

中文引用格式:李显,焦宇,史旦达,等. 基于二元 Logistic 回归船舶修造事故伤害模型[J]. 中国安全科学学报,2024,34(12):120-128.

英文引用格式:LI Xian,JIAO Yu,SHI Danda,et al. Accident injuries model of ship repair and building enterprises based on binary Logistic regression [J]. China Safety Science Journal,2024,34(12):120-128.

## 基于二元 Logistic 回归船舶修造事故伤害模型\*

李显<sup>1</sup>,焦宇<sup>\*\*1</sup>副教授,史旦达<sup>1</sup>教授,吴建军<sup>2</sup>高级工程师,康与涛<sup>1</sup>

(1 上海海事大学 海洋科学与工程学院,上海 201306;

2 上海船舶工艺研究所 研发中心,上海 200032)

中图分类号:X928.4

文献标志码:A

DOI: 10.16265/j.cnki.issn1003-3033.2024.12.0350

基金项目:国家自然科学基金资助(51109127)。

**【摘要】** 为探究船舶修造企业(SRBE)事故伤害的影响因素,基于1411起某造船集团事故数据,构建二元 Logistic 回归(LR)模型,并借助让步比(OR)量化企业、时间、地点、人员、环境和事故类型等因素对事故伤害的影响。结果表明:共有11个因素显著影响事故伤害。非合同工事故伤害风险显著高于合同工;非审批的危险作业项目事故伤害风险是审批项目的3.246倍;男性肇事者事故伤害风险显著高于女性;事故中作业人员受教育水平集中在初中,相较于初中学历,作业人员受教育程度越高,事故伤害风险越低;事故伤害呈现季节性特征,第2季度事故频发,第1季度事故较少且风险较低;高峰时段及工作日事故伤害风险分别显著高于非高峰时段和非工作日;物体打击是主要事故类型,而机械伤害显著增加作业人员受伤风险;厂区内船舶和车间是事故伤害高发地点;随着工龄增加,事故造成人员伤害风险降低,然而温度上升,发生事故伤害风险增加。

**【关键词】** Logistic 回归(LR); 船舶修造企业(SRBE); 事故伤害; 影响因素; 让步比(OR)

### Accident injuries model of ship repair and building enterprises based on binary Logistic regression

LI Xian<sup>1</sup>,JIAO Yu<sup>1</sup>,SHI Danda<sup>1</sup>,WU Jianjun<sup>2</sup>,KANG Yutao<sup>1</sup>

(1 College of Ocean Science and Engineering, Shanghai Maritime University, Shanghai 201306, China;

2 Research and Development Center, Shanghai Shipbuilding Technology Research Institute,  
Shanghai 200032, China)

**Abstract:** In order to explore the factors influencing accident injuries in SRBE, a binary LR model was constructed based on data from 1411 accidents from a Chinese shipbuilding group. OR were used to quantify the impact of factors such as enterprise, time, location, personnel, environment, and accident types to accident injuries. The result identifies 11 significant factors influencing accident injuries. Non-contract workers face significantly higher risks of injury compared to contract workers. Non-approved hazardous operations pose 3.246 times the risk of accident injuries compared to approved hazardous operations. Male perpetrators present a significantly higher risk of causing accident injuries than female

\* 文章编号:1003-3033(2024)12-0120-09; 收稿日期:2024-07-14; 修稿日期:2024-09-18

\*\* 通信作者:焦宇(1981—),男,河南汝州人,博士,副教授,主要从事船舶与海工安全与应急方面的研究。E-mail: yujiao@shmtu.edu.cn。

perpetrators. The education level of the workers is predominantly at junior high school, and higher education levels are associated with a lower risk of causing accident injuries compared to those with junior high school education. Accident injuries exhibit seasonal characteristics, with accidents occurring frequently in the second quarter, while relatively fewer and less risky in the first quarter. Peak working hours and workdays significantly increase injury risks compared to off-peak hours and non-working days. Object strikes are the primary type of accidents, and mechanical injuries significantly increase the risk of accident injuries. Ships and workshops are the most likely locations for accident injuries within the shipyards. Longer years of service reduce the risk of accident injuries, while higher temperatures increase it.

