第十届湖南省研究生数学建模竞赛承诺书

我们仔细阅读了湖南省高校研究生数学建模竞赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们完全清楚，在竞赛中必须合法合规地使用文献资料、软件工具和 AI 工具，不能有任何侵犯知识产权的行为。否则我们将失去评奖资格，并可能受到严肃处理。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们授权湖南省研究生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号是（从组委会提供的赛题中选择一项填写）：

我们的参赛编号（请填写完整参赛编号）：

所属学校（请填写完整的全名）：

参赛队员（打印后签名）：1.

2.

3.

指导教师或指导教师组负责人（打印后签名）：

日期：2025 年 8 月 27 日

第十届湖南省研究生数学建模竞赛

题　目：民用航空运输机场安全保卫防范系统脆弱性评估

摘要：

现代机场安检系统作为保障航空安全的关键屏障，其运行效能与脆弱性评估是安全管理领域的核心议题。本文针对某机场旅检大队面临的质控、绩效分配及系统脆弱性等复杂问题，构建了一系列相互关联的数学模型，旨在实现从数据驱动诊断到智能化决策支持的闭环管理。

针对质控情况评估（问题一），本研究首先构建了质量分数(QS)与风险收益比(RRR)评价模型，对各组织单元的静态绩效进行了量化。进而，通过建立时序特征模型并结合卡方独立性检验（$\chi^2=104.93$, $p<0.01$），揭示了不同违规类型在一天内存在高度显著的时空演化规律，并利用卡方贡献值热图精准识别出早高峰操作风险、深夜流程风险等关键风险点。此外，通过对内外测数据的皮尔逊相关性分析（$\rho=-0.21$, $p=0.45$），证实了“霍桑效应”的存在，论证了暗测在评估真实绩效中的不可替代性。

针对绩效奖金分配（问题二），本文运用多元线性回归模型，以极高精度（$R^2=0.9932$）解构了薪酬系数的内在决定机制，量化了岗位晋升与档位提升的激励价值。同时，通过对缺勤数据的帕累托与长尾分布分析，揭示了“病假”主导和个体极端化的核心问题。基于此，我们提出了一个包含非线性分配函数、差异化扣罚规则和风险调整系数的绩效奖金分配优化模型。

针对系统性问题诊断与脆弱性评估（问题三、四），本研究首先运用数据包络分析（DEA-BCC）模型，从安全、能效、资源三维度对各勤务组的相对运行效率进行了评估，识别出效率洼地与投入冗余。最后，为解决数据收集难题，我们设计了一套标准化的数据收集表格体系。在此基础上，构建了基于 VSD 理论框架的系统脆弱性评估模型，通过熵权法-TOPSIS 对各分队的脆弱性进行量化排序，并引入障碍度模型诊断出制约系统韧性的关键障碍因子（如内测违规率、缺勤天数），最终形成了“绩效-脆弱性”四象限差异化管控策略。

综上，本研究建立了一套从微观行为诊断到宏观系统评估的综合性量化分析框架。针对赛题要求，我们从量化角度提出以下核心建议：

质控方面：实施基于时空风险预警的动态督导机制，并重构绩效评估模型，显著提升“暗测”结果权重以抵消“霍桑效应”，实现精准质控。

绩效奖金方面：引入基于综合绩效得分(CPS)的非线性分配函数以强化激励，建立差异化的缺勤扣罚规则提升公平性，并应用岗位风险调整系数实现跨团队的横向可比。

系统性问题方面：DEA 模型定量识别出“勤务二组”为非效率单元，存在投入冗余，应作为管理能效和资源分配优化的首要目标。

脆弱性评估与数据收集方面：推行我们设计的标准化数据收集表格；并将最终的绩效奖金分配与 VSD-熵权-TOPSIS 模型评估出的系统脆弱性挂钩，通过**“绩效-脆弱性”四象限矩阵**，将个人奖励与对系统整体韧性的贡献度相关联，实现安全与激励的统一。

关键词： 质控评估；绩效奖金；数据包络分析（DEA）；系统脆弱性；熵权法-TOPSIS；卡方检验

目录

一、问题重述

1.1 问题背景和意义

现代机场集成安检、行李处理、航班调度等数十个子系统，任一环节漏洞（如围界入侵、安检失效）均可能引发连锁风险。自“9·11”事件后，民航成为恐怖分子重点目标，劫机、爆炸、生化袭击等新型威胁不断涌现，迫使安保系统需持续升级应对能力。然而，目前民航安检存在包括走私、劫持航空器、网络攻击等的非法干扰行为多样化，传统安防手段难以覆盖全风险谱系。安检疏漏、内部人员作案、应急响应迟缓等人为失误占安全事故的 30%以上。部分机场仍依赖传统模拟监控和人工安检，而新型威胁（如隐形爆炸物、无人机干扰）需智能识别、大数据分析等新技术支撑。在此背景下，国际民航组织（ICAO）标准要求缔约国建立系统化的航空保安管理体系（SEMS），定期开展脆弱性评估并优化措施。同时，中国民航法规也进一步强化，如《国家处置劫机事件预案》《航空保安审计制度》等，明确将脆弱性评估纳入安全审计核心内容。

1.2 问题提出

1、使用统计分析方法，分析民航机场某旅检大队的岗位违规、X 光机、内部测试、重点违规等质控情况，并从量化角度给出质控建议。

2、使用统计分析方法，从缺勤原因、实发金额、各档奖金、奖惩情况、二次分配、过检率与绩效等方面分析该大队的绩效奖金数据，并从量化角度给出绩效奖金分配建议。

3、从安全漏洞、管理能效、资源分配角度，使用定量化的方法分析该大队目前存在的问题。

4、附件中的数据是使用传统手工操作的方式收集的，请优化设计数据收集表格。进一步地，请将定性与定量相结合，构建系统分析模型，从整体上对该大队的安检系统脆弱性进行评估，并给出绩效奖金分配建议。

二、模型假设

本研究建立的风险评估模型基于以下七个方面的基本假设：

2.1 数据质量假设

在题目提供的附件数据中，基准数据完整性假设。组织架构及人员名单数据完整且准确，包含所有相关人员的编制信息、岗位分配及组织结构关系，可作为人员信息分析的基准依据。

关键字段有效性假设。质控记录中的日期、岗位、人员姓名及问题描述等核心字段记录准确可靠，不同记录员之间的记录标准相对一致。

跨源数据一致性假设。不同数据源中的人员标识信息具备相对可比性，可通过姓名、岗位等关键字段进行有效关联与匹配，当匹配不上时，以基准数据为准。

2.2 风险动态性假设

事件独立性假设。各质控事件对人员风险值的影响相互独立，且同类型、同等级事件的影响程度一致，满足线性可加性原则。

影响瞬时性假设。质控事件对个人风险值的影响在事件发生时即刻生效，不存在延迟效应或累积阈值效应。

时间衰减特性假设。个体的风险水平会随时间变化，但仅可通过质控事件监测到，个人风险值可直接代表当前时刻个人风险状态的观测值。

2.3 人员行为假设

有限理性人假设。人员在知悉质控规则的前提下，会主动调整行为以降低个人风险水平，但调整过程存在一定的惯性延迟，且个人风险水平会以未知规则波动。

学习效应假设。人员能够从质控反馈中学习并改进操作，随着时间推移，重复性错误的发生概率会逐渐降低。

2.4 组织结构假设

管理有效性假设。分队、小组等管理单元能够有效传达和执行质控要求。

岗位风险均质假设。同一岗位面临的客观风险环境基本相同，观测到的风险差异主要源于个体行为而非环境因素。

组织稳定性假设。在研究期间，组织架构和人员岗位保持相对稳定，不发生重大结构调整或人员大规模变动。

2.5 抽样与测量假设

质控抽样代表性假设。质控记录为系统性抽样结果，抽样覆盖面足够广泛，能够有效反映整体安全水平，不存在明显的抽样偏差。

评估标准统一性假设。风险等级评估客观科学，其标准在不同时间、不同评估者之间保持一致，评价结果具有纵向和横向可比性。

测量误差随机性假设。数据记录和采集过程中的误差是随机的，不存在系统性偏差，且误差幅度在可接受范围内。

2.6 模型结构假设

线性可加性假设。不同类型事件对风险值的贡献满足线性叠加原理，且不同风险维度（岗位风险、设备操作风险等）之间相互独立。

参数稳定性假设。模型中的关键参数（如衰减因子、风险增量系数等）在研究期间保持稳定，不随时间发生显著变化。

申诉机制合理性假设。成功申诉的事件表明原始记录存在误差，其风险影响应当予以消除，且申诉过程本身不引入新的偏差。

2.7 局限性认知假设

外生变量恒定假设。模型未考虑的外部因素（如季节性变化、特殊事件影响等）在研究期间保持相对稳定，不对风险评估产生系统性干扰。

模型简化可接受性假设。为建立可操作的风险评估框架，对复杂现实情况的必要简化不会实质性影响模型的政策含义和实践指导价值。这些基本假设构成了本研究的理论基础，虽然现实情况可能不完全符合所有假设条件，但这些假设为建立简化而有效的风险评估模型提供了必要的框架。

三、 数据预处理

本研究的数据预处理旨在将原始多源数据转化为适合风险建模的结构化数据集，确保数据的完整性、一致性和可靠性，处理方法包括异常值处理、缺失值填充、关键数据校对并对重要数据进行量化，过程遵循完整性、一致性和可靠性原则，确保后续分析的准确性和有效性。

3.1 预处理目标与原则

本研究采用系统化的数据预处理方法，将来自多个来源的原始数据转化为适合风险建模的结构化数据集。处理过程遵循完整性、一致性和可靠性原则，确保后续分析的准确性和有效性。

3.2 多源数据整合

首先对四个主要数据源进行整合，包括组织架构信息、绩效奖金数据、考勤记录以及质控情况汇总表。通过建立统一的识别标识，实现不同数据源间的人员信息关联与匹配。

3.3 数据清洗与质量控制

采用分级处理策略对数据进行清洗。对关键字段进行完整性校验，剔除姓名、日期等核心信息缺失的记录。同时建立系统的异常值检测机制，识别并处理不符合逻辑关系或超出合理范围的数据条目。

3.4. 数据标准化处理

实施全面的数据标准化流程，将各类文本信息和分类变量转换为统一的数值表示。包括将勤务组、分队等组织单元进行编码映射，对风险等级、测试结果等评估指标进行标准化量化，并统一所有日期信息的格式规范。

3.5 人员信息重构

针对人员信息中的重名现象，建立多特征消歧机制，通过结合岗位、分队和入职时间等辅助信息进行身份识别。为每位人员构建从入职日期到分析截止日的完整时间序列框架。

3.6 风险量化体系

基于三类主要质控事件（岗位违规、设备操作和能力评估）建立风险影响量化模型。根据不同事件类型和严重程度，制定差异化的风险值变化标准，确保风险度量的科学性和一致性。

3.7 风险动态更新机制

采用事件驱动的时间序列更新方法，将质控事件按照发生时间匹配到对应人员的风险轨迹中。同时引入时间衰减效应，模拟风险随时间的自然消退过程。

3.8 质量评估与验证

通过交叉验证和一致性检查确保数据处理质量，建立完整的处理日志和元数据记录，保证整个预处理过程的可追溯性和可重现性，为后续风险评估建模提供可靠的数据基础。

四、问题一

4.1 问题分析

本题要求使用统计方法对分析民航机场某旅检大队的岗位违规、 光机、内部测试、重 X 点违规等质控情况并给出质控建议，首先根据附件
名单明确人员组织架构，如图 4-1 所示。对各层级人员违规次数、明暗测通过率等指标进行统计，进而通过这些指标完成对整体指控情况的评估与分析。

