

基于多因素Logistic回归分析的广西低等级公路交通事故风险评估研究

黄静洁, 程 瑞*, 纪 宇, 张燕熙

桂林电子科技大学建筑与交通工程学院, 广西 桂林

收稿日期: 2023年1月31日; 录用日期: 2023年3月7日; 发布日期: 2023年3月15日

摘 要

为了探究低等级公路交通事故致因, 预防重大事故及减少事故伤害后果, 基于2012~2021年间广西低等级公路交通事故数据, 从人-车-道路-环境因素四方面选取48个潜在影响因素, 采用多因素Logistic回归分析筛选得到影响不同水平事故严重度的显著性因素。研究结果表明: 导致事故严重度为死亡或重伤的显著性影响因素有驾驶人驾龄为49~60个月, 参与方车辆类型为半挂车、客车、摩托车, 发生事故时车速为60 km/h以上, 车辆与道路固定物碰撞, 事故发生月份在4月、5月、6月, 事故发生时间段在0:00~3:00; 事故严重度为未受伤或轻微受伤的显著性影响因素有参与方车辆类型为小型轿车、事故时车速为0~40 km/h、车辆状态为静止、道路灯点亮、无纵坡、晴天、事故发生时间段在5:00~6:00。研究结果可为广西低等级公路进行交通事故预防和风险评估提供参考和依据, 对于提高广西低等级公路运输安全水平具有现实意义。

关键词

低等级公路, 事故严重度, 影响因素, 多因素Logistic回归分析

Risk Assessment Studies of Low-Grade Highway Traffic Accidents in Guangxi Based on Multivariate Logistic Regression Analysis

Jingjie Huang, Rui Cheng*, Yu Ji, Yanxi Zhang

School of Architecture and Transportation, Guilin University of Electronic Technology, Guilin Guangxi

Received: Jan. 31st, 2023; accepted: Mar. 7th, 2023; published: Mar. 15th, 2023

*通讯作者。

文章引用: 黄静洁, 程瑞, 纪宇, 张燕熙. 基于多因素 Logistic 回归分析的广西低等级公路交通事故风险评估研究[J]. 交通技术, 2023, 12(2): 93-104. DOI: 10.12677/ojtt.2023.122012

Abstract

In order to explore the causes of low-level highway traffic accidents, prevent major accidents and reduce the consequences of accident injuries, based on the data of low-level highway traffic accidents in Guangxi from 2012 to 2021, 48 potential influencing factors were selected from the four aspects of human-vehicle-road-environmental factors, and multivariate logistic regression analysis was used to screen the significant factors affecting the severity of accidents at different levels. The results show that the significant influencing factors leading to the death or severe injury accidents are 49~60 months of driving age, the type of vehicle of the participant is semi-trailer, bus, motorcycle, the speed at the time of the accident is more than 60 km/h, the vehicle collides with the road fixture, the accident occurs in April, May and June, and the accident time period is 0:00~3:00; The significant influencing factors of accident severity as no injury or minor injury are the vehicle type of the participant is a small car, the speed at the time of the accident is 0~40 km/h, the vehicle state is stationary, the road lights are lit, there is no longitudinal slope, the weather is sunny, and the time period of the accident is 5:00~6:00. The research results can provide a reference and basis for traffic accident prevention and risk assessment on low-grade highways in Guangxi, and have practical significance for improving the safety level of low-grade road transport in Guangxi.

Keywords

Low-Grade Highway, The Severity of the Accident, Influencing Factors, Multivariate Logistic Regression Analysis

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

广西低等级公路道路条件复杂、混合交通现象较多，交通违法现象普遍存在，极易导致各种交通事故。根据《广西综合交通运输发展“十四五”规划》可知，广西公路通车总里程达到 13.16 万公里，其中低等级公路里程将近 10 万公里，占比约 76%。据统计，2012~2021 年期间，广西交通事故发生数量多于 1 万起，主要归因于低等级公路交通事故发生率的逐年增加，因此应当予以重视。

道路安全与交通事故分析一直是交通研究的重要方向。例如，蔡晓男等[1]利用有序 Logit 模型研究了雨天条件下高速公路驾驶风险。张小玉[2]对低等级公路交通安全设施存在问题进行研究。周穆雄等[3]分析了农村低等级公路重大事故的事故形态与各项事故要素间的相关性。毛力增等[4]采用了调查样本与分析数据的方法，研究了广西的农村交通事故与影响因素。司远[5]分析了低等级公路的安全影响因素，研究了低等级公路交通安全的预警方法。