**Keywords:** Logistic regression (LR); ship repair and building enterprises (SRBE); accident injuries; influencing factor; odds ratio (OR)

## 0 引言

船舶修造具有高发事故风险的行业特性,易造成人员伤害,给作业人员安全带来威胁。深入研究船舶修造企业(Ship Repair and Building Enterprises, SRBE)事故,识别并量化其伤害影响的因素,对于预防事故具有重要意义。

以往针对 SRBE 事故伤害研究中,多集中于事故特征和事故致因研究。王颖丽<sup>[1]</sup>、丁永明<sup>[2]</sup>等采用频率统计方法,分析造船厂发生的工伤事故,涵盖事故发生次数、事故严重性以及受害员工年龄分布等多个方面;YONG<sup>[3]</sup>、YOU<sup>[4]</sup>等进一步探讨了事故中作业人员的技术熟练程度在事故伤害和事故数量方面的差异;LEE<sup>[5]</sup>总结了韩国造船厂行业事故特征;BARLAS<sup>[6]</sup>、IZCI<sup>[7]</sup>等分别研究了土耳其造船厂致死事故和非致死事故的特征,并探究事故致因。Logistic 回归(Logistic Regression, LR)模型用于分析一个或多个自变量和一个分类因变量之间的关系,确定自变量对事件发生概率的影响大小。在事故研究领域,其主要被应用于交通(海<sup>[8]</sup>、陆<sup>[9]</sup>、空<sup>[10]</sup>)方面,如预测事故发生的概率、分析事故发生的风险因素等。然而,少有研究将 LR 模型用于 SRBE 事故中。

鉴于此,笔者拟基于 2013—2022 年我国某造船集团(THX)生产事故调查报告和网络爬虫获取的气象环境数据为研究对象,在描述性统计和编码影响因子分析的基础上,将因变量分类类别划分为有/无人员伤害,借助二元 LR 模型建立 SRBE 事故伤害的识别模型,并探讨显著影响因素中连续变量分箱对模型拟合和预测性能的影响,以期 SRBE 事故伤害预防提供模型理论支撑,并为相关领域的从业者及监管部门提供决策参考。

## 1 船舶修造事故伤害数据

### 1.1 事故伤害数据来源

SRBE 事故数据主要来源于 THX 事故数据库。数据库中包含旗下各船舶修造子公司发生的所有事故调查报告。数据采集途径有 3 种:①2013—2022 年,THX 数据库的 1 411 起事故数据;②通过网络爬虫技术,基于企业的地理位置与国家气象网站,获取事故发生时气象环境信息,包含天气、风向、风级和温度;③基于事故发生日期,提取事故发生的年、月、日和星期数据,进一步结合我国法定节假日调休信息,衍生出工作日和非工作日数据。

### 1.2 事故伤害分类及伤害因素处理

#### 1.2.1 事故伤害分类

企业在进行事故调查时,需按照《企业职工伤亡事故分类》<sup>[11]</sup>(简称分类标准)对事故后果进行分类。分类标准<sup>[11]</sup>中将事故伤害划分为轻伤、重伤和死亡 3 类,然而在 SRBE 事故中,有许多的人员伤害程度达不到轻伤的标准,但又区别于无伤害,此类伤害被划分为轻微伤类别。在提取的数据中共有 5 类事故伤害类型,分别为无人员伤害、轻微伤、轻伤、重伤和死亡。考虑到文中研究的目的是探究 SRBE 事故伤害的影响因素,将因变量事故造成人员伤害分为无人员伤害和有人员伤害 2 类。

#### 1.2.2 事故伤害因素处理

SRBE 事故伤害的发生是由人、机、环、管 4 因素共同作用的结果。基于数据库中记录内容和借鉴其他事故领域的研究,从环境、时间、空间、人员、事故和企业等信息中选取 23 个变量。

环境变量中风力、风向和天气数据在一天中呈现出变化的趋势,数据处理时保留最后部分,如阴转

晴会被划分为晴。风力划分为6类,分别为1—5级和6级往上;风向按照8个方向进行划分;天气可分为晴、阴和雨雪雾霾3类;根据一天中最高和最低温度以及事故发生时间,采用正弦函数模拟温度的周期性变化,如下式:

$$\theta(t) = \frac{\theta_{\max} + \theta_{\min}}{2} + \frac{\theta_{\max} - \theta_{\min}}{2} \times \sin\left(\frac{2\pi}{24}(t - t_0)\right) \quad (1)$$

式中: $\theta(t)$ 为时间 $t$ 时刻的温度,℃; $\theta_{\max}$ 为温度的最大值,℃; $\theta_{\min}$ 为温度的最小值,℃; $t_0$ 为时间偏移量,其影响正弦函数的相位。

时间变量中,是否为工作日需考虑国家法定节假日和调休<sup>[12]</sup>,将其划分为是(正常工作日和调休)和否(法定节假日和正常周末);依据2 h划分时刻变量,统计发现SRBE事故时刻特征也符合“M”分布<sup>[12]</sup>,将其按照是否为事故高峰时段分为2类:是(8:00—12:00和14:00—18:00)和否(其他时间段);星期变量划分为星期一至星期日7个类别;季节变量划分为4个季度。

空间变量中,根据事故发生在SRBE的具体位置,将事故地点变量划分为厂区内 and 厂外2大类,其中,厂区内进一步进行细分为内场、船舶、码头和坞、车间4类<sup>[13]</sup>。