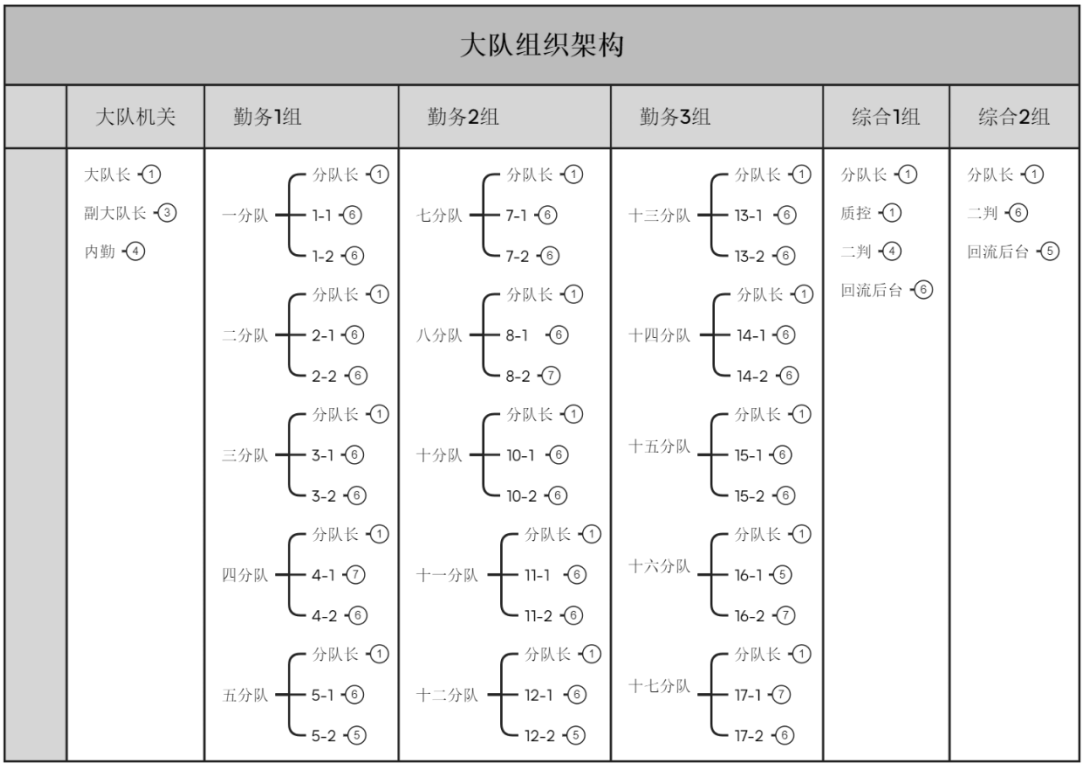

图 4.1 某旅检大队组织架构图

4.2 基于质量分数与风险收益比的岗位违规评价模型

为实现对绩效的综合度量，我们构建以下两个关键衍生指标：

1. **质量分数 (Quality Score, QS)**: 该指标旨在衡量在所有已记录事件中，积极事件所占的比率，反映了工作的“质量”或“效率”。其数学表达式为：

$$QS = \frac{PPI}{NPI + PPI}$$

QS 值域为[0, 1]，值越高，表示在同等事件总量下，该单元的正面表现越突出。

2. **风险收益比 (Risk-Reward Ratio, RRR):** 该指标用于量化获得一次“优秀”表现所需付出的“违规”代价，直观反映了工作的“风险性”。其表达式为：

$$RRR = \frac{NPI}{PPI} \quad (PPI > 0)$$

RRR 值越低，表明该单元的绩效模式越稳健、风险越低。

表 4.1 某旅检大队质检统计结果表

组别	违规次数	优秀次数	小队	违规次数	优秀次数
勤务一组	592	57	一分队	119	11
			二分队	84	10
			三分队	135	12
			四分队	149	1
			五分队	105	23
勤务二组	622	26	七分队	104	3
			八分队	149	1
			十分队	160	14
			十一分队	103	3
			十二分队	106	5
勤务三组	529	39	十三分队	92	10
			十四分队	89	4
			十五分队	116	10
			十六分队	113	9
			十七分队	119	6
综合组	40	0	综合一组	37	0
			综合二组	3	0
总计	1783	122	\	\	\

各单元在违规总数和质量分数上表现出显著的异质性。勤务二组的违规总数最高（622 次），是最大的风险源头。勤务一组的质量分数(QS)最高（0.0878），表明其在产生正面绩效方面的效率最高。相反，勤务二组的 QS 最低，意味着其工作模式缺漏较多，违规最严重的 3 个小队（十分队、四分队、八分队），占业务小队总数(17 个)的 17.6%，其违规总和为 458 次，占所有业务小队违规总数(1783 次)的 25.6%。这表明违规事件存在显著的集中效应，可能小队内人员素质较低，少数“高风险”单元是导致整体风险水平上升的主要原因。

表 4.2 各组质量分数表

组别	违规次数 (NPI)	优秀次数 (PPI)	质量分数 (QS)
勤务一组	592	57	0.0878

勤务二组	622	26	0.0401
勤务三组	529	39	0.0681
综合组	40	0	0.0000

4.3 违规事件的时间序列特征建模与对比分析

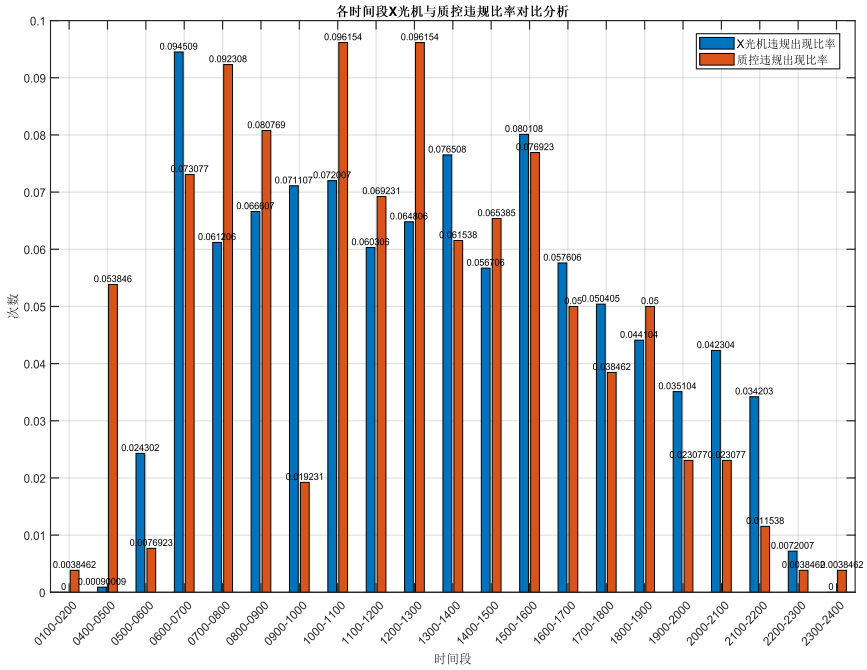

图 4.2 各时间段 X 光机与指控违规比率对比分析图

该分组柱状图直观地展示了 24 小时内不同时间段“X 光机违规出现比率”与“质控违规出现比率”的动态变化，展示了各时段产生的违规问题与总违规次数的比率。通过对比分析，旨在揭示两类违规行为在时间分布上的规律、峰谷特征以及彼此间的关联性，为质控资源的优化配置和风险预警提供数据支持。

从总体上看，两类违规比率均呈现出明显的非均匀分布特征，在特定时间段内集中爆发，而在深夜和凌晨时段则处于极低水平，这与机场旅客流量的潮汐规律基本吻合。然而，两者在峰值出现的具体时间、峰值高度以及波动模式上存在显著差异，表明它们可能受到不同驱动因素的影响。

X 光机违规呈现出典型的“双峰”分布。第一个且最高的峰值出现在 06:00-07:00 的早高峰时段，比率高达 0.0945。这通常是机场一天中首个客流高峰，安检压力骤增，同时时间较早，对于判图而言可能由于精神不集中、任务繁重而出现疏漏。午后高峰： 第二个峰值出现在 15:00-16:00 时段，比率为 0.0801，对应下午的客流高峰。

质控违规同样呈现多峰分布，但其峰值模式与 X 光机违规有显著不同。质控违规的最高峰值出现在 10:00-11:00 和 12:00-13:00 两个时段，比率均达到了惊人的 0.0962。值得注意的是，其早高峰出现在 07:00-08:00（比率 0.0923），比 X 光机违规的早高峰晚一个小时。

从总体上看，两类违规比率均呈现出明显的非均匀分布特征，在特定时间段内集中爆发，而在深夜和凌晨时段则处于极低水平，这与机场旅客流量的潮汐规律基本吻合。

计算得出，本模型的卡方统计量 $\chi^2 \approx 104.93$，在自由度 df=40 的情况下，对应的 P-value ≈ 1.34e-07。由于 P-value 远小于显著性水平 α=0.05，我们强烈拒绝零假设，得出结论：违规类型的分布与时间段之间存在高度显著的统计学关联。

接下来的分析将通过对可视化结果的深度解读，揭示这种关联的具体模式，并识别出导致这种关联性的关键风险点。

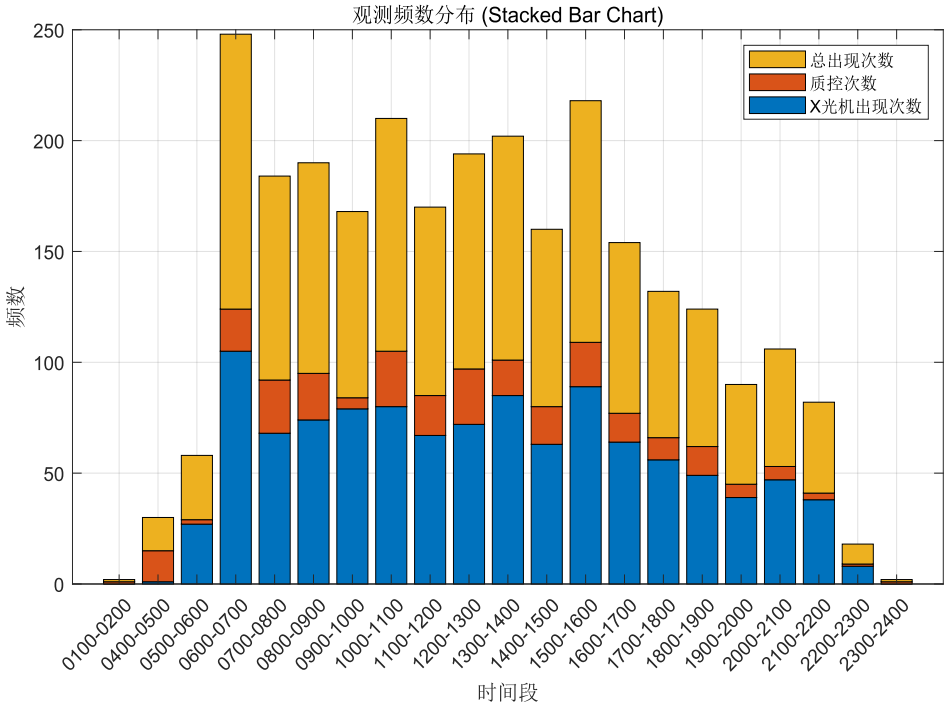

图 4.3 堆叠条形图

图 4.3 直观地展示了违规事件在 24 小时内的总体分布与内部构成。违规总频数呈现出明显的非均匀、多峰分布特征，与机场客流量的潮汐规律高度吻合。违规事件集中爆发于 06:00-17:00 的业务高峰期，而在 01:00-05:00 的深夜及凌晨时段则处于极低水平。这验证了业务负荷是驱动违规总数变化的首要外部变量。

图表揭示了不同违规类型的构成比例在不同时间段存在动态变化。在 06:00-09:00 的早高峰，蓝色部分占据了绝对主导地位，表明此阶段的主要风险集中在核心技术操作上。在 10:00-13:00 的上午平峰期，橙色部分的相对占比显著提升，甚至在某些时段（如 10:00-11:00）与蓝色部分形成鲜明对比。在 04:00-05:00 等极端低业务量时段，橙色部分（质控次数）反而成为违规的主要构成，程序性、规范性问题在低负荷下更为凸显。

堆叠条形图从宏观上证实了违规类型的分布并非一成不变，而是随着时间动态演化，这为我们拒绝 H_0 提供了直观的证据支持。然而，要精确定位导致这种关联性的“异常点”，我们需要引入卡方贡献值分析。

图 4.4 是本模型的核心输出，它量化了每个单元格（特定时间段的特定违规类型）的“异常程度”，即其观测频数与在独立假设下的期望频数之间的偏差大小。颜色越亮（数值越大），代表该点对“违规类型与时间相关”这一结论的贡献越大，是我们需要重点关注的关键风险点。

早高峰时段的“X 光机操作风险”高度集中热图中，06:00-07:00 时段与“X 光机出现次数”交叉的单元格呈现出极高亮度（卡方贡献值约为 28.26），是整个图表中最显著的异常点。这表明，在该时段观测到的 105 次 X 光机违规远超统计期望值，是导致模型显著性的首要驱动因素。这强烈暗示了在早高峰客流压力骤增的初期，X 光机操作岗位的风险被急剧放大，可能是由于人员尚未完全进入高效工作状态、设备预热或流程启动阶段的磨合问题。

深夜时段的“质控流程风险”异常凸显 04:00-05:00 时段与“质控次数”交叉的单元格同样呈现出高亮度（卡方贡献值约为 19.53）。尽管该时段的绝对违规数（14 次）不高，但其构成比例严重偏离了期望分布。

在业务量极低的情况下，本应极少出现的质控类违规（如流程、规范问题）却反常地高发。这可能揭示了在无外部压力、人员精神状态最松懈的时段，基础性的、程序性的规范执行能力是最大的短板。

上午平峰期的“风险模式转换”09:00-10:00 时段，热图呈现出有趣的反向模式：与“X 光机出现次数”交叉的单元格贡献值较低，而与“质控次数”交叉的单元格贡献值相对较高（约 6.10）。这印证了我们在宏观分析中的观察。上午时段，随着业务压力趋于平稳，X 光机操作的风险得到控制，但管理和流程上的疏漏开始成为主要的风险来源。热图通过量化的方式，精确指出了 09:00-10:00 是这种风险模式转换的关键节点。

图 4.4 卡方贡献值热图

4.4 监督模式对绩效评估有效性的影响建模分析

我们采用测试通过率 (Pass Rate, PR)作为衡量各分队绩效表现的核心指标，以避免测试总数不同带来的偏倚。其数学表达式为：

$$PR = \frac{\text{通过次数}}{\text{测试总数}}$$

通过为每个组、分队计算其在“明测”和“暗测”下的 PR，我们将得到两组对应的绩效数据向量，并利用皮尔逊相关系数模型来量化它们之间的线性关系。

设立三个研究假设：

H_1 (表现一致性假设 - 正相关): 旅检人员的真实能力是内在且稳定的。在明测中表现出色的成员，在暗测中同样会表现出色。

H_2 (“霍桑效应”假设 - 负相关或不相关): 监督模式是决定绩效的关键变量。在明测中的高通过率并不能预测其在暗测中的表现。

H_3 (模式独立性假设 - 不相关): 两种测试评估的是不同维度的能力，因此绩效无显著线性关系。

表 4.3 各组明暗测失败率表

组别	测试结果	明测	暗测
勤务一组	通过	32	11
	未通过	16	6
失败率 FR	/	33.3%	35.3%
勤务二组	通过	34	10
	未通过	8	10
失败率 FR	/	19.0%	50.0%
勤务三组	通过	32	8
	未通过	9	5
失败率 FR	/	22.0%	38.5%

