j.cnki.issn1000-7180.2018.06.014).
- 刘红华, 曹连众, 2020. 价值工程视角下大型体育赛事赞助功能分析与功能实现路径[J/OL]. 北京体育大学学报, 43(09): 65-73. DOI:[10.19582/j.cnki.11-3785/g8.2020.09.008](https://doi.org/10.19582/j.cnki.11-3785/g8.2020.09.008).
- 刘萍, 胡月红, 2012. 基于 FCA 和关联规则的情报学本体构建[J]. 现代图书情报技术(02): 34-40.
- 刘潇, 2009. 关联规则挖掘在 DM 研究热点演化分析中的应用[J]. 情报杂志, 28(12): 143-147.
- 孟秀丽, 周林泉, 杨进, 等, 2016. 基于价值工程的客户订单分离点定位冲突消解[J]. 技术经济, 35(03): 88-94.
- 王朝, 黄慧涛, 张晶, 等, 2023. 基于解空间降维的大规模约束多目标进化算法[J]. 电子学报, 51(11): 3120-3127.
- 王蕾, 郭妍, 曹建华, 等, 2023. 基于 K-H-Mine 算法的废旧机械产品失效信息与再制造加工方案关联规则挖掘[J/OL]. 现代制造工程(08): 134-140. DOI:[10.16731/j.cnki.1671-3133.2023.08.019](https://doi.org/10.16731/j.cnki.1671-3133.2023.08.019).
- 王乃静, 时文庆, 2008. 价值工程在合理质量分析中的应用[J]. 管理学报(02): 228-232.
- 王倩楠, 王治国, 杨阳, 等, 2024. 基于多特征融合自编码器的无监督地震相分类研究[J]. 地球物理学报, 67(01): 370-378.
- 王振宇, 向泽锐, 支锦亦, 2024. 离散小波变换和自编码器耦合的脑电信号异常检测方法[J/OL]. 北京邮电大学学报, 47(02): 66-73. DOI:[10.13190/j.jbupt.2023-078](https://doi.org/10.13190/j.jbupt.2023-078).
- 吴青, 罗儒国, 王权于, 2015. 基于关联规则的网络学习行为实证研究[J]. 现代教育技术, 25(07): 88-94.
- 谢胜利, 陈泓达, 高军礼, 等, 2023. 基于分布对齐变分自编码器的深度多视图聚类[J]. 计算机学报, 46(05): 945-959.
- 许孝元 闵华清, 2005. 预测型关联规则演化学习的适应值函数[J]. 华南理工大学学报(自然科学版)(05): 1-6.
- 杨继红, 陈玲, 王晓龙, 等, 2024. 基于深度自编码器的核动力历史异常数据检测技术研究[J/OL]. 核动力工程, 45(02): 207-213. DOI:[10.13832/j.jnpe.2024.02.0207](https://doi.org/10.13832/j.jnpe.2024.02.0207).
- 杨建平, 向月, 刘俊勇, 2022. 面向配电网投资决策的小样本关联规则自适应迁移学习方法[J/OL]. 中国电机工程学报, 42(16): 5823-5834+6159. DOI:[10.13334/j.0258-8013.pcsee.210734](https://doi.org/10.13334/j.0258-8013.pcsee.210734).
- 张彩江 马庆国, 2005. 价值工程应用系统的复杂性机理认识[J/OL]. 科学学研究(02): 242-248. DOI:[10.16192/j.cnki.1003-2053.2005.02.018](https://doi.org/10.16192/j.cnki.1003-2053.2005.02.018).
- 张彩江 徐咏梅, 2002. 对我国价值工程理论与实践的回顾和影响降低的深层原因分析[J]. 南开管理评论(01): 14-19.
- 张思慧, 吴云龙, 张毅, 等, 2024. 一种基于联合变分自编码器的卫星重力数据粗差探测方法研究[J/OL]. 武汉大学学报(信息科学版), 49(06): 986-995. DOI:[10.13203/j.whugis20230226](https://doi.org/10.13203/j.whugis20230226).
- 周家玉, 侯慧娟, 盛戈皞, 等, 2023. 状态参量关联规则挖掘及深度学习融合的变压器故障诊断算法[J/OL]. 高压电器, 59(03): 108-115. DOI:[10.13296/j.1001-1609.hva.2023.03.015](https://doi.org/10.13296/j.1001-1609.hva.2023.03.015).