人员变量中,性别、婚姻状态和新老工人均为二分类,其中,新老工人划定基于事故人员入职的登记表和在企业工作时长。人员的受教育水平,划分为5类,分别是小学、初中、高中、专科和大学及以上,其中,高中类别中包含高中、职专和中技。

事故变量中,按照分类标准<sup>[11]</sup>统计,除其他伤害外共有10种事故类型。由于部分类别出现的频次较少,可能会引起模型估计不稳定,导致模型估计的回归系数具有较大的标准误,从而引入较大的偏差和方差,影响模型的解释性。合并火灾、灼烫、车辆伤害、交通事故、中毒和窒息、触电等6类事故类型,标记为合并类别。事故性质确定与企业事故安全管理理念或事故调查执行标准挂钩,分为责任和非责任事故2类。界定标准为:如果事故调查发现当事人未能遵守相关安全规定或存在明显的疏忽,则归为责任事故;反之,如果环境条件(作业环境和自然环境)不佳、当事人已经尽力避免事故或意外发生,则归为非责任事故。

企业变量中,由于THX下属的造修船厂均为中大型企业,人员规模可划分为 $[0, 1\ 000)$ 和 $[1\ 000, +\infty)$ 2类;危险作业审批项划分为是和否;SRBE会

根据项目的具体需求和成本效益,选择不同的用工形式人员来进行生产作业。在船舶修造中,大部分工程项目或部分工程任务往往分包给专业公司,船厂中的工人大多来自不同的分包公司,而合同制员工更多地负责维持企业的基础运转。以韩国的造船业为例,大规模的装配操作中,约有80%的员工是分包工人<sup>[14]</sup>。THX事故涉及人员的用工形式被划分为合同制与非合同制2类,其中,非合同制包含外包、协力、劳务及外协制等4种不同的用工形式。

### 1.3 变量处理编码与统计分析

按照变量类型可将变量划分为分类变量和连续变量。分类变量需要经过数值编码转化后才可纳入到模型中,根据类别多少的不同,采用不同的编码方式。其中,二分类变量使用二进制编码,将频次较高的类别编码为1,较少的类别编码为0;多分类变量采用虚拟变量(哑变量)编码,即一个变量有 $m$ 种分类,则将其转化成为 $m - 1$ 种二分类的虚拟变量,选择频次最多的类别为对照组。具体操作过程为:①将多分类变量根据事故发生的频次进行排序,从高到低;②对每个分类进行编码,编码从1到 $m$ ;③在创建虚拟变量时,删除编码为1的分类,即排名最高的分类不单独设立虚拟变量。连续变量描述统计见表1,分类变量描述统计与编码结果见表2。

表1 连续变量描述性统计

Table 1 Continuous variable descriptive statistics

变量维度	变量名称	平均值	标准差	最小值	中位数	最大值
人员	年龄/岁	38.04	9.16	19	38	60
	工龄/a	10.22	8.48	1	8	40
	本工种工龄/a	6.83	5.74	1	5	38
环境	最低温度/℃	13.66	9.46	-17	15	36
	最高温度/℃	20.90	9.61	-13	23	40
	正弦温度/℃	18.64	9.66	-14.5	20.5	39.8

## 2 SRBE事故伤害模型

### 2.1 SRBE事故伤害二元LR模型

在事故领域的研究中,通常需要分析事故或伤害严重程度的发生概率 $P(0 < P < 1)$ ,并识别对其有显著影响的因素。然而,由于事故或伤害严重程度与众多因素的关系难以用常规线性模型准确描述,直接分析概率 $P$ 面临一定挑战。此外,当 $P$ 接近边界值时,普通方法难以有效处理微小变化。LR模型是一种研究分类因变量的模型,可解释因变量