相关性强度与方向: 计算得出的皮尔逊相关系数 $\rho \approx 0.32$。这是一个中等强度的正相关。其方向为正，表明明测失败率越高的组别，其暗测失败率也倾向于越高。这初步支持了 H_1 (表现一致性假设)。

统计显著性检验: P-value 高达 0.7915，远大于传统的显著性水平 $\alpha=0.05$ 甚至 $\alpha=0.10$。这意味着，尽管我们观察到了一个中等强度的正相关趋势，但由于样本量极小 (N=3)，我们无法在统计学上拒绝“两个变量不相关”的原假设。换言之，这个观测到的相关性有很大可能仅仅是由于偶然因素造成的。

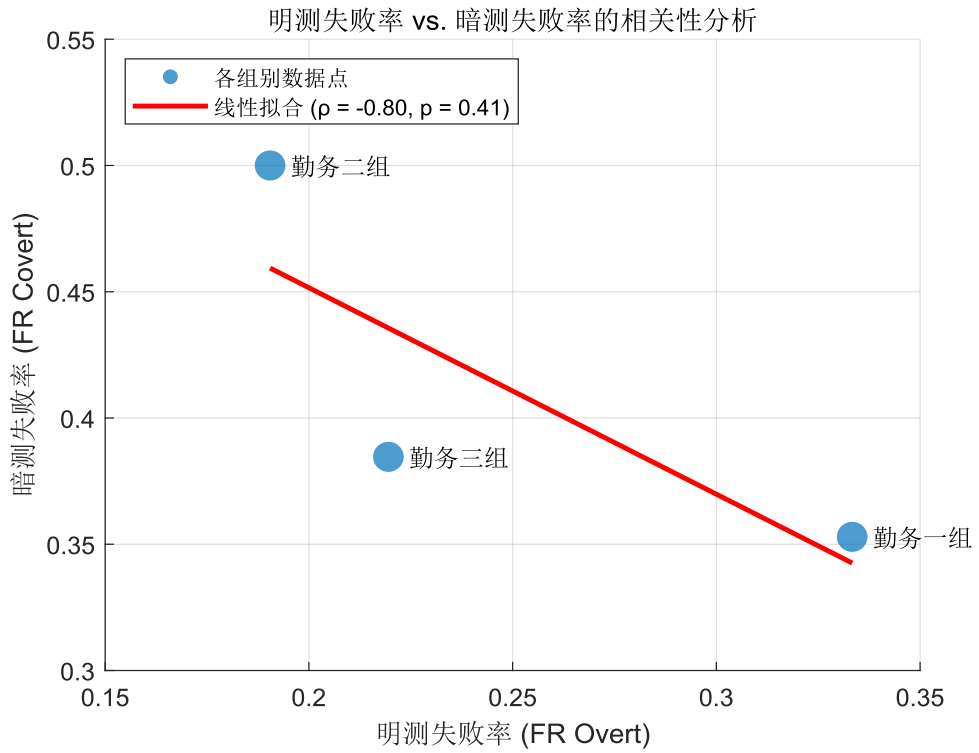

图 4.5 失败率散点图

如图 4.5 所示，勤务一组和勤务三组在两种测试模式下的失败率较为接近，表现出一定的稳定性。勤务二组则是一个显著的异常点，其明测失败率是三组中最低的，而暗测失败率却是最高的，点的存在极大地影响了相关性的分析。

表 4.4 各小队明暗测失败率表

分队	明测总数	明测通过	明测通过率 (PR_Overt)	暗测总数	暗测通过	暗测通过率 (PR_Covert)
1	8	6	75.0%	3	2	66.7%
2	7	5	71.4%	1	1	100.0%
3	10	8	80.0%	5	3	60.0%
4	16	11	68.8%	6	5	83.3%
5	12	7	58.3%	3	1	33.3%
7	6	4	66.7%	4	0	0.0%
8	10	9	90.0%	4	3	75.0%
10	7	4	57.1%	0	0	NaN (剔除)
11	11	10	90.9%	5	2	40.0%
12	9	7	77.8%	6	4	66.7%
13	9	7	77.8%	2	1	50.0%

14	5	2	40.0%	4	3	75.0%
15	10	9	90.0%	2	0	0.0%
16	7	5	71.4%	1	1	100.0%
17	4	4	100.0%	4	3	75.0%

旨在通过对 16 个独立分队的内外测数据进行建模分析，深入探究在微观组织单元层面，“明测”与“暗测”两种监督模式下绩效表现的线性相关性。与宏观组别分析相比，细化到分队级别的数据集提供了更大的样本量，从而能够得出更可靠的统计推断。

计算得出的皮尔逊相关系数 $\rho \approx -0.21$。这是一个弱负相关。其方向为负，表明明测中通过率越高的分队，在暗测中的通过率有微弱的下降趋势。这一结果初步倾向于支持 H_2 (“霍桑效应”假设)，即监督的存在可能掩盖了部分分队的真实表现。

统计显著性检验: P-value 高达 0.4503，远大于显著性水平 $\alpha=0.05$。这意味着，尽管我们观察到了一个弱负相关的趋势，但我们没有足够的统计学证据来拒绝“两个变量不相关”的原假设。这个观测到的相关性很可能是由随机波动造成的。

模型验证与可视化分析:散点图直观地展示了 15 个分队在二维绩效空间中的分布。数据点整体呈现出高度离散的状态，没有形成明显的线性趋势，这与计算出的低相关系数相符。

第 14 分队表现出显著的“明差暗优”模式，其明测通过率最低(40%)，而暗测通过率却相对较高(75%)。第 11 分队和第 15 分队则表现出强烈的“明优暗差”模式，两者明测通过率均高达 90%，而暗测通过率却分别为 40%和 0%。

这些行为模式迥异的异常点的存在，削弱了整体的线性关系，并进一步印证了不同分队对监督模式的反应是异质的。

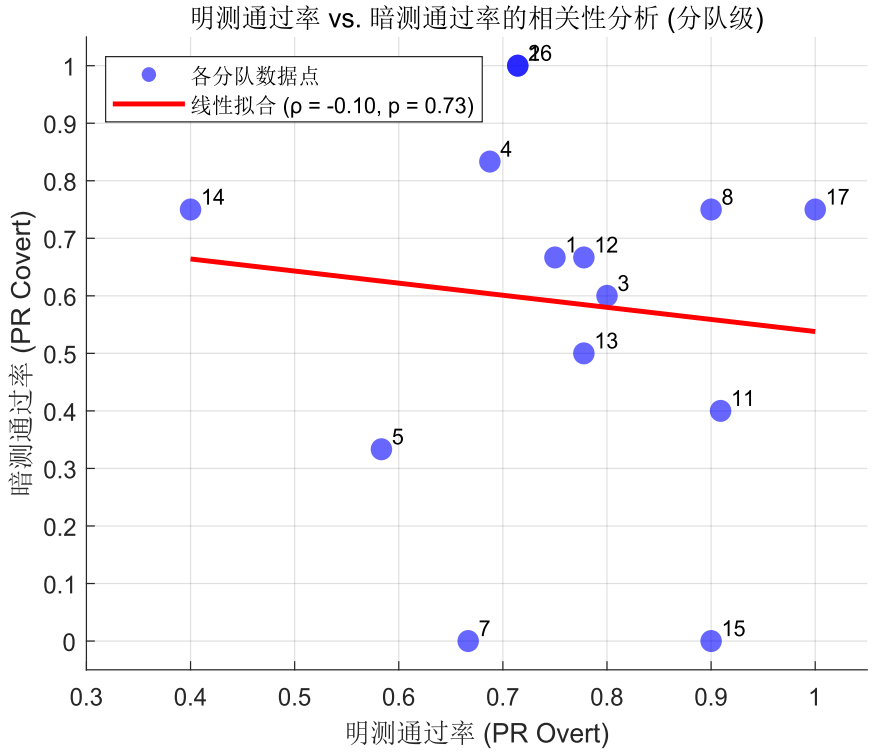

图 4.6 失败率散点图（分队级）

在绩效评估体系中，应提高暗测结果的权重，将其作为评估团队真实风险水平和工作质量的核心依据。

应为每个分队建立一个动态的“明暗测绩效对比档案”。通过追踪其在散点图中的位置变化，可以大致分为四种类型：

1、明暗测均表现优异（如 8、17 分队）。

2、明测优异而暗测较差（如 11、15 分队），需重点关注其责任心与工作习惯。

3、明测较差暗测优异（如 14 分队），可能存在应试能力不足但基础扎实、责任心强的情况。

4、 明暗测均表现不佳（如 5、7 分队），能力不足可能性较大，需要系统性培训提升，促进成长。

4.5 结论与建议

根据卡方贡献值热图揭示的风险模式，建立一个动态的督导排班系统。在 06:00-09:00 早高峰时段，派遣技术骨干或图像识别专家重点巡查 X 光机岗位，进行实时指导和质量把关。

通过上述分析，我们提出以下质控建议：

在 04:00-05:00 深夜时段，加强对基础流程、台账记录等规范性操作的突击检查。

在 09:00-10:00 风险转换节点，督导重点应从“技术操作”转向“岗位流程与服务规范”。

重构绩效评估模型，提升暗测权重 (Reconstruct Performance Evaluation Model, Increase Covert Test Weighting)目标：建立一个更能反映真实工作质量的绩效评估体系。

在个人及团队的绩效考核中，显著提升“暗测”结果的权重，降低“明测”成绩的比重，以激励员工将高标准内化为日常工作习惯，而非应试行为。

将“明暗测表现差异度”作为一个独立的负向考核指标。对于“明优暗差”表现显著的个人或团队，应视为风险警示信号，启动针对性的培训和思想教育。

鼓励并奖励那些“明暗测表现一致”且均为优秀的员工，将其作为“工作作风扎实”的典范进行表彰。通过实施以上建议，该旅检大队能够从一个“问题驱动”的被动管理模式，转型为一个“数据驱动”的主动、精准、高效的现代化质控管理体系，从而在保障安全与提升服务品质之间达到更优的平衡。

五、问题二

5.1 问题分析

本题旨在从绩效奖金分配的角度，识别影响奖金差异的关键因素，提出更公平、激励性更强的奖金分配机制，属于多属性决策问题，由于实习生不参与绩效奖金分配，我们在数据预处理中，将实习生数据剔除，并进行数据异常值处理。针对本题，我们首先提取并量化影响绩效分配的关键元素，通过均值、中位数、标准差、极差等指标描述性统计分析整体情况，接着使用皮尔逊相关系数分析过检率、绩效分数、缺勤天数与实发金额之间的相关性。为优化绩效奖金分配，提升公平性……，我们基于 APH+Tosis 方法建立了绩效奖金分配优化模型，并为绩效奖金分配提供相应的建议。

5.2 描述性统计分析

（1）实发金额

表 5.1 薪资分配统计量表

岗位	平均值	众数	方差	中位数	标准差	最小值	最大值
整体	2159.944	2064	180710.1	2084.6	425.1001	1468.1	4200
基础岗位	2001.976	1693	85553.02	1944	292.4945	1468.1	3023
开机员	2164.341	2064	65360.24	2173	255.6565	1573	2819
大队管理	3861.25	3565	77406.25	3840	278.2198	3565	4200

大队管理层的平均薪酬（3861.25）显著高于开机员（2164.34）和基础岗位（2001.98），形成了清晰的“管理-技术-基础”三级薪酬阶梯。中位数（3840, 2173, 1944）与平均值的接近，表明各层级内部的薪酬分布相对对称，不存在极端高值或低值的严重扭曲。

值得注意的是，开机员的平均薪酬略高于基础岗位，这符合技术岗位薪酬更高的普遍规律，但两者差距（约 162 元）相对较小。

方差与标准差是衡量薪酬分布稳定性的关键指标。基础岗位的方差（85553.02）和标准差（292.49）均远高于开机员（65360.24, 255.66）和大队管理（77406.25, 278.22）。

这一现象揭示了基础岗位人员的薪酬不确定性最高。换言之，同为基础岗位，个体间的最终收入差异最大。这可能源于该层级人员数量多、绩效表现差异大，或是奖惩机制在该层级的作用更为显著。这种高波动性既可能起到激励作用，也可能引发内部不公平感，是一个需要重点关注的管理信号。

岗位之间的薪酬范围（1468.1 - 3023）远宽于开机员（1573 - 2819），进一步印证了其内部收入的巨大差异。

（2）二次分配的档次

表 5.2 二次分配定档分布表

岗位	A	B	C	D	E	总人数
基础岗位	7	8	12	9	9	45
开机员	12	12	17	14	14	69
大队管理	0	0	4	0	0	4
整体	19	20	33	23	23	118

岗位	A 占比	B 占比	C 占比	D 占比	E 占比	总人数
基础岗位	0.1556	0.1778	0.2667	0.2	0.2	45
开机员	0.1739	0.1739	0.2464	0.2029	0.20299	69
大队管理	0	0	1	0	0	4
整体	0.1610	0.1694	0.2796	0.1949	0.1949	118