已有大部分研究中仅分析事故数据的整体规律及特性，或在特定条件下对交通事故致因风险进行研究，对于不同水平严重程度事故之间显著性影响因素的差异性分析研究较少，导致其显著影响因素会被减弱或隐藏[6]。另外，专门针对低等级公路的交通事故风险评价研究相对较少。因此，本文通过整理 2012~2021 年间广西交通管理部门提供的交通事故统计信息，从人员、车辆、道路、环境四个方面提取 48 个潜在影响因素，利用多因素 Logistic 回归分析挖掘影响广西低等级公路不同水平事故严重度的显著

性因素，由此提出减少广西低等级公路交通事故伤亡的建议和措施。

2. 低等级公路交通事故统计

2.1. 事故数据

本文选取 2012~2021 年间广西低等级公路发生的 1368 起交通事故数据(含驾驶人特性、车辆特性、道路特性、环境特性、事故发生时间、发生地点、事故严重程度等信息)。在数据预处理过程中，共剔除 144 起不完整的信息数据，剩余 1224 起事故纳入下一步数据分析中。不同水平事故严重度对应的事故数量及占比如表 1 所示。

Table 1. Accident severity statistics of low-grade highway in Guangxi between 2012~2021

表 1. 2012~2021 年广西低等级公路事故严重度统计数据

事故严重度	事故起数	
	数量	占比(%)
死亡	645	47.1
中度受伤	240	17.5
重度受伤	180	13.1
未受伤或轻微受伤	159	11.6
/	144	10.7

2.2. 变量选取

本文以交通事故严重度为因变量，从人员 - 车辆 - 道路 - 环境因素 4 方面选择了 48 个潜在自变量开展广西低等级公路交通事故严重程度评估研究。各变量的分类及分组编码见表 2。

Table 2. Independent variable statistics

表 2. 自变量统计

事故变量	分类描述	变量分组编码	
	年龄(岁)	1: 0~25	2: 26~40
		3: 41~64	4: 65~
	性别	1: 男	2: 女
		驾驶人特性	1: 0~12
驾龄(月)	3: 25~36	4: 37~48	
	5: 49~60	6: 61~72	
	7: 73~		
文化程度	1: 小学	2: 初中	
	3: 高中	4: 大学	

Continued

	是否饮酒	1: 是	2: 否
	是否吸毒	1: 是	2: 否
	心情是否正常	1: 是	2: 否
	注意力是否集中	1: 是	2: 否
	视野是否受限	1: 是	2: 否
	是否超速	1: 是	2: 否
	是否熟悉环境	1: 是	2: 否
	是否闯红灯	1: 是	2: 否
	是否按车道行驶	1: 是	2: 否
	是否做好安全防护	1: 是	2: 否
	是否有违规经历	1: 是	2: 否
		1: 半挂车	2: 电动二轮车
		3: 货车	4: 客车
	参与方车辆类型	5: 摩托车	6: 三轮车
		7: 小型轿车	8: 自行车
		9: 行人	
	是否为自动挡	1: 是	2: 否
		1: 左侧碰撞	2: 右侧碰撞
	碰撞类型	3: 正碰	4: 刮碰
		5: 追尾	6: 未碰撞
车辆特性	事故时车速(km/h)	1: 0~20	2: 21~40
		3: 41~60	4: 61~
车辆服役时间(月)		1: 0~12	2: 13~24
		3: 25~36	4: 37~48
		5: 49~60	6: 61~72
		7: 73~	
车辆行驶状态		1: 直行	2: 左转
		3: 右转	4: 静止
与道路基础设施碰撞		1: 撞护栏端头	2: 撞灯杆
		3: 撞墙体	4: 撞树
		5: 撞石头	