表 2 分类变量描述性统计与编码结果

Table 2 Descriptive statistics and encoding results for categorical variables

变量维度	变量名称	编码=类别(频数,占比/%)	变量维度	变量名称	编码=类别(频数,占比/%)				
事故伤害	人员伤害	1=人员伤害(1 078, 76.40) 0=无人员伤害(333, 23.60)	企业	时间	星期	1=星期三(233, 16.51) 2=星期五(217, 15.38) 3=星期六(217, 15.38) 4=星期四(205, 14.53) 5=星期二(200, 14.17) 6=星期一(182, 12.90) 7=星期日(157, 11.13)			
	用工形式	1=非合同制(1 206, 85.47) 2=合同制(205, 14.53)				季节	1=第二季度(412, 29.20) 2=第三季度(370, 26.22) 3=第一季度(327, 23.18) 4=第四季度(302, 21.40)		
企业	人员规模	1=[1 000, +∞)(862, 61.09) 2=[0, 1 000)(549, 38.91)					是否工作日	1=是(1 007, 71.37) 0=否(404, 28.63)	
	危险作业审批项	1=否(1 130, 80.09) 0=是(281, 19.91)					高峰事故时刻	1=[8, 12) & [14, 18)(1 024, 72.57) 0=其他(387, 27.43)	
人员	性别	1=男(1 331, 94.33) 0=女(80, 5.67)				环境	天气	1=晴(876, 62.08) 2=雨雪雾霾(283, 20.06) 3=阴(252, 17.86)	
	婚姻状态	1=已婚(1 219, 86.39) 0=未婚(192, 13.61)						风向	1=东南风(377, 26.72) 2=西北风(197, 13.96) 3=南风(192, 13.61) 4=北风(173, 12.26) 5=东北风(172, 12.19) 6=东风(152, 10.77) 7=西南风(90, 6.38) 8=西风(58, 4.11)
	受教育水平	1=初中(966, 68.46) 2=高中(287, 20.34) 3=专科(69, 4.89) 4=小学(53, 3.76) 5=本科及以上(36, 2.55)							风力
	新工人	1=否(1 148, 81.36) 0=是(263, 18.64)							
事故	事故性质	1=责任事故(1 029, 72.93) 0=非责任事故(382, 27.07)	空间	事故地点	1=内场(468, 33.17) 2=船舶(406, 28.77) 3=码头或坞(390, 27.64) 4=车间(132, 9.36) 5=厂区外(15, 1.06)				
	事故类型	1=物体打击(462, 32.74) 2=其他伤害(374, 26.51) 3=机械伤害(162, 11.48) 4=高处坠落(144, 10.21) 5=起重伤害(102, 7.23) 6=合并类别(167, 11.83)							

和自变量之间的相互关系,因此,构建 LR 模型。

SRBE 事故依据是否造成人员伤害,其结果为有人员伤害和无人员伤害 2 类,采用二元 LR 模型进行研究,其表达式为:

$$P(Y = 1 | X) = \frac{\exp(\alpha + \beta_1 x_1 + \dots + \beta_i x_i)}{1 + \exp(\alpha + \beta_1 x_1 + \dots + \beta_i x_i)} \quad (1)$$

式中:因变量  $Y$  为二分类变量,0 表示无人员伤害,1 表示有人员伤害; $\alpha$  为常数; $\beta_i$  为使用最大似然估

计方法估计的系数; $x_i$  为自变量。

让步比(Odds Ratio, OR)是一种量化因变量的 2 种可能结果之间关系指标<sup>[13]</sup>,其中,0 指的是事件发生概率与不发生概率之间的比值。在构建 SRBE 事故伤害二元 LR 模型中,OR 描述了事故造成人员伤害概率(Odds<sub>1</sub>,  $O_1$ )相对于无人员伤害概率(Odds<sub>0</sub>,  $O_0$ )的比率,其表达式为:

$$O = \frac{p}{1-p} = \exp(\alpha + \beta_1 x_1 + \dots + \beta_i x_i) \quad (2)$$

$$OR = \frac{O_1}{O_0} \quad (3)$$

## 2.2 模型变量检验

在回归模型中选取自变量需确保其与因变量显著相关且自变量间相互独立,根据自变量的数据类型及其分布特性,选择合适的假设检验方法如下:对于分类自变量,使用卡方检验来评估与因变量的关联;对于服从正态分布的连续自变量,采用  $t$  检验;而对于不符合正态分布的连续自变量,则使用斯皮尔曼等级相关系数。此外,为避免自变量多重共线性问题影响模型的准确性和可解释性,采用方差膨胀因子(Variance Inflation Factor, VIF)进行多重共线性诊断,剔除  $VIF > 5$  的自变量。

## 2.3 显著性因素筛选

以往研究常利用单因素分析、逐步回归或其组合筛选显著性影响因素<sup>[15-16]</sup>。首先,单因素分析通过假设检验筛选与因变量显著相关的自变量,并利用 VIF 剔除多重共线性变量。再将显著变量纳入回归模型。逐步回归作为一种模型显著因素的选择方法,通过向前添加或向后剔除变量来优化模型。

## 2.4 模型检验

在二元 LR 中,霍斯默-莱梅肖(Hosmer-Lemeshow, H-L)检验常用于评估模型的总体拟合效果。二分类模型预测性能判定借助接受者操作特性(Receiver Operating Characteristic, ROC)曲线下的面积(Area Under Curve, AUC)。AUC 值范围为 0.5~1,越接近 1 代表预测效果越好,通常 AUC 值大于 0.7 时结果较准确<sup>[17]</sup>。此外,模型优劣通过负二倍对数似然函数值(-2 Log-Likelihood, -2LL)进行衡量。在比较不同的二元逻辑回归模型时,-2LL 值越小,说明模型对数据的拟合程度越高。

# 3 二元 LR 模型指标和结果分析

## 3.1 模型评价指标分析

在建立回归模型后,先评估其准确性和拟合优度;只有在这些检验结果符合标准的情况下,才能进一步分析模型的具体结果。

对模型变量进行单因素和共线性诊断分析,结果显示:共有 15 个显著相关变量,且 VIF 值均小于 5,说明自变量之间不存在强的多重共线性,可作为回归模型的变量,变量单因素和共线性诊断分析结