无论是基础岗位、开机员还是整体，人员分布均向 C 档高度集中。C 档人员占比（基础岗位 26.7%，开机员 24.6%，整体 28.0%）显著高于其他档位，形成了典型的“中间大、两头小”的纺锤形结构。

将 A、B 档视为“高收入群体”，D、E 档视为“低收入群体”。A、B 两档合计占比为 33.0% (16.1%+16.9%)。D、E 两档合计占比为 39.0% (19.5%+19.5%)。

当前的二次分配机制并未呈现出强烈的“赢家通吃”的帕累托效应，而是倾向于平均主义。大部分人员（近 1/3）集中在中间档位，高收入群体与低收入群体的比例也相对均衡。这种模式优点在于稳定性高，不易引发剧烈矛盾；缺点在于激励性不足，未能有效拉开绩优与绩差人员的收入差距，可能导致高绩效员工的激励感不强。大队管理层的二次分配档次 100%集中于 C 档，这表明二次分配机制对管理层完全无效，其薪酬主要由基本盘决定，不受此调节机制影响。

（3）不同勤务组的绩效奖金分配分布

表 5.3 不同勤务组绩效奖金定档分布表

勤务组	A 占比	B 占比	C 占比	D 占比	E 占比
勤务一组	0.1388	0.1944	0.2222	0.1666	0.2777
勤务二组	0.1470	0.2058	0.2647	0.2647	0.1176
勤务三组	0.175	0.15	0.275	0.2	0.2

勤务三组的分布模式与整体的“纺锤形”最为接近，表现为最均衡的分配结构。

勤务一组呈现出显著的“U 型”或“两极分化”特征：其最低档 E 的占比（27.8%）是三组中最高的，而最高档 A 的占比（13.9%）则是最低的。这暗示勤务一组内部的收入差距最大，且整体偏向低收入端。

勤务二组则呈现出向中高档（B、C、D）集中的趋势，其最低档 E 的占比（11.8%）显著低于其他两组。

卡方独立性检验预判:研究假设:“薪酬档次的分布与所属勤务组是否相关？”

从上表观察到的显著分布差异来看，我们有理由相信卡方检验的结果将是显著的（即拒绝“薪酬档次分布与组别无关”的原假设）。这意味着，一个员工最终落入哪个薪酬档位，不仅取决于其个人表现，还与其所在的勤务组有很强的关联。

5.3 基于多元线性回归的薪酬系数结构化建模与验证

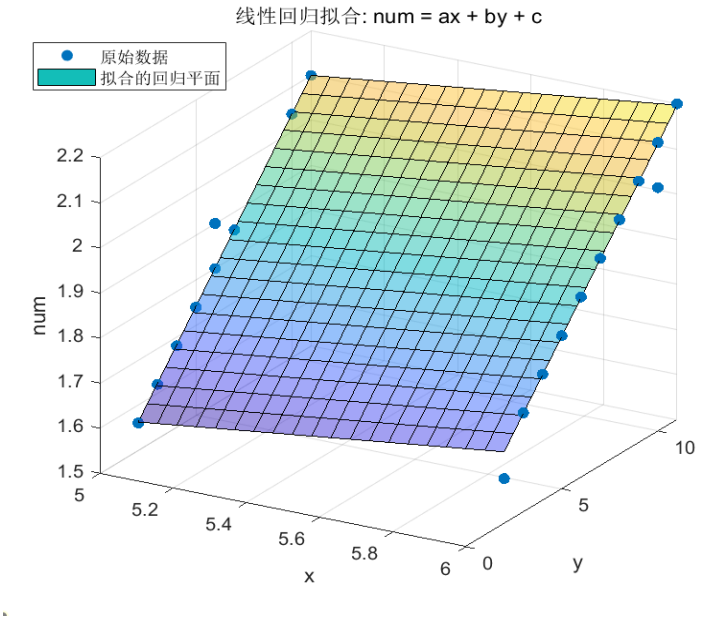

图 5.1 系数 1 的线性回归拟合模型图

为了探究并量化组织内部薪酬系数的决定机制，本研究构建了一个多元线性回归模型(Multiple Linear Regression Model)。该模型旨在揭示员工的两个核心职业属性——“现岗位分级 (x)” 和 “现薪档 (y)”——对其最终“现系数 (num)”的综合影响。

模型的理论假设是，“现系数”并非随机设定，而是由岗位级别和薪酬档位这两个自变量通过一个线性的、结构化的数学关系共同决定的。模型的数学表达式为：

$$num = a * x + b * y + c$$

基于 182 个观测样本数据，模型拟合结果如下：

最终拟合方程:

$$num = (0.0971) \cdot x + (0.0598) \cdot y + (0.9572)$$

岗位分级系数 (a = 0.0971)具有显著的统计学意义和管理学内涵。它表明，在保持薪档 (y) 不变的情况下，员工的岗位分级 (x) 每提升一级，其薪酬系数平均增加约 0.0971。这个值可以被视为对员工岗位职责、技能复杂性与经验价值的量化定价。

现薪档系数 (b = 0.0598)表明在保持岗位分级 (x) 不变的情况下，员工的薪档 (y) 每提升一档，其薪酬系数平均增加约 0.0598。这个值可以被解读为对员工在同一岗位级别内，因绩效表现、工龄或能力提升而获得的增量回报。

截距 (c = 0.9572)项代表了模型的理论起点，即一个岗位分级为 0、薪档为 0 的员工的理论基础系数。在实际应用中，它更多地是作为调整模型整体拟合度的基准值。

系数对比分析:比较两个系数的绝对值，我们发现 a (0.0971) 显著大于 b (0.0598)，其比值约为 1.62。这揭示了一个重要的薪酬设计哲学：在该组织的薪酬体系中，岗位晋升（纵向发展）带来的薪酬回报，远大于在同岗位内的档位提升（横向发展）。具体而言，提升一级岗位的价值约等于提升 1.62 个薪档。

5.3 缺勤行为的帕累托构成与长尾分布特征分析

图 5.2 通过频数直方图与核密度估计曲线，描绘了全体员工“总缺勤天数”的概率分布形态。该图揭示了一个典型的、高度显著的**“长尾分布” (Long-tail Distribution)，也称为重尾分布 (Heavy-tailed Distribution)**。

核密度估计曲线在 0 天附近形成了一个极其尖锐的、高耸的峰值（概率密度接近 0.18）。这表明，绝大多数员工的缺勤天数都非常少，集中在 0 到 10 天这个极窄的区间内。这些员工构成了组织的“稳定核心”。在 20 天之后，概率密度迅速下降并趋近于零，但并未完全消失，而是形成了一条“厚重”且一直延伸到 120 天的长尾。虽然处于尾部的员工人数极少（由直方图的低矮条形块可知），但他们每个人都贡献了极高的缺勤天数。

缺勤行为在员工个体间并非正态分布，而是存在极端的不均衡。组织的整体缺勤率并非由“平均水平”决定，而是被少数处于“长尾”末端的高缺勤员工严重拉高。

这个模型清晰地告诉我们，管理策略必须是二元化的：对于“头部”的大多数员工，管理策略应以维持和激励为主，确保他们继续保持高出勤率。

对于“尾部”的少数员工，需要对这些员工进行个体化的深度分析，探究其长期缺强背后的具体原因（如慢性病、特殊工伤等），并提供针对性的支持与管理方案。

“长尾分布”的一个重要数学特征是均值的高度敏感性。少数极端值（长尾）会严重扭曲平均数，使其无法真实反映群体的典型情况。因此，在制定基于缺勤的绩效评估或奖金分配政策时，使用“平均缺勤天数”作为衡量标准是极具误导性的。应更多地考虑中位数，或者设计能够稳健处理极端值的非线性模型。

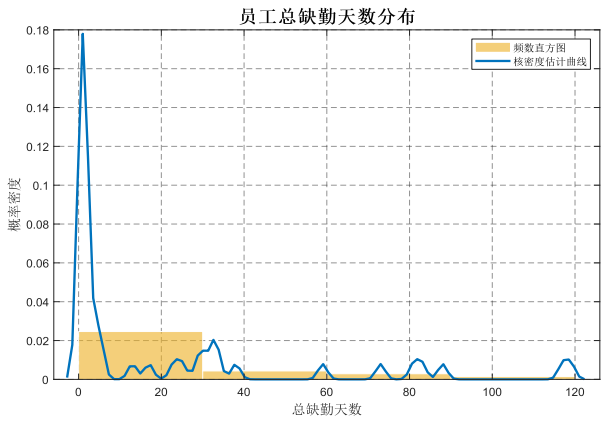

图 5.2 员工总缺勤天数分布图

图 5.3 通过饼图与水平条形图的组合，直观地展示了不同缺勤类型的总天数构成。该图揭示了一个极其显著的**帕累托效应**，即所谓的“80/20 法则”在此处表现得淋漓尽致。

病假总天数高达 935.5 天，在所有缺勤事件中占比达到了压倒性的 99.7%。事假总天数仅为 3 天，占比仅 0.3%。病假的总天数是事假的 311.8 倍，两者在量级上存在巨大差异。 数据清晰地表明，当前组织面临的主要缺勤问题几乎完全是“病假”问题。任何旨在降低整体缺勤率的管理干预，如果不能有效针对病假，其收效将微乎其微。这为问题的定性与资源投入指明了方向。“病假”通常被视为一种非主观意愿的、不可抗力的缺勤。其占比畸高，暗示了组织的缺勤风险主要来源于员工的健康状况，而非工作态度或纪律松懈。这与以“事假”为主导的缺勤模式相比，其管理策略的着力点应有根本性的不同。

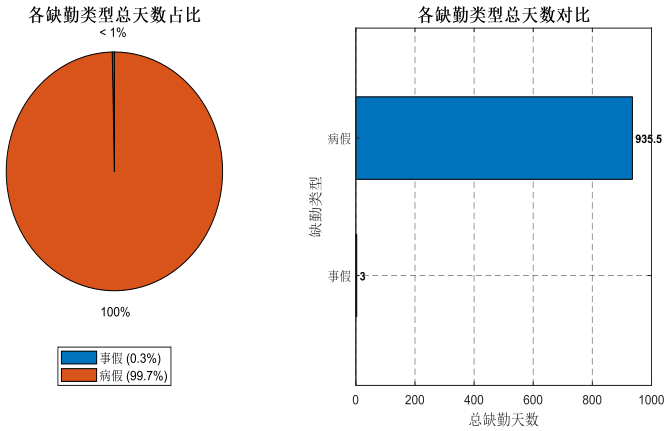

图 5.3 缺勤类型统计饼图与直方图

5.4 薪酬生成机制的结构化建模与偏差分析

我们观察到基数与在岗天数 1 存在严格的线性关系。我们建立模型：

$$\text{基数} = \alpha \cdot \text{在岗天数}$$

通过数据(120, 1200), (115, 1150)可精确求解出 $\alpha = 10$

基于第一层模型，我们构建标准额的决定模型。标准额是基数和系数 1 的函数。

$$\text{标准额} = \text{基数} \cdot \beta = (\alpha \cdot \text{在岗天数}) \cdot \beta$$

其中 β 为系数 1，这个模型定义了在不考虑任何扣除项的情况下，一个员工的理论应得薪酬。它由两个核心变量驱动：出勤（在岗天数）和个人价值（系数）。

如图 5.4 所示的彩色的理论薪酬曲面代表了由我们的核心公式定义出的理论薪酬空间。红色的实际数据点代表了表格中每个员工的真实情况。我们可以清晰地看到，几乎所有的红色数据点都精确地、无偏差地落在了理论曲面上。这为我们的薪酬基准模型提供了强有力的视觉证明。“标准额”计算规则是高度确定和刚性的。不存在模糊地带或人为调整空间，其透明度和公平性在这一环节得到了保证。薪酬增长路径清晰，由“在岗天数”和“个人系数”决定。员工可以通过保证出勤和提升个人价值（通过晋升、技能提升等方式提高系数）来获得更高的理论薪酬。

尽管“缺勤天数”是扣罚的主要依据，但 P27200 的案例（73 天缺勤，0 扣罚）揭示了存在一个更上位的、隐藏的规则。结合缺勤原因列（病假 33 天；产假 40 天），我们可以大胆推断：并非所有类型的缺勤都会触发扣罚机制。特别是产假、长期病假等受法律保护或公司政策支持的缺勤，可能被豁免或采用另一套完全不同的计算方法。

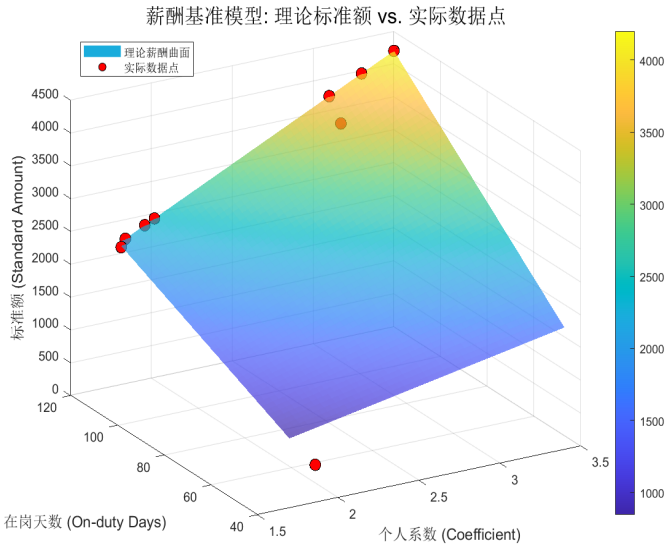

图 5.4 薪酬基准模型图

5.5 结论与建议

基于上述分析，我们对绩效奖金分配提出以下建议：

1.重构“二次分配”模型，强化激励梯度

建议构建综合绩效得分 (Composite Performance Score, CPS): 放弃固定的档位，改为计算每个员工的连续性综合绩效得分。参考“4.3 薪酬系数结构化模型”的思路，使用层次分析法 (AHP) 或专家打分法，为二次分配的各项指标（高峰期过检率、查堵得分、质控得分等）确定一个明确的、公开的权重 w。

$$CPS_i = w_1 \cdot (\text{过检率}_i) + w_2 \cdot (\text{查堵分}_i) + w_3 \cdot (\text{质控分}_i) + \dots$$