Continued

道路特性	道路线型	1: 直线路段 3: 十字路口	2: 转弯路段 4: T形路口
	路面材料	1: 沥青	2: 水泥
	路基类型	1: 路堤	2: 路堑
	道路表面状况	1: 良好	2: 坑洼
	道路是否湿滑	1: 是	2: 否
	有无障碍物	1: 有	2: 无
	有无信号控制	1: 有	2: 无
	有无防护栏	1: 有	2: 无
	路段车流量	1: 较多 3: 一般	2: 较少
	有无非机动车道	1: 有	2: 无
	有无纵坡	1: 有	2: 无
	机非混行是否存在标线	1: 是	2: 否
	机非混行是否存在标志	1: 是	2: 否
	有无路口横向减速线	1: 有	2: 无
	有无路口提醒标志	1: 有	2: 无
	有无限速标志	1: 无 3: 有, 限速 30 km/h 5: 有, 限速 50 km/h 7: 有, 限速 70 km/h	2: 有, 限速 20 km/h 4: 有, 限速 40 km/h 6: 有, 限速 60 km/h 8: 有, 限速 80 km/h
	有无警示柱	1: 有	2: 无
中央分隔情况	1: 无 3: 绿化带	2: 分割线 4: 中央护栏	
有无道路灯	1: 有	2: 无	
道路灯密度	1: 成列的道路灯	2: 无道路灯	
环境特性	天气	1: 晴 3: 阴 5: 雪	2: 雨 4: 雾 /
	照明条件	1: 日间 3: 晨	2: 夜间 4: 昏

Continued

是否降水	1: 是	2: 否
年份	1: 2011 年	2: 2012 年
	3: 2013 年	4: 2014 年
	5: 2015 年	6: 2016 年
	7: 2017 年	8: 2018 年
	9: 2019 年	10: 2021 年
	月份	1: 1 月
3: 3 月		4: 4 月
5: 5 月		6: 6 月
7: 7 月		8: 8 月
9: 9 月		10: 10 月
11: 11 月		12: 12 月
具体时间段	0:00~1:00:1	1:00~2:00:2
	2:00~3:00:3	3:00~4:00:4
	4:00~5:00:5	5:00~6:00:6
	6:00~7:00:7	7:00~8:00:8
	8:00~9:00:9	9:00~10:00:10
	10:00~11:00:11	11:00~12:00:12
	12:00~13:00:13	13:00~14:00:14
	14:00~15:00:15	15:00~16:00:16
	16:00~17:00:17	17:00~18:00:18
	18:00~19:00:19	19:00~20:00:20
	20:00~21:00:21	21:00~22:00:22
	22:00~23:00:23	23:00~24:00:24

3. 多因素 Logistic 回归统计模型

评估一个因变量和多个自变量间的相关关系通常使用多因素线性回归的方法。Logistic 回归的因变量可以是两分类的,也可以是多分类的。考虑到本文因变量为 4 分类,故采用多因素 logistic 回归分析筛选确定低等级公路交通事故严重度显著性影响因素。多因素有序 Logistic 模型的表达式[7]为:

$$P(Y \leq m|X) = \frac{\exp(\alpha_m + \sum_{k=1}^n \beta_k x_k)}{1 + \exp(\alpha_m + \sum_{k=1}^n \beta_k x_k)} \quad (1)$$

$$P(Y = 2|X) = P(Y \leq 2|X) - P(Y \leq 1|X) \quad (2)$$

$$P(Y \leq 2|X) = 1 - P(Y \leq 2|X) \quad (3)$$

式中： Y 为事故严重程度等级； X 为自变量对应的向量； $P(Y \leq m|X)$ 为第 m 事故严重程度等级的累积概率； β_k 为第 k 个自变量的回归系数； K 为自变量的总数；为第 m 事故严重程度等级的常数项；为第 k 个自变量， $k=1,2,\dots,K$ 。

4. 模型分析评估

4.1. 模型初步检验

由表3模型拟合信息可知，显著性值为0，小于0.05，说明模型具有统计意义，该模型通过检验。

Table 3. Model fit information

表 3. 模型拟合信息

模型	模型拟合条件		似然比检验	
	-2 对数似然	卡方	自由度	显著性
仅截距	3884.665	/	/	/
最终	1927.359	1957.306	1365	0.000

4.2. 模型拟合度检验

由表4可知，假设模型对原始数据的拟合优度较高，其中显著性为1，表明假设成立，模型对原始数据的拟合通过检验。由表5可知，依次列出的3个伪R方值(可决系数)最高为0.808，说明模型拟合度高，只有一小部分信息无法解释。