果见表 3。

表 3 变量单因素和共线性诊断分析结果

Table 3 Analysis results of univariate variables and collinearity diagnostics

变量名称	卡方值	自由度	显著性	VIF
年龄	—	—	0.140	1.757
工龄	—	—	0.025*	2.089
本工种工龄	—	—	0.047*	1.598
正弦温度	—	—	0.030*	1.415
危险作业审批项	63.308	1	0.000*	1.047
用工形式	7.722	1	0.005*	1.015
人员规模	0.685	1	0.408	1.197
事故地点	31.121	4	0.000*	1.045
事故性质	1.019	1	0.313	1.168
事故类型	59.263	5	0.000*	1.018
性别	9.087	1	0.003*	1.057
婚姻状态	0.944	1	0.331	1.401
新老工人	2.132	1	0.144	1.071
教育水平	16.520	4	0.002*	1.042
季节	13.882	3	0.003*	1.330
工作日与休息日	6.694	1	0.010*	1.073
星期	14.870	6	0.021*	1.069
高峰事故时刻	4.788	1	0.029*	1.014
风级	15.231	5	0.009*	1.030
风向	23.545	7	0.001*	1.086
天气	3.542	2	0.170	1.042

注:\*代表显著性  $p$  值小于 0.05。

将显著相关变量纳入到二元 LR 模型中,建立向前 LR 及向后 LR 模型。模型 H-L 检验  $p$  值、-2LL 值和 AUC 值,见表 4。

表 4 两模型检验比较结果

Table 4 Results of comparison between two model test

模型	H-L 检验 $p$ 值	-2LL	AUC
向前 LR	0.123	1 114.263	0.847
向后 LR	0.407	1 102.038	0.852

由表 4 可知:两模型 H-L 检验  $p$  值均大于 0.05,模型拟合效果均达标,但向后 LR 模型的 -2LL 值小于向前 LR 模型,说明向后 LR 模型拟合效果更好;同时,向后 LR 模型的 AUC 值也大于后者,说明向后 LR 模型预测的准确度更高。

## 3.2 模型结果分析

为保证模型的简洁性,采用向后逐步回归的方法筛选变量,最终模型只保留对事故人员伤害有显著影响( $p < 0.05$ )的变量,见表 5。

根据模型分析结果,得到 11 个对 SRBE 事故伤害有显著影响的因素,其中,企业维度有 2 个,人员和时间维度各有 3 个,事故、空间和环境维度各有 1 个。

与企业相关的 2 个显著性影响因素分别为用工

表 5 SRBE 事故伤害二元 LR 分析结果

Table 5 Binary logistic regression analysis results for SRBE accident casualty

变量维度	变量名称	变量类别	回归系数	标准误差	df	Sig.	OR	OR 的 95%		
企业	用工形式		0.355	0.181	1	0.047	1.427	1.000	2.036	
	危险作业审批项		1.177	0.160	1	0.000	3.246	2.373	4.441	
人员	性别		0.956	0.264	1	0.000	2.602	1.550	4.368	
	工龄		-0.030	0.013	1	0.026	0.970	0.945	0.996	
	受教育程度	初中 <sup>a</sup>				4	0.001			
		高中		-0.478	0.170	1	0.005	0.620	0.444	0.866
		专科		-0.907	0.289	1	0.002	0.404	0.229	0.710
大学		-0.965	0.432	1	0.025	0.381	0.163	0.888		
时间	季节	第二季度 <sup>a</sup>			3	0.001				
		第一季度	-0.499	0.184	1	0.007	0.607	0.424	0.871	
	高峰事故时刻		0.376	0.166	1	0.023	1.457	1.053	2.017	
	是否工作日		0.356	0.145	1	0.015	1.462	1.073	1.898	
事故	事故类型	物体打击 <sup>a</sup>			5	0.000				
		其他伤害	-0.825	0.176	1	0.000	0.438	0.310	0.619	
		合并类别	-0.477	0.227	1	0.035	0.621	0.398	0.968	
		机械伤害	0.616	0.303	1	0.042	1.851	1.022	3.352	
空间	事故地点	内场 <sup>a</sup>			4	0.000				
		船舶	0.383	0.185	1	0.039	1.467	1.020	2.108	
		码头或坞	-0.353	0.169	1	0.036	0.702	0.504	0.978	
		车间	0.852	0.297	1	0.004	2.345	1.310	4.198	
		厂区外	-1.611	0.733	1	0.028	0.200	0.047	0.840	
环境	正弦温度		0.058	0.009	1	0.000	1.060	1.041	1.079	
常数项			-0.293	0.450	1	0.515	0.746			