引入非线性分配函数: 将 30%的奖金池，根据每个员工的 CPS 得分，通过一个非线性函数进行分配，以拉开差距。例如，可以使用平方分配法

$$\text{个人二次分配奖金}_i = (\text{总奖金池}) \cdot \frac{(CPS_i)^2}{\sum_{j=1}^{n} (CPS_j)^2}$$

在这种模型下，绩效得分是别人两倍的员工，其获得的奖金将是别人的四倍，从而极大地放大了激励效果。

2.明确并量化缺勤扣罚的差异化规则

将缺勤明确分类，并设定不同的扣罚系数 k。主观意愿类缺勤 (如事假、无故缺勤)设置高惩罚系数 (例如 k=1.5)。不可抗力类短期缺勤 (如短期病假): 设置标准系数 (例如 k=1.0)。受保护类长期缺勤 (如产假、经核定的长期病假): 设置零或极低的豁免系数 (例如 k=0.1)，或直接不纳入 70%部分的考勤扣罚计算，改为发放基本生活补贴。

将此规则明确写入分配方案，消除模糊地带。

3.引入“风险调整系数”，实现横向公平

消除因岗位/任务难度不同而造成的系统性评价偏差，对岗位难度评估:，基于历史数据，对不同勤务组或不同安检通道的工作负荷（如人均过检量）、复杂性进行量化评估，得出一个岗位风险/难度系数 (Difficulty Adjustment Factor, DAF)。例如，以全大队平均水平为 1，最繁忙的勤务一组的 DAF 可能为 1.2，较清闲的组可能为 0.9。

在计算二次分配的综合绩效得分 CPS 时，对部分指标进行校准。

$$\text{校准后过检率} = \frac{\text{实际过检率}}{DAF}$$

通过这种方式，勤务一组的员工即使过检率略低于其他组，其校准后的得分也可能更高，从而补偿了其承担的额外压力，实现了更深层次的公平。

六、问题三

6.1 问题分析

问题三要求从安全漏洞、管理能效、资源分配三个角度定量分析该大队存在的问题，属于多属性绩效评估与诊断问题。传统的单一指标分析（难以全面、综合地反映一个复杂系统的整体效能。因此，需要一种能够同时处理多重影响因素、多类输出成果的综合评价方法。基于此，我们选择 DEA 方法作为本问的核心分析工具。我们将该大队下辖的勤务一组、勤务二组、勤务三组定义为三个决策单元（DMU）从安全漏洞、管理能效、资源分配进行效率评估。

6.2 基于 DEA 的运行效率评估模型构建

6.2.1 数据包络分析（DEA）

数据包络分析（Data Envelopment Analysis, DEA）是一种非参数的线性规划方法，用于评价具有多输入多输出的决策单元（DMU）之间的相对有效性。其核心是构建一个“前沿面”，有效率的 DMU 位于此前沿面上，效率值为 1；无效率的 DMU 则位于前沿面内部，效率值介于 0 和 1 之间。

本研究选用 BCC 模型（Banker, Charnes, and Cooper, 1984），该模型假设规模报酬可变（VRS），可将综合技术效率（TE）分解为纯技术效率（PTE） 和规模效率（SE），即：

$$TE = PTE \times SETE = PTE \times SE$$

● 技术效率（TE）：衡量在最优规模下的投入产出效率，TE=1 表示 DMU 同时达到技术有效和规模有效。

● 纯技术效率（PTE）：剔除规模因素，纯粹反映 DMU 的管理和技术水平，PTE=1 表示技术有效。

● 规模效率（SE）：反映 DMU 的规模是否处于最优状态，SE=1 表示规模有效。

BCC 模型数学形式（投入导向）：

对于每个 DMU，我们求解以下线性规划问题：

$$\begin{aligned}
&\min_{\theta,\lambda} \theta \\
\text{s.t.} \quad &\sum_{j=1}^{n} \lambda_j x_{ij} \le \theta x_{i0}, \quad i = 1, \dots, m \\
&\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{r0}, \quad r = 1, \dots, s \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \ge 0, \quad j = 1, \dots, n
\end{aligned}$$

其中：

- θ 为 DMU_0的效率值；
- x_{ij}表示第j个 DMU 的第i项投入；
- y_{rj}表示第j个 DMU 的第r项产出；
- $lambda_j$为权重变量；
- 约束$\sum \lambda_j = 1$ 确保模型为 VRS。

6.2.1 指标选取与数据说明

根据附件数据，我们将三个勤务组（勤务一组、二组、三组）作为决策单元（DMU），构建如下输入输出指标体系：

图 6.1 DEA 分析输入输出指标体系

类型	指标名称	变量符号	计算方式	含义
投入 Inputs	在岗人员数量	I_1	各组总人数	人力资源投入
	总工作天数	I_2	∑(每人出勤天数)	时间资源投入
	绩效奖金总额	I_3	各组奖金总和	资金资源投入
期望产出 Outputs	总过检旅客量	O_1	估算值（基于过检率）	工作量的正面产出
	平均过检率	O_2	各组过检率均值	工作质量的正面产出
	平均绩效分数	O_3	各组绩效分均值	个人绩效的正面产出
非期望产出 Undesirable Outputs	岗位违规总次数	U_1	各组违规记录求和	安全漏洞/管理失误
	重点违规次数	U_2	高风险违规记录数	重大安全风险

通过数据预处理，三个勤务组各自的指标数值如下：

其中，过检旅客量为该勤务组 4 到 6 月过检旅客数总和，重点违规次数为该勤务组中高风险违规次数。

DMU	I_1 人数	I_2 总工时(天)	O_1 过检旅客量(估算)	O_2 平均过检积分(%)	O_3 平均绩效积分	U_1 违规次数	U_2 重点违规次数
勤务一组	65	310	68715.5	84.25694	78.57371	197	6

勤务二组	65	321	66831.5	82.64216	73.39779	131	2
勤务三组	66	301	68468.5	82.49375	84.70438	175	2

6.2.3 DEA 效率计算及分析

使用 Python 对上述数据进行 BCC 模型计算，结果如下：

决策单元	技术效益（PTE）	规模效益（SE）	综合效益（TE）	松弛变量 S-	松弛变量 S+	有效性
勤务一组	1.000	1.000	1.000	0.000	0.000	DEA 强有效
勤务二组	1.000	0.981	0.981	10.789	570.741	非 DEA 有效
勤务三组	1.000	1.000	1.000	0.000	0.000	DEA 强有效

表中，技术效益反映在当前规模下决策单元的技术效率；规模效益反映决策单元是否处于最优生产规模；综合效益反映总体效率水平；松弛变量 S-标识投入可减少量，松弛变量 S+表示产出可增加量。

6.3 结果分析

首先从安全漏洞、管理能效、资源分配方面对结果进行分析：

1.安全漏洞方面

效率值最低的勤务二组，其单位工时的过检旅客量最少。勤务一组的违规次数（U_1=197）远高于其他两组，其安全率最低。

在质控结果不合格的时间段中，早上 6：00-7：00，下午 13：00-14：00，15;00-16：00 为高发时段。

在各个质控不合格的岗位中，不合格率最多的是日常和安全门岗位。

2. 管理能效方面

三个勤务组的纯技术效率差不多，PTE 均为 1 ，表明其在管理流程、人员培训、操作规程执行等方面基本一致。该大队整体平均纯技术效率为 1，管理效率存在一定提升空间，可实现更大的效益。

在问题类型中，出现问题最多的是 C1，即有完备的制度/程序/标准单未执行。

在各分队中，第七分队的内部测试通过率最低，第八分队在实际工作中的犯错频次最高。

第五分队和第八分队整改后，平均风险等级增加。

3.资源分配方面

勤务二组规模效益略低于一，说明规模略有不足或者过大，综合效益为 0.981，未达到有效，S-为 10.789 表示投入有冗余，S+为 570.741，表示产出严重不足。因此，勤务二组存在规模不经济和产出不足的问题，应优化资源配置或调整规模。

6.4 结论与建议

本研究运用 DEA-BCC 模型对某旅检大队三个勤务组进行了效率评估，从安全漏洞、管理能效、资源分配三个维度分析了当前存在的问题：勤务二组管理效能低下导致安全风险高发；勤务一组规模过大，资源利用率不高；整体存在一定的管理效率提升空间。研究成果为该大队的安全管理优化和资源精准配置提供了定量依据。

基于上述分析，我们提出以下管理建议：

对勤务二组开展专项整顿培训，重点提升安检操作规范性，降低违规次数；同时优化其资源分配模式，在提升管理效能的基础上适当增配资源。

建立长效效率监测机制，定期开展 DEA 效率评估，将效率指标与绩效考核挂钩，实现资源动态优化配置。

针对早上 6:00-7:00、下午 13:00-14:00、15:00-16:00 等高发时段，临时增派质控员、流动班长进行高频次巡查和提醒，打破固定的检查模式。

鉴于日常和安全门岗位不合格率最高，应对这两个岗位增加专项技能强化培训。

针对有制度但不执行的问题，加强人员管理，从大队领导到分队长，层层加强管理。

针对整改后风险等级提升的问题，建议大队管理层成立专项小组定期检查，并提高整改后风险等级提升的罚金。

七、问题四

7.1 问题分析

问题四要求优化设计数据收集表格，并将定性与定量相结合，构建系统分析模型，从整体上对该大队的安检系统脆弱性进行评估，并给出绩效奖金分配建议。

7.2 数据收集表格优化

在完成前面问题的过程中，我们可以感受到，现有的表格体系是为**手工记录和逐级上报**设计的，存在数据孤岛、结构不一、冗余严重等问题，无法有效支持系统性的脆弱性评估和精准的绩效分配。

以下是对这些表格的优化方案、原因及建议。

7.2.1 优化核心思路

数据融合与关联：打破表格孤岛，建立以人员和分队为核心的数据模型，使所有数据可以通过关键 ID（如员工 ID、分队 ID、日期）进行关联分析。

结构化与标准化：将非结构化的文本描述（如“原因分析”、“整改跟踪”）转化为结构化的分类数据（如使用下拉菜单选择），便于量化分析。

减少冗余：相同的基础信息（如人员、分队、岗位）只在一个主表中维护，其他表格通过 ID 引用。

为分析服务：表格设计可直接服务于数据分析的指标计算和绩效奖金分配算法的需求。

7.2.2 优化方案详情

新增核心主表《人员主数据表》。原架构中人员信息分散在多个表，存在不一致和冗余，此表将作为所有数据的人员维度基础。新增的员工 ID 是唯一标识，用于关联所有其他业务表。

简化并标准化《组织架构表》。原表仅为列表，无法关联。此表定义组织单元，分队长 ID 关联至人员主数据表，可准确追溯分队长责任。

重构《安全事件记录表》。原《质控情况汇总表》内的三张表本质都是记录“安全事件”，结构相似但割裂，难以从整体记录分析一个分队或人员的违规趋势，且存在重复记录、错记漏记现象。建议用事件类型和问题类型字段区分不同来源和具体问题，问题类型必须使用标准分类（如：漏查、程序不符、迟到、误判等），淘汰自由文本，这是后续量化分析（计算频次、类型占比）的基石。 通过员工 ID 和分队 ID 关联回主数据，可轻松聚合计算“高发岗位占比”、“违规频次”、“内测违规率”、“事件等级”等脆弱性指标，便于动态监控大队质控情况。

整合《绩效与考勤记录表》。将原先的《绩效奖金计算过程》、《绩效奖金分配明细》、《绩效奖金发放表》、《实习生考核》等多张表整合成《绩效奖金核算表》。以上多张表记录同一件事（月度绩效）的不同侧面，流程割裂，可整合为一张核心表，记录每月每人绩效的计算过程和最终结果。结合绩效奖金分配方案，二次分配原因必须标准化（如：奖励-查获违禁品、扣罚-违规、扣罚-缺勤、团队激励等），这是确保奖金分配公平性和分析绩效和奖金关联度的关键。

新增《员工考勤明细表》。原缺勤数据散落在绩效表或备注中，无法精确分析，建立独立的考勤流水表，可精准计算每个员工、每个分队的“缺勤天数”和“缺勤类型”，为分析评估提供准确数据。通过员工 ID 可关联其所属分队，用于计算分队整体绩效和奖金情况。