Table 4. Goodness-of-fit

表 4. 拟合优度

名称	卡方	自由度	显著性
相关系数	3323.345	5425	1.000
偏差	1927.359	5425	1.000

Table 5. Pseudo-R square

表 5. 伪R方

评价指标	指标值
Cox 及 Snell	0.761
Nagelkerke	0.808
McFadden	0.504

4.3. 模型预测

参考表6中预测总体正确百分比达73.9%，且各分类事故严重程度预测值正确百分比均等于高于50%，最高为82.8%，说明模型预测能力较高。

Table 6. Classification (accident severity sub division)
表 6. 分类(事故严重度合计划分)

实测值	预测值					正确百分比
	重度受伤	中度受伤	未受伤或轻微受伤	死亡	/	
重度受伤	90	28	13	45	4	50.0
中度受伤	13	148	37	29	13	61.7
未受伤或轻微受伤	9	19	113	11	7	71.1
死亡	44	26	14	534	27	82.8
总体百分比/%	11.5	16.5	13.1	45.9	13	73.9

5. 事故风险评估

5.1. 风险影响因素统计结果

经过统计筛选，提取得到影响不同水平事故严重度的显著性影响因素各项指标值见图 1~4。图 1~4 中 B 为模型变量的回归系数，回归系数 B 的取值范围为 $|B| \leq 1$ ，B 的绝对值越趋近于 1，变量间的相关关系越强；反之，相关关系越弱。Exp(B) 为每个自变量的优势比，即事件发生概率 P 与事件不发生概率 $1 - P$ 之比。

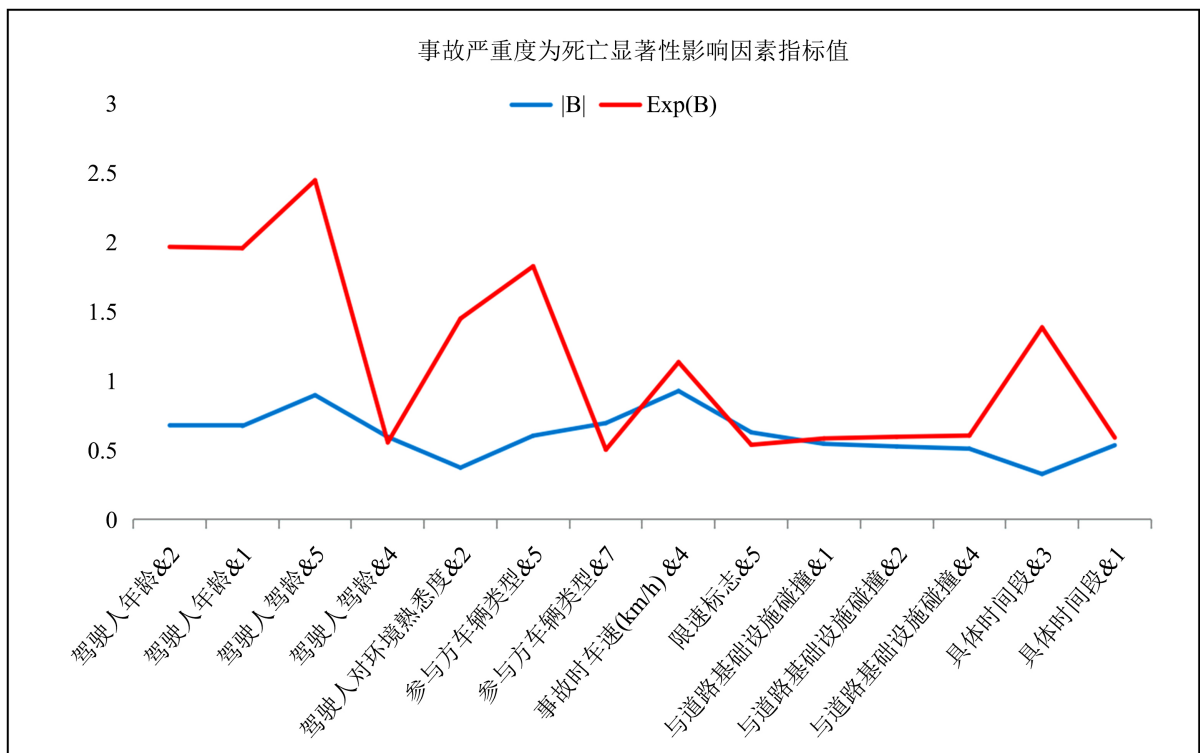


Figure 1. Significance influence factor index value causing death
图 1. 事故严重度为死亡的显著性影响因素指标值