注:a 作为分类变量中的基准变量。

形式、危险作业审批项。企业用工形式中合同工占全部涉事人员的 14.53%，非合同工相对于合同工的 OR 值为 1.427，说明事故造成非合同工伤害的风险是合同工的 1.427 倍。在船舶修造行业，合同工主要负责企业的基础运转，享有更系统的安全培训和更严格的工作规程遵守，而非合同工频繁地从事临时或较危险的任务，面临更高的事故伤害风险。危险作业审批项和未审批项的事故起数比约为 2:8，危险作业审批变量的 OR 值为 3.246，说明危险作业非审批项目发生事故造成人员伤害的风险要远大于危险作业审批项目。危险作业审批项目通常伴随详细的风险评估和严格的监控措施，能识别出潜在危险并提供有效的防护措施，且经审批的作业需要人员更严格地遵循安全规范和操作标准，从而减少由操作不当导致的事故伤害。

与人员相关的 3 个显著性影响因素分别是工龄、性别和受教育程度。事故中涉及人员工龄范围为 1~40 年，平均工龄为 10.22 年，中位数为 8 年。工龄的 OR 值为 0.970，说明随着工龄的增加，事故造成人员伤害的风险降低。男性在船舶修造作业中

占有主导地位，其肇事数量高达 94.33%，性别的 OR 值为 2.602，说明男性肇事者在事故中造成人员伤害的风险是女性肇事者的 2 倍多，即男性相对于女性易造成事故伤害。事故中人员受教育程度主要集中在初中和高中，两者的合计占比达到 88.80%，而低学历（小学）和高学历（本科及以上）的人员占比均不到 4%。受教育程度以初中为基准，除小学类别，剩余各个类别均达到显著性水平，OR 值小于 1，说明受教育水平是初中的人员最易造成事故人员伤害。进一步对比显著性水平的 OR 值可发现，高中、专科和大学的 OR 值呈现递减的趋势，说明作业人员受教育程度相较于初中水平越高，造成事故伤害的风险越低。

与时间相关的 3 个显著性影响因素分别是：季节、高峰事故时刻和是否工作日。船舶修造行业受到季节因素影响显著，第 2 季度发生船舶修造事故最多，达到 412 起，占比 29.2%。第 2 季度的 4 月份常常对应我国春节后大规模的人员流动潮，SRBE 在此期间通常面临经验丰富的工人离职和新员工上岗的情况。新员工常常缺乏充分的安全培训，对即

将从事的工作不够熟悉,或对工作环境的适应不良,从而导致第2季度事故的频繁发生。第1季度涵盖我国传统节日—春节,在此期间,企业会加强岁末安全检查、应对春节假期的生产停滞以及进行平安开工检查,旨在确保安全工作能够得到有效收尾和顺利开端。从事故数量(最少)和OR值(0.607)的数据分析结果可得到印证。事故高峰时刻的OR值为1.457,说明在[8, 12) & [14, 18)时刻区间中发生事故伤害的风险比其他时段高出约45.7%。在这2个时间段内,员工可能处于不同的生理和心理状态。例如:上午刚开始工作时的调整期和下午疲劳时段,导致集中注意力的能力下降,从而增加事故伤害风险。工作日发生事故起数和非工作日事故起数比约为7:3,工作日发生的事故伤害的风险是非工作日的1.428倍。

与事故、空间和环境相关的显著性影响因素分别是:事故类型、事故地点和正弦温度。物体打击是最常见的事故类型,发生462起,占比32.74%,机械伤害OR值为1.851,表明机械伤害造成的事故伤害风险显著高于物体打击。根据分类标准<sup>[11]</sup>中的事故类型,SRBE事故难以精确归类,导致其他伤害类型的事故占比高达26.51%。与物体打击相比,其他伤害类型的事故伤害风险显著较低,这可能与其它伤害类别包括的事故种类和性质较为广泛有关。此外,合并类别的事故类型的风险也相对较低。尽管合并类别覆盖了多种事故类型,其综合风险水平低于物体打击。绝大多数事故发生在造船厂内部,其中,厂区外发生的事故仅占1.06%。在事故地点的分析中,船舶和车间区域的事故伤害风险显著高于内场,OR值分别为1.467和2.345。这表明:船舶和车间区域的事故伤害发生风险显著增加。相对而言,码头或坞造成的事故伤害风险仅为内场的70%,而厂区外的事故伤害风险仅为内场的20%。事故发生时的温度范围为-14.5~39.8℃,平均温度为18.64℃。温度的OR值为1.060,即随着温度的增加,作业过程中造成事故伤害的风险增加。

#### 4 显著连续变量的分箱讨论

为评估显著影响因素中的连续变量分箱对模型拟合和预测性能的影响,考虑自定义分箱和等距分箱。假定工龄变量不分箱为 $\bar{A}$ ,自定义分箱为 $A_1$ ,等距分箱为 $A_2$ ;温度变量不分箱为 $\bar{B}$ ,自定义分箱为 $B_1$ ,等距分箱为 $B_2$ 。连续变量分箱描述性统计与编码结果见表6。