表 7.1 表格优化方案

表名	主要字段
人员主 数据表	员工 ID、姓名、岗位、所属大队、所属分队、所属小组、 员工类型、入司时间、岗位系数、X 光机操作资质

组织 架构表	分队 ID、分队名称、勤务组、分队长 ID
安全事件记录表	事件 ID、日期、员工 ID、分队 ID、时间、地点、事件类型（岗位违规/X 光违规/内部测试）、问题类型（标准化分类）、严重程度、风险 R 值、是否下发整改、整改完成状态、验证结果
分队过检率日记录表	记录 ID、日期、分队 ID、过检率
绩效奖金核算表	记录 ID、年月、员工 ID、初始金额、过检率得分、绩效分数、考勤扣罚、二次分配金额、最终金额、二次分配原因（标准化分类）
员工考勤明细表	记录 ID、日期、员工 ID、班次、缺勤类型（事假/病假/旷工等）、缺勤时长

7.2.3 实施建议

一是分步实施。根据程度，首先应当推行主数据表和安全事件记录表的标准化。二是结合工具支持。即使暂时不能开发专业系统，也应使用专业工具来模拟关系型数据库，强制数据规范录入，可考虑结合手机小程序，app 等进行开发，推进无纸化办公。三是培训与文化。对分队长和质控员进行培训，让他们理解数据标准化录入的重要性，这直接决定了后续分析的质量和可靠性。

通过以上优化，数据将从繁琐的“记录”资产，转变为有价值的“分析”资产，真正为管理决策提供强大支撑。

7.3 基于 VSD 理论框架的系统脆弱性评价模型构建

7.3.1 系统脆弱性评估模型构建

为科学评估机场安检大队运行系统的脆弱性（Vulnerability），提升其抵御违规事件与内部扰动的能力，本研究基于 VSD（Vulnerability Scoping Diagram）理论框架，构建了一个融合“暴露度（Exposure）-敏感性（Sensitivity）-适应能力（Adaptive Capacity）”三维度的综合评价模型，模型将系统脆弱性定义为：系统因自身结构特性与外部环境扰动共同作用而导致功能受损的潜在属性。我们通过建立融合熵权法（Entropy Weight Method）与 TOPSIS（Technique for Order Preference by Similarity to Ideal Solution）的综合评价模型，结合安检大队运行实际，对 15 个分队的脆弱性进行精准量化排序；并引入障碍度模型（Obstacle Degree Model），诊断制约系统稳健性的关键指标，为提升安检系统韧性提供数据驱动的决策依据。评价指标体系如表 1 所示，包含 3 个要素层及 12 个具体指标，每个指标均明确其指向（效益型或成本型）。

表 7.2 机场安检系统脆弱性评价指标体系

目标层	要素层	指标层	属性	说明
安全漏洞	违规岗位分析	a11 高发岗位占比	-	反映违规行为的集中程度
		a12 违规频次	-	反映违规事件的发生频率
		a13 内测违规率	-	反映内部管控的有效性
	违规事件分析	a21 事件等级（平均风险值）	-	反映违规事件的严重程度
		a22 违规类型多样性	-	反映违规模式的复杂程度
管理能效	分队人员工作量安排	b11 分队过检率	+	反映核心工作效率与负荷
	问题整改效果	b21 内部测试整改变化率	-	反映学习与改进能力
	员工缺勤情况	b31 缺勤天数	-	反映队伍稳定性
		b32 缺勤类型（加权值）	+	反映缺勤的结构性影响
	岗位绩效计算	b41 开机员与基础岗位绩效对比	+	反映激励制度的合理性
资源分配	奖金分配	c11 二次分配金额	+	反映激励资源的投入水平
	岗位配置	c21 分队内人员配置（X 光机/基础岗）	+	反映人力资源的结构优化
		c22 过检轮班设置合理性	+	反映勤务安排的科学程度

7.3.2 熵权法计算指标权重

为消除人为主观因素干扰，采用客观赋权法——熵权法确定指标权重。熵值越小，表明指标值的变异程度越大，提供的信息量越多，其权重也应越大。计算过程如下：

(1) 计算第j项指标下，第i个评价对象的特征比重P_{ij}：

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{m} Y_{ij}}, \quad (i = 1, 2, \ldots, m; , j = 1, 2, \ldots, n)$$

其中，Y_{ij} 为标准化后的数据矩阵。

(2) 计算第j项指标的信息熵值e_j：

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^{m} P_{ij} \ln(P_{ij})$$

规定当$P_{ij} = 0$时，$P_{ij} \ln(P_{ij}) = 0$。

(3) 计算第 j项指标的差异系数g_j：

$g_j = 1 - e_j$ 差异系数越大，表明该指标在综合评价中的重要性越强。

(4) 确定各指标的熵权 w_j：

$$w_j = \frac{g_j}{\sum_{j=1}^{n} g_j}$$

7.3.3 基于 TOPSIS 法的综合评价

TOPSIS 法通过计算评价对象与理想解的相对贴近度进行排序。具体步骤如下：

(1) 构建加权规范化决策矩阵（Z）：

$$Z = (z_{ij})m \times n = w_j \cdot Yij$$

(2) 确定正理想解Z^+与负理想解 Z^-：

其中，J_+为效益型指标集，J_-为成本型指标集。

(3) 计算各评价对象到正、负理想解的欧氏距离D_i^+与D_i^-：

$$D_i^+ = \sqrt{\sum_{j=1}^{n} (z_{ij} - z_j^+)^2}$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} (z_{ij} - z_j^-)^2}$$

(4) 计算各评价对象的相对贴近度 C_i：

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

其中，$C_i \in [0,1]$，其值越接近于 1，表明该分队越接近正理想解，即系统脆弱性越高；反之，脆弱性越低。

7.3.4 引入障碍度模型诊断障碍因子

为识别制约系统脆弱性降低的关键因素，引入障碍度模型。该模型通过计算因子贡献度（即指标权重 w_j）、指标偏离度（ $1-Y_{ij}$ ）和障碍度O_{ij}来实现：

$$O_{ij} = \frac{(1-Y_{ij}) \cdot w_j}{\sum_{j=1}^{n}[(1-Y_{ij}) \cdot w_j]} \times 100$$

O_{ij} 值越大，表明该指标是制约评价对象 i 脆弱性降低的主要障碍因子。

7.4 实证分析与结果讨论

7.4.1 权重分析

利用熵权法计算出的各指标权重分布如图 7.1 所示。由图可知，“内测违规率(a13)”与 “缺勤天数(b31)” 的权重显著高于其他指标，表明这两个指标在评估系统脆弱性时提供的信息量最大，其波动对最终评价结果的影响最为敏感。

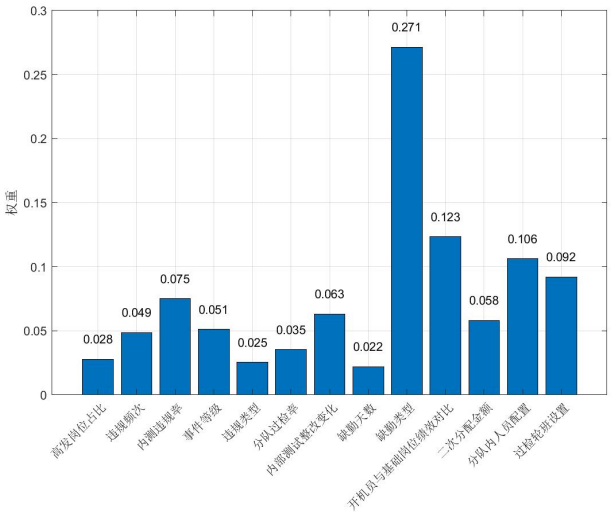

图 7.1 指标权重分布图

7.4.2 脆弱性排序与空间分布

15 个分队的脆弱性贴近度C_i排序如图 7.1 所示。分队间脆弱性水平差异显著，其中分队 1 的脆弱性最高（$C_1 = 0.645$），而分队 5、8、11 等则表现出了较好的稳健性。这直观反映了不同分队在安全管理、资源配置、运行效率等方面存在的不平衡性。

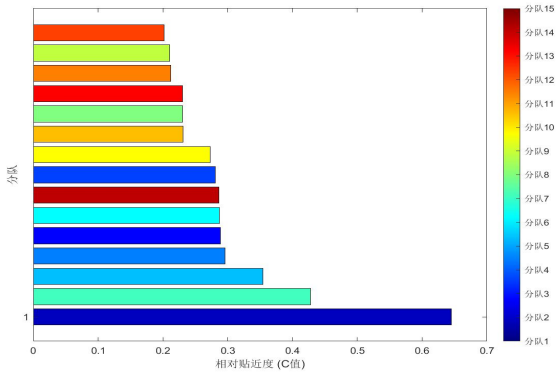

图 7.2 各分队安检系统脆弱性排序

7.4.3 关键障碍因子诊断

根据障碍度模型计算结果图 7.2，制约各分队脆弱性降低的障碍因子存在显著差异。例如：分队 3 的脆弱性首要障碍因子为“内测违规率(a13)”（障碍度达 24.4%），暴露出其内部质量控制机制存在严重短板，需重点关注整改措施的落实与有效性验证。

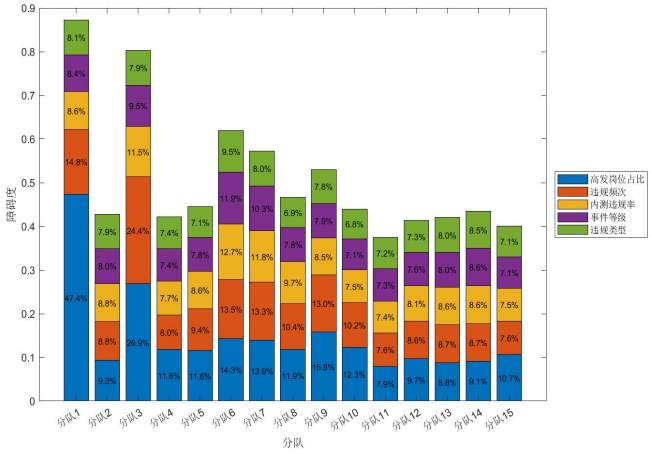

图 7.3 障碍因子诊断图

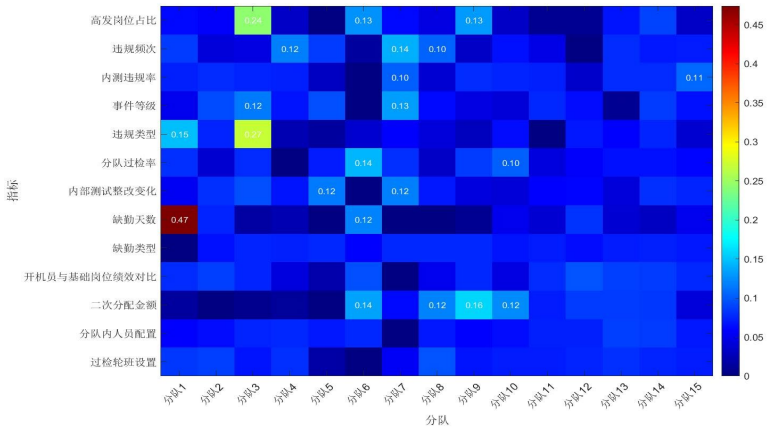

图 7.4 障碍度热力图

分队 1 则主要受“缺勤天数(b31)” 和 “缺勤类型(b32)” 的困扰，高缺勤率直接导致了队伍不稳定、人力不足，从而增加了运行风险，反映出其在人员排班、福利保障或团队凝聚力建设方面可能存在缺陷。

7.4.4 四象限策略分析

图 3 的“绩效-脆弱性”四象限分析为进一步制定差异化管控策略提供了清晰指引：第一象限（高绩效-高脆弱性）：如分队 X，虽当前绩效表现良好，但系统脆弱性高，犹如“高压下的繁荣”。需警惕过度透支带来的风险，重点保障人员休息，优化流程以维持可持续性。

第二象限（低绩效-高脆弱性）：如分队 1、3，此为“重点整改区”，必须进行全面干预，从人员、管理、资源多方面系统性地降低脆弱性、提升绩效。

第三象限（低绩效-低脆弱性）：如分队 Y，属于“潜力待激活区”。应在保持当前稳健性的基础上，通过技能培训、激励机制等措施挖掘潜力，提升绩效。

第四象限（高绩效-低脆弱性）：如分队 5、8，是“标杆示范区”。其管理实践应被总结和推广，作为其他分队学习的榜样。

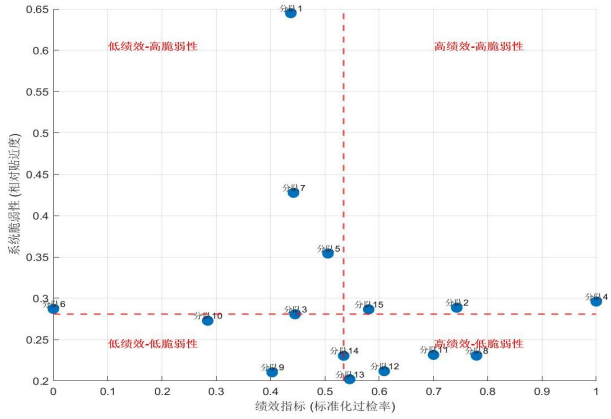

图 7.5 脆弱性四象限散点图

7.5 结论与建议

综上，实证结果总结如下：

1. 系统脆弱性在不同分队间呈现显著差异性，需实施差异化管控。

2. “人”的因素（如缺勤、内测违规）是制约系统韧性的最关键障碍。

3. 四象限分析法能为管理者提供清晰的决策矩阵，实现精准施策。

据此，提出以下对策建议：

1.对于高脆弱性分队：立即开展针对性的障碍因子整治，如优化排班制度以降低缺勤影响、加强质量控制与培训以降低内测违规率。

2.对于全体系统：建立健全基于脆弱性评估结果的动态监测与预警机制，将熵权 TOPSIS 模型与障碍度诊断嵌入日常安全管理流程，实现从“事后补救”向“事前预警、事中控制”的转变。