表6 连续变量分箱描述性统计与编码结果

变量名称	分箱方式	编码=类别(频数,占比/%)
$A_1$	自定义	1=熟练工(1 240, 87.88) 0=新手(171, 12.12)
$A_2$	等距	1=0~10(909, 64.42) 2=10~20(313, 22.18) 3=20~30(157, 11.13) 4=30~40(33, 2.27)
$B_1$	自定义	1=非高温(1 353, 95.89) 0=高温(58, 4.11)
$B_2$	等距	1=20~30(603, 42.74) 2=10~20(379, 26.86) 3=0~10(252, 17.86) 4=30~40(125, 8.86) 5=小于0(52, 3.69)

自定义分箱的方式往往依据普遍接受的认知(经验)或已有研究。在我国气象上,以日最高气温达到或超过35℃作为高温的标准<sup>[18]</sup>。参考YONG<sup>[3]</sup>研究,工龄在1年以内的工人被标记为新手,超过1年的被标记为熟练工。等距分箱需根据连续变量最值进行考虑,分箱的个数不宜过多,且各类别中的频数不应该较少。按照10作为分箱的依据对工龄和温度变量进行分箱,考虑发生时温度在-20~-10℃的事故仅有2起,将此分箱类别与-10~0℃合并为小于0。

采用向后逐步回归,以 $\bar{A}\bar{B}$ 为参考模型,剩余组合模型为对比模型。9个模型的分箱变量的组合方式、模型H-L检验 $p$ 值、-2LL值和AUC值见表7。

表7 分箱变量组合模型对比结果

编号	模型	H-L 检验 $p$ 值	-2LL	AUC
1	$\bar{A}\bar{B}$	0.407	1 102.038	0.852
2	$A_1\bar{B}$	0.273	1 108.014	0.850
3	$A_2\bar{B}$	0.273	1 108.014	0.850
4	$\bar{A}B_1$	0.252	1 116.673	0.847
5	$A_1B_1$	0.082	1 121.395	0.845
6	$A_2B_1$	0.082	1 121.395	0.845
7	$\bar{A}B_2$	0.079	1 107.755	0.850
8	$A_1B_2$	0.559	1 111.497	0.848
9	$A_2B_2$	0.559	1 111.497	0.848

由表7可知:以 $\bar{A}\bar{B}$ 组合的参考模型的-2LL得分最低,且AUC值最大,说明此模型的拟合效果相较于对比模型较好,预测的效果也相对较好。分箱结果中,未进行分箱的变量组合模型相对较优;2种

工龄变量的分箱方式对于模型拟合优度和预测结果相同;在考虑温度分箱时,温度等距分箱结合工龄变量不分箱( $\bar{A}B_2$ )模型效果最优。

## 5 结 论

1) 借助单因素分析和向后逐步回归构建 SRBE 事故伤害二元 LR 模型。模型在显著性水平 0.05 下通过拟合优度检验, AUC 值为 0.852, 表明其具有良好的影响因素识别有效性和预测结果正确性。

2) 识别并量化对 SRBE 事故伤害有显著影响的因素为用工形式、危险作业审批项、作业人员工龄、性

别、受教育程度、事故发生季节、是否在工作日、高峰事故时刻、正弦温度、事故类型和事故地点等 11 个。

3) 对比讨论显著影响因素中连续变量分箱的组合模型结果。结果发现:连续变量不分箱时构建组合模型的拟合优度和预测效果较好。

4) 研究中的数据基于 THX 的事故调查报告,可能限制了研究的全面性和指标设置的通用性。尽管未来随着数据内容的增加,研究结果有望得到进一步的完善和优化,但目前仍然迫切需要建立全国性统一且标准化的船舶修造行业事故数据库,以便更系统地进行事故研究。

## 参 考 文 献

- [1] 王颖丽, 夏昭林, 庄惠民. 某造船企业 1994—2005 年职业伤害流行病学调查[J]. 环境与职业医学, 2008, 25(1): 49-50.  
WANG Yingli, XIA Zhaolin, ZHUANG Huimin. Epidemiological study of occupational injuries in a shipyard during the period of 1994-2005[J]. Journal of Environmental & Occupational Medicine, 2008, 25(1): 49-50.
- [2] 丁永明, 朱新良, 傅华. 造船企业工伤事故调查[J]. 中国工业医学杂志, 2000, 13(5): 284-285.  
DING Yongming, ZHU Jinliang, FU Hua. A survey on occupational injuries in a shipyard[J]. Chinese Journal of Industrial Medicine, 2000, 13(5): 284-285.
- [3] YONG J B. Prevalence of occupational accidents and factors associated with deaths and disabilities in the shipbuilding industry: comparisons of novice and skilled workers[J]. WORK-A Journal of Prevention Assessment & Rehabilitation, 2021, 69(3): 997-1 005.
- [4] YOU W S, JEONG B Y, LEE D K. Comparison of occupational accidents and illnesses in shipbuilding industry[J]. Journal of the Ergonomics Society of Korea, 2019, 38(5): 385-392.
- [5] LEE K T. The characteristics of industrial accidents in shipbuilding industry[J]. Journal of the Ergonomics Society of Korea, 2012, 31(1): 137-142.
- [6] BARLAS B, IZCI F B. Individual and workplace factors related to fatal occupational accidents among shipyard workers in Turkey[J]. Safety Science, 2018, 101: 173-179.
- [7] IZCI F B, GOKYAY O, BARLAS B. Investigation of non-fatal occupational accidents and their causes in Turkish shipyards[J]. International Journal of Occupational Safety and Ergonomics, 2024, 30(1): 33-40.
- [8] 蔡玉良, 于淳, 孙旭, 等. 我国沿海船舶事故特征及影响因素分析[J]. 中国航海, 2021, 44(2): 45-52.  
CAI Yuliang, YU Chun, SUN Xu, et al. Characteristics of ship accidents in coastal waters of China and analysis on influencing factors[J]. Navigation of China, 2021, 44(2): 45-52.
- [9] 王磊, 吕璞, 林永杰. 高速公路交通事故影响因素分析及伤害估计[J]. 中国安全科学学报, 2016, 26(3): 86-90.  
WANG Lei, LYU Pu, LIN Yongjie. Traffic accidents on freeways: influencing factors analysis and injury severity evaluation[J]. China Safety Science Journal, 2016, 26(3): 86-90.
- [10] 于洪霞, 季建华, 李兴. 一种基于 Logistic 回归的全损飞行事故分析与预测[J]. 中国安全科学学报, 2010, 20(3): 34-38.  
YU Hongxia, JI Jianhua, LI Xing. An analysis and forecast model of total loss flight accidents based on logistic regression method[J]. China Safety Science Journal, 2010, 20(3): 34-38.
- [11] GB/T 6441—1986, 企业职工伤亡事故分类标准[S].  
GB/T 6441—1986, Classification for casualty accidents of enterprise staff and workers[S].
- [12] 焦宇, 李显, 刘琦, 等. 2005—2019 年我国重特大安全生产事故特征分析[J]. 安全与环境学报, 2021, 21(6):

2 875-2 882.

JIAO Yu, LI Xian, LIU Qi, et al. Analysis of the characteristics of major accidents and particularly serious accidents from 2005 to 2019 in China[J]. Journal of Safety and Environment, 2021, 21(6): 2 875-2 882.

- [13] 李显,焦宇,陈文涛,等. 船舶修造企业火灾事故特征挖掘与致因分析[J]. 中国安全科学学报, 2024, 34(3): 29-38.

LI Xian, JIAO Yu, CHEN Wentao, et al. Feature mining and causative analysis of fire accidents in ship repair and building enterprises[J]. China Safety Science Journal, 2024, 34(3): 29-38.

- [14] LEE S Y, KIM E, PARK K. Subcontracting workers outside of Korean social safety net: a case study of subcontracting workers in Ulsan shipbuilding industry [J]. Social Welfare Policy, 2017, 44(2): 111-144.

- [15] 王卫杰,沈轩霆,王贵彬,等. 电动自行车骑行者事故伤害程度影响因素分析[J]. 中国安全科学学报, 2019, 29(2): 20-25.

WANG Weijie, SHEN Xuanning, WANG Guibin, et al. Analysis of factors affecting injury to electric bicycle rider in crash[J]. China Safety Science Journal, 2019, 29(2): 20-25.

- [16] 戡晓峰,李德林,杨文臣. 山区二级公路交通事故致因的时间演化机制[J]. 中国安全科学学报, 2019, 29(4): 31-36.

JI Xiaofeng, LI Delin, YANG Wenchen. Temporal evolution analysis on causes of traffic accidents occurring on secondary highways in mountain areas[J]. China Safety Science Journal, 2019, 29(4): 31-36.

- [17] 李海萍,龙志航,杨滋恒,等. 基于 Logistic 模型的四川凉山州森林火灾风险分析[J]. 安全与环境学报, 2021, 21(2): 498-505.

LI Haiping, LONG Zhihang, YANG Ziheng, et al. Analysis of forest fire risk in Sichuan Liangshan based on Logistic model[J]. Journal of Safety and Environment, 2021, 21(2): 498-505.

- [18] QX/T 280—2015, 极端高温监测指标[S].

QX/T 280—2015, Monitoring indices of high temperature extremes[S].

作者简介: 李显 (1995—),男,河南信阳人,博士研究生,主要研究方向为船舶修造事故数据挖掘、作业过程风险评估等。E-mail: xianl828@163.com。