推广最佳实践：总结低脆弱性高分队的成功管理经验（如有效的团队激励、高效的轮班设置等），并在整个安检大队内进行推广，全面提升系统整体韧性。

附录 A，B

A-1 数据预处理代码

```
import pandas as pd
import re
from collections import Counter


df = pd.read_excel('1X 光机质控总表.xlsx')
problem_column = '时间段'
problems = df[problem_column].dropna()  # 删除空值
time_pattern = re.compile(r'(\d{4})-(\d{4})')
time_ranges = []
for problem in probAlems:
    matches = time_pattern.findall(str(problem))
    for match in matches:
        start_time, end_time = match
        time_range = f"{start_time}-{end_time}"
        time_ranges.append(time_range)
time_range_counts = Counter(time_ranges)
sorted_time_ranges = sorted(time_range_counts.items(), key=lambda
x: x[1], reverse=True)
print("时间段出现次数统计:")
print("时间段\t\t 出现次数")
for time_range, count in sorted_time_ranges:
    print(f"{time_range}\t{count}")
total_count = sum(time_range_counts.values())
```

```
    print(f"\n 总共发现了 {len(time_range_counts)} 个不同的时间段，总计 {total_count} 次出现")
    results_df = pd.DataFrame(sorted_time_ranges, columns=['时间段', '出现次数'])
    results_df.to_csv('时间段统计.csv', index=False, encoding='utf-8-sig')
    print("\n 结果已保存到 '时间段统计.csv' 文件")
```

```
    import pandas as pd
    absence_df = pd.read_excel('绩效奖金_缺勤.xlsx')
    bonus_df = pd.read_excel('绩效奖金_总表.xlsx')
    absence_df.rename(columns={'姓名': '姓名'}, inplace=True)
    bonus_df.rename(columns={'Unnamed: 2': '姓名', 'Unnamed: 8': '最终实发总额'}, inplace=True)
    merged_df = pd.merge(absence_df, bonus_df[['姓名', '最终实发总额']], on='姓名', how='left')
    merged_df.to_excel('绩效奖金_更新.xlsx', index=False)
    print("更新完成，结果已保存到'绩效奖金_更新.xlsx'")
```

```
    import pandas as pd
    import numpy as np
    file1_path = '1-4 月绩效奖金计算过程（2）.xlsx'
    overview_df = pd.read_excel(file1_path, sheet_name='过检率积分')
    file2_path = '组织构架及人员名单.xlsx'
    org_df = pd.read_excel(file2_path, sheet_name='Sheet1')
    jan_data = overview_df.iloc[1:, [0, 1]].dropna()
```

```
    feb_data = overview_df.iloc[1:, [3, 4]].dropna()
    mar_data = overview_df.iloc[1:, [6, 7]].dropna()
    apr_data = overview_df.iloc[1:, [9, 10]].dropna()
    jan_data.columns = ['姓名', '1 月过检率积分']
    feb_data.columns = ['姓名', '2 月过检率积分']
    mar_data.columns = ['姓名', '3 月过检率积分']
    apr_data.columns = ['姓名', '4 月过检率积分']
    merged_df = jan_data.merge(feb_data, on='姓名', how='outer')
    merged_df = merged_df.merge(mar_data, on='姓名', how='outer')
    merged_df = merged_df.merge(apr_data, on='姓名', how='outer')
    org_info = org_df[['勤务组', '分队', '小组', '姓名']].copy()
    final_df = merged_df.merge(org_info, on='姓名', how='left')
    final_df = final_df.fillna('')
    column_order = ['姓名', '勤务组', '分队', '小组', '1 月过检率积分', '2
月过检率积分', '3 月过检率积分', '4 月过检率积分']
    final_df = final_df[column_order]
    output_path = '人员过检率积分汇总.xlsx'
    final_df.to_excel(output_path, index=False)
```

B-1 数据预处理代码

```
clc;
clear;

% 文件路径
filepath = 'E:\202508 湖南省数模竞赛\代码\';
file_qc4 = '总_质控情况汇总表（清洗后）.xlsx';
fileName = [filepath, file_qc4];
sheetName_all={'岗位违规','X 光机','内部测试'};
events_Position = struct();
```

```
events_XRay = struct();
events_Test = struct();
for t=1:length(sheetName_all)
    sheetName=string(sheetName_all(t));
switch sheetName
    case '岗位违规'
        %% 设置导入选项
            opts = detectImportOptions(fileName,'Sheet',sheetName);
            opts.DataRange = 3;  % 数据从第三行开始
            opts.VariableNamesRange = '2:2';  % 表头在第二行
            opts.Sheet=sheetName;%  读取第二个表格
            data= readtable(fileName, opts);
            % 读取数据

            % 读取第二行作为表头
            header = readcell(fileName,'Sheet',sheetName, 'Range', '2:2');
            data.Properties.VariableNames = strtrim(string(header));

            % 显示表头
            disp('所有列的表头名称:');
            disp(data.Properties.VariableNames);

            % 初始化结构体数组
            n = height(data);



            col2_missing = ismissing(data.(2));    % 或者 ismissing(data{:,2})
            col6_missing = ismissing(data.(6));

            % 两列都缺失时才标记为 true
            rows2drop = col2_missing | col6_missing;

            % 删除这些行
            data(rows2drop, :) = [];
```

```
                dutyGroupMap = containers.Map({'勤务一组','勤务二组','勤务三组','综合一组','综合二组'}, ...
                    [1 2 3 4 5]);

                % 分队映射（将分队名称映射为数字）
                teamMap = containers.Map( ...
                    {'一分队', '二分队', '三分队', '四分队', '五分队', '七分队', '八分队', '十分队', ...
                    '十一分队', '十二分队', '十三分队', '十四分队', '十五分队', '十六分队', '十七分队', ...
                    '综合一组', '综合二分队', '综合一分队', '综合 2', '综合二', '综合', '综服 1', '运行组', '综合一'}, ...
                    [1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 18, 19, 19, 18, 18,4,18]);

                % 物品类型映射（A->1, B->2, C->3）
                itemTypeMap = containers.Map(...
                    {'A', 'B', 'C'}, ...
                    [3, 2, 1]);

                % 问题严重程度映射（A->1, B->2, C->3）
                severityMap = containers.Map(...
                    {'A', 'B', 'C'}, ...
                    [3, 2, 1]);
                issueMap    = containers.Map(...
                    {'C1', 'C2', 'C3'}, ...
                    [1, 2, 3]);

                RMap    = containers.Map(...
                    {'1C', '2C', '3C'}, ...
                    [1, 2, 3]);

                RiskMap    = containers.Map(...
                    {'低风险', '中风险', '高风险'}, ...
                    [1, 2, 3]);
                        % 测试方式映射（A->1, B->2, C->3）
```

```
                Job_Position_Map = containers.Map(...
                    {'安全门', 'X 光机','开包','验证','开机','日常','前传','回流','服务
','二判'}, ...
                    [1, 2,3,4,5,6,7,8,9,10]);

                % 遍历每一行数据
                for i = 1:height(data)
                    if isdatetime(data.("日期")(i))
                        date_str = datestr(data.("日期")(i), 'yyyymmdd');
                    else
                        date_str = '000000'; % 默认值
                    end
                    name_str = char(data.("姓名")(i));
                    Job_Position_str = string(data.("岗位")(i));
                    if isKey(Job_Position_Map, Job_Position_str)
                        Job_Position_num = Job_Position_Map(Job_Position_str);
                    else
                        Job_Position_num = 0; % 默认最轻微
                    end
                    team_leader_str = string(data.("责任分队长")(i));


                issue=string(data.("具体问题(分队号-勤务组：起始 hhmm-结束
hhmm 问题描述)")(i));

                    % 问题类型（使用映射）
                    issue_str = string(data.("问题类型")(i));
                    if isKey(issueMap, issue_str)
                        issue_num = issueMap(issue_str);
                    else
                        issue_num = 0; % 默认最轻微
                    end

                    % 问题严重程度（使用映射）
                    severity_str = string(data.("问题严重程度")(i));
```

```
if isKey(severityMap, severity_str)
    severity_num = severityMap(severity_str);
else
    severity_num = 0; % 默认最轻微
end

frequency_num=string(data.("发生频次")(i));
risk_lvl=string(data.("风险等级")(i));


% R 值（使用映射）
R_str = string(data.("R 值")(i));
if isKey(RMap, R_str)
    R_num = RMap(R_str);
else
    R_num = 0; % 默认最轻微
end
% 风险等级（使用映射）
Risk_str = string(data.("风险等级")(i));
if isKey(RiskMap, Risk_str)
    Risk_num = RiskMap(Risk_str);
else
    Risk_num = 0; % 默认最轻微
end

% 构建结构体
events_Position(i).Identifier = 2;
events_Position(i).date = str2double(date_str);
events_Position(i).ID = name_str;
events_Position(i).post = Job_Position_num;
events_Position(i).TeamLeader = team_leader_str;
events_Position(i).issue = issue;
events_Position(i).issue_num = issue_str;
events_Position(i).level = severity_str;
events_Position(i).frequency_num = frequency_num;
```

```
                events_Position(i).risk_lvl = risk_lvl;
                events_Position(i).R_num = R_num;
                events_Position(i).Risk_num = Risk_num;
            end
        case 'X 光机'
            % 设置导入选项
            opts = detectImportOptions(fileName,'Sheet',sheetName);
            opts.DataRange = 3;  % 数据从第三行开始
            opts.VariableNamesRange = '2:2';  % 表头在第二行
            opts.Sheet=sheetName;%  读取第二个表格
            data= readtable(fileName, opts);
            % 读取数据

            % 读取第二行作为表头
            header = readcell(fileName,'Sheet',sheetName, 'Range', '2:2');  % 表头在第二行
            % T.Properties.VariableNames = header;
            data.Properties.VariableNames = strtrim(string(header));

            % 显示表头
            disp('所有列的表头名称:');
            disp(data.Properties.VariableNames);

            % 初始化结构体数组
            n = height(data);



            col2_missing = ismissing(data.(2));    % 或者 ismissing(data{:,2})
            col6_missing = ismissing(data.(6));

            % 两列都缺失时才标记为 true
            rows2drop = col2_missing | col6_missing;

            % 删除这些行
            data(rows2drop, :) = [];
```

```
                dutyGroupMap = containers.Map({'勤务一组','勤务二组','勤务三组','综合一组','综合二组'}, ...
                    [1 2 3 4 5]);

                % 分队映射（将分队名称映射为数字）
                teamMap = containers.Map( ...
                    {'一分队', '二分队', '三分队', '四分队', '五分队', '七分队', '八分队', '十分队', ...
                    '十一分队', '十二分队', '十三分队', '十四分队', '十五分队', '十六分队', '十七分队', ...
                    '综合一组', '综合二分队', '综合一分队', '综合 2', '综合二', '综合', '综服 1', '运行组', '综合一'}, ...
                    [1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 18, 19, 19, 18, 18,4,18]);

                % 物品类型映射（A->1, B->2, C->3）
                itemTypeMap = containers.Map(...
                    {'A', 'B', 'C'}, ...
                    [3, 2, 1]);

                % 问题严重程度映射（A->1, B->2, C->3）
                severityMap = containers.Map(...
                    {'A', 'B', 'C'}, ...
                    [3, 2, 1]);

                % 遍历每一行数据
                for i = 1:height(data)
                    % 日期处理：转换为 yymmdd 格式
                    if isdatetime(data.("日期")(i))
                        date_str = datestr(data.("日期")(i), 'yyyymmdd');
                    else
                        date_str = '000000'; % 默认值
                    end
```

```
                % 姓名处理
                name_str = char(data.("姓名")(i));

                % 勤务组处理
                duty_group_str = string(data.("勤务组")(i));
                if dutyGroupMap.isKey(duty_group_str)
                    duty_group_num = dutyGroupMap(duty_group_str);
                else
                    duty_group_num = 0; % 未知组
                end


                % 分队处理
                team_str = string(data.("所属分队")(i));
                if teamMap.isKey(team_str)
                    team_num = teamMap(team_str);
                else
                    team_num = 0; % 未知分队
                end

                % 责任分队长
                team_leader_str = string(data.("责任分队长")(i));

                % 通道处理
                channel_str = string(data.("通道")(i));

                % 时间段处理：提取开始时间并转换为数字
                time_slot_str = data.("时间段")(i);
                if ~(isempty(time_slot_str) || (iscell(time_slot_str) && isempty(time_slot_str{1})) || strcmp(time_slot_str, ''))
                    start_time = str2double(extractBefore(time_slot_str, '-'));
                else
                    start_time = 0;
                end
```

```
                    if    ~(isempty(time_slot_str) || (iscell(time_slot_str) && is-
empty(time_slot_str{1})) || strcmp(time_slot_str, ''))
                    % 提取开始时间的小时部分
                    time_parts = split(time_slot_str, '-');
                    start_time_str = time_parts{1};
                    colon_pos = 3;
                    hour_str = start_time_str(1:colon_pos-1);
                        hour_num = str2double(hour_str);
                        % 将时间映射为数字 (04:00=1, 05:00=2, ..., 23:00=19)
                        if hour_num >= 4 && hour_num <= 23
                            start_time = hour_num - 3; % 04:00 -> 1, 05:00 -> 2, ...,
23:00 -> 19
                        else
                            start_time = 0; % 超出范围的时间段
                        end
                    else
                    start_time = 0;
                    end


                    % 物品类型处理：从描述中提取类型
                    item_desc = data.("物品类型")(i);
                    item_type_num = 0; % 默认未知
                    if contains(item_desc, 'C:')
                        item_type_num = 3;
                    elseif contains(item_desc, 'B:')
                        item_type_num = 2;
                    elseif contains(item_desc, 'A:')
                        item_type_num = 1;
                    end

                    % 问题严重程度（使用映射）
                    severity_str = string(data.("问题严重程度")(i));
                    if isKey(severityMap, severity_str)
                        severity_num = severityMap(severity_str);
```

```
                else
                    severity_num = 0; % 默认最轻微
                end

                frequency_num=string(data.("发生频次")(i));
                risk_lvl=string(data.("风险等级")(i));


                % 构建结构体
                events_XRay(i).Identifier = 1;
                events_XRay(i).date = str2double(date_str);
                events_XRay(i).ID = name_str;
                events_XRay(i).DutyGroup = duty_group_num;
                events_XRay(i).Team = team_num;
                events_XRay(i).TeamLeader = team_leader_str;
                events_XRay(i).Channel = channel_str;
                events_XRay(i).TimeSlot = start_time;
                events_XRay(i).ItemType = item_type_num;
                events_XRay(i).level = severity_str;
                events_XRay(i).frequency_num = frequency_num;
                events_XRay(i).risk_lvl = risk_lvl;
            end
        case '内部测试'
                % 设置导入选项
            opts = detectImportOptions(fileName,'Sheet',sheetName);
            opts.DataRange = 3;  % 数据从第三行开始
            opts.VariableNamesRange = '2:2';  % 表头在第二行
            opts.Sheet=sheetName;%  读取第二个表格
            data= readtable(fileName, opts);
            % 读取数据

            header = readcell(fileName,'Sheet',sheetName, 'Range', '2:2');
            data.Properties.VariableNames = strtrim(string(header));

            % 显示表头
```

```
disp('所有列的表头名称:');
disp(data.Properties.VariableNames);

% 初始化结构体数组
n = height(data);



col2_missing = ismissing(data.(2));     % 或者 ismissing(data{:,2})
col6_missing = ismissing(data.(6));

% 两列都缺失时才标记为 true
rows2drop = col2_missing | col6_missing;



data(rows2drop, :) = [];



% 映射定义
% 勤务组映射（将勤务组名称映射为数字）
dutyGroupMap = containers.Map({'勤务一组','勤务二组','勤务三组','综合一组','综合二组'}, ...
    [1 2 3 4 5]);

% 分队映射（将分队名称映射为数字）
teamMap = containers.Map( ...
    {'一分队', '二分队', '三分队', '四分队', '五分队', '七分队', '八分队', '十分队', ...
    '十一分队', '十二分队', '十三分队', '十四分队', '十五分队', '十六分队', '十七分队', ...
    '综合一组', '综合二分队', '综合一分队', '综合 2', '综合二', '综合', '综服 1', '运行组', '综合一'}, ...
    [1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 18, 19, 19, 18, 18,4,18]);

% 物品类型映射（A->1, B->2, C->3）
```

```
        itemTypeMap = containers.Map(...
            {'A', 'B', 'C'}, ...
            [3, 2, 1]);

        % 问题严重程度映射（A->1, B->2, C->3）
        severityMap = containers.Map(...
            {'A', 'B', 'C'}, ...
            [3, 2, 1]);

        % 测试方式映射（A->1, B->2, C->3）
        TestMap = containers.Map(...
            {'明测', '暗测'}, ...
            [1, 0]);
    % 测试方式映射（A->1, B->2, C->3）
        Test_Locate_Map = containers.Map(...
            {'安全门', 'X 光机','开包','验证','开机'}, ...
            [1, 2,3,4,5]);
        % 测试方式映射（A->1, B->2, C->3）
        Test_Result_Map = containers.Map(...
            {'通过', '未通过'}, ...
            [1, 0]);
        % 遍历每一行数据
        for i = 1:height(data)
            % 日期处理：转换为 yymmdd 格式
            if isdatetime(data.("日期")(i))
                date_str = datestr(data.("日期")(i), 'yyyymmdd');
            else
                date_str = '000000'; % 默认值
            end

            % 测试方式处理
            test_mode_str= string(data.("测试方式")(i));
            if TestMap.isKey(test_mode_str)
                test_num = TestMap(test_mode_str);
            else
```

```
        test_num = 0; %  未知分队
    end

    %  姓名处理
    name_str = char(data.("姓名")(i));

    %  勤务组处理
    duty_group_str = string(data.("勤务组")(i));
    if dutyGroupMap.isKey(duty_group_str)
        duty_group_num = dutyGroupMap(duty_group_str);
    else
        duty_group_num = 0; %  未知组
    end



    %  分队处理
    team_str = string(data.("所属分队")(i));
    if teamMap.isKey(team_str)
        team_num = teamMap(team_str);
    else
        team_num = 0; %  未知分队
    end

    %  责任分队长
    team_leader_str = string(data.("责任分队长")(i));

    %  测试岗位
    Test_Locate_Str = string(data.("测试岗位")(i));
    if Test_Locate_Map.isKey(Test_Locate_Str)
        Test_Locate_num = Test_Locate_Map(Test_Locate_Str);
    else
        Test_Locate_num = 0; %  未知分队
    end

    %  测试结果
```

```
                    Test_Result_str = string(data.("测试结果")(i));
                    if Test_Result_Map.isKey(Test_Result_str)
                        Test_Result_num = Test_Result_Map(Test_Result_str);
                    else
                        Test_Result_num = 0; % 未知分队
                    end
                    events_Test(i).Identifier = 3;
                    events_Test(i).date = str2double(date_str);
                    events_Test(i).Testmode = test_num;
                    events_Test(i).ID = name_str;
                    events_Test(i).DutyGroup = duty_group_num;
                    events_Test(i).Team = team_num;
                    events_Test(i).TeamLeader = team_leader_str;
                    events_Test(i).Test_Locate = Test_Locate_num;
                    events_Test(i).Test_Result = Test_Result_num;
                end
            end
        end
```

A-2 假设检验代码

```
dataTable = table(time_slots, counts_data(:,1), counts_data(:,2), counts_data(:,3), ...
    'VariableNames', {'时间段', 'X 光机出现次数', '质控次数', '总出现次数'});

disp('成功创建 Table:');
disp(dataTable);

fprintf('\n--- 2. 检验 Table 的有效性 ---\n');

disp('结论: 我们创建的 dataTable 的结构与 X2kafang 函数的输入要求【完全匹配】。');
fprintf('现在将调用您的函数进行分析...\n\n');
```

```
results = X2kafang(dataTable);

disp(results.conclusion);

function results = X2kafang(resultsTable)

disp('---  开始关联性分析  ---');

dataTable = resultsTable;
rowLabels = dataTable{:, 1};
if iscell(rowLabels)
    totalRowIndex = find(strcmp(rowLabels, 'Total'));
else
    totalRowIndex = find(rowLabels == 'Total');
end
if ~isempty(totalRowIndex)
    dataTable(totalRowIndex, :) = [];
    disp("已自动移除  'Total'  行。");
end
contingencyTable = dataTable{:, 2:end};
if isempty(contingencyTable) || any(contingencyTable < 0, 'all')
    error('清理后的数据为空或包含负数，无法进行卡方检验。');
end
disp('用于分析的列联表  (Observed Frequencies):');
disp(contingencyTable);


totalObservations = sum(contingencyTable, 'all');
if totalObservations == 0
    warning('列联表总和为 0，无法进行检验。');
    results.contingencyTable = contingencyTable;
    results.chi2_stat = NaN;
    results.p_value = NaN;
```

```
        results.df = NaN;
        results.conclusion = '数据总和为 0，无法分析。';
        return;
    end
    rowTotals = sum(contingencyTable, 2);
    colTotals = sum(contingencyTable, 1);
    expectedTable = (rowTotals * colTotals) / totalObservations;
    disp('计算出的期望频数表 (Expected Frequencies):');
    disp(expectedTable);
    expectedTable(expectedTable == 0) = 1e-9;
    chi2_stat = sum(((contingencyTable - expectedTable).^2) ./ expectedTable, 'all');
    df = (size(contingencyTable, 1) - 1) * (size(contingencyTable, 2) - 1);
    p_value = 1 - chi2cdf(chi2_stat, df);
    cleanRowLabels = dataTable{:, 1};
    cleanColLabels = dataTable.Properties.VariableNames(2:end);



    alpha = 0.05;
    fprintf('\n--- 卡方检验结果 ---\n');
    fprintf('卡方统计量 (Chi-Square): %.4f\n', chi2_stat);
    fprintf('自由度 (df): %d\n', df);
    fprintf('P 值 (P-value): %e\n', p_value);
    fprintf('显著性水平 (Alpha): %.2f\n', alpha);
    if p_value < alpha
        conclusion_text = sprintf('结论: P 值 < %.2f，拒绝零假设。\n 违规类型与时间段之间存在显著的统计学关联。', alpha);
    else
        conclusion_text = sprintf('结论: P 值 >= %.2f，无法拒绝零假设。没有足够证据表明违规类型与时间段之间存在关联。', alpha);
    end
    disp(conclusion_text);



    results.contingencyTable = contingencyTable;
    results.chi2_stat = chi2_stat;
```

```
    results.p_value = p_value;
    results.df = df;
    results.conclusion = conclusion_text;



        disp('正在生成可视化图表...');



        fig1 = figure('Name', '观测频数分布', 'NumberTitle', 'off', 'Position', [100, 400,
800, 500]);
        bar(contingencyTable, 'stacked');
        title('观测频数分布 (Stacked Bar Chart)');
        xlabel('时间段');
        ylabel('频数');
        set(gca, 'XTick', 1:length(cleanRowLabels), 'XTickLabel', cleanRowLabels);
        xtickangle(45);
        legend(cleanColLabels, 'Location', 'northeast');
        grid on;

        fig2 = figure('Name', '', 'NumberTitle', 'off', 'Position', [950, 400, 600, 500]);
        chi2_contributions = ((contingencyTable - expectedTable).^2) ./ expectedTable;



        h = heatmap(cleanColLabels, cleanRowLabels, chi2_contributions);



        h.Title = '';
        h.XLabel = '违规类型';
        h.YLabel = '时间段';
        h.ColorbarVisible = 'on';
        h.Colormap = parula;
        h.CellLabelFormat = '%.2f';

        sgtitle(fig2, conclusion_text, 'FontSize', 12, 'FontWeight', 'bold');
```

```
        results.figureHandles = [fig1, fig2];

    disp('---  分析结束  ---');
    end
```

A-3 线性回归拟合代码

```
    clc;
    clear;
    Tl = readtable('\1-4 月绩效奖金分配明细.xlsx', 'Sheet', 1);

    pattern = '\d+';

    level_text = cellfun(@(s) regexp(s, pattern, 'match', 'once'), Tl{:,6}, 'UniformOutput',
false);

    level_nums = str2double(level_text);

    dang_text = cellfun(@(s) regexp(s, pattern, 'match', 'once'), Tl{:,7}, 'UniformOutput',
false);
    dang_nums = str2double(dang_text);

    extracted_data_matrix = [level_nums, dang_nums];
    extracted_data_matrix = extracted_data_matrix(5:end,:);
    disp('提取出的纯数字矩阵:');
    disp(extracted_data_matrix);

    num = Tl{:,9};
    value = num(5:end,:);
    x = extracted_data_matrix(:,1);
    y = extracted_data_matrix(:,2);
    data_table = table(x,y,value);



    mdl = fitlm(data_table, 'value ~ x + y');
```

```
disp('--- 使用 fitlm 的完整回归分析结果 ---');
disp(mdl);

coefficients = mdl.Coefficients.Estimate; % 提取所有系数值

c = coefficients(1);
a = coefficients(2);
b = coefficients(3);

fprintf('\n--- 拟合结果 ---\n');
fprintf('方程: num = a*x + b*y + c\n');
fprintf('系数 a (x 的系数): %.4f\n', a);
fprintf('系数 b (y 的系数): %.4f\n', b);
fprintf('常数项 c (截距): %.4f\n', c);
fprintf('最终拟合方程为: num = (%.4f) * x + (%.4f) * y + (%.4f)\n', a, b, c);

r_squared = mdl.Rsquared.Ordinary;
fprintf('模型的 R-squared (决定系数): %.4f\n', r_squared);
disp('R-squared 值越接近 1，说明模型对数据的拟合效果越好。');

figure('Name', '多元线性回归可视化');
scatter3(data_table.x, data_table.y, data_table.value, 'filled', 'DisplayName', '原始数据');
hold on;

x_fit = linspace(min(data_table.x), max(data_table.x), 20);
y_fit = linspace(min(data_table.y), max(data_table.y), 20);
[X_fit, Y_fit] = meshgrid(x_fit, y_fit);
Z_fit = a * X_fit + b * Y_fit + c;

surf(X_fit, Y_fit, Z_fit, 'FaceAlpha', 0.5, 'DisplayName', '拟合的回归平面');

xlabel('x');
ylabel('y');
```

```
zlabel('num');
title('线性回归拟合: num = ax + by + c');
legend;
grid on;
view(30, 25);
```