图 1 为事故严重度为死亡显著性影响因素指标值。驾龄为 49~60 个月的驾驶人、参与方车辆类型为摩托车、发生事故时车速为 60 km/h 以上和事故时间段为凌晨 2:00~3:00 是导致事故严重度为死亡的主要原因,其优势比分别为 2.446、1.825、1.132、1.383, $|B|$ 分别为 0.894、0.601、0.924、0.324。同时,死亡事故与驾驶人年龄及对其环境熟悉度、车辆是否与道路基础设施碰撞、限速 50 km/h 的路段有很大的关联。

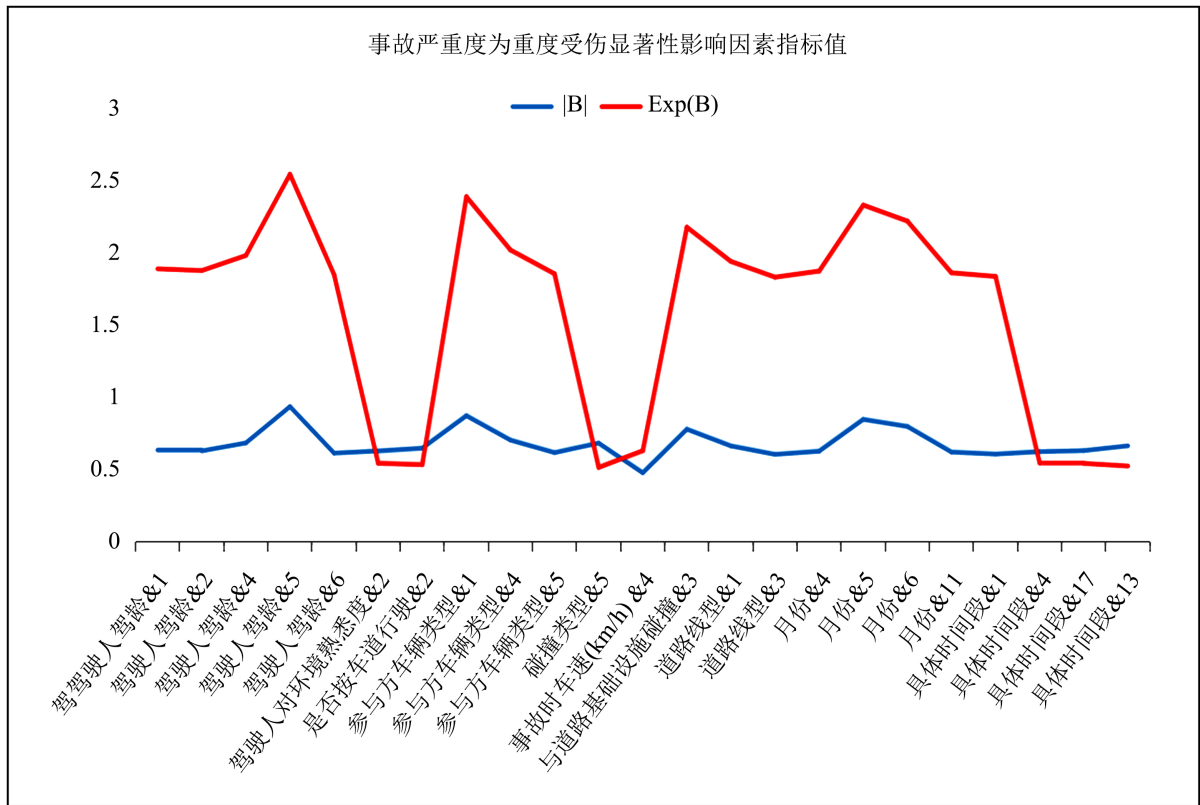


Figure 2. Significance influence factor index value causing severe injuries

图 2. 事故严重度为重度受伤的显著性影响因素指标值

图 2 为事故严重度为重度受伤显著性影响因素指标值。造成重度受伤的主要影响因素有驾驶人驾龄为 49~60 个月,参与方车辆类型为半挂车、客车和摩托车,事故发生月份为 4 月、5 月、6 月,事故时间段 0:00~1:00,车辆与墙体碰撞 5 个因素,其优势比分别为 2.526、2.372、2.005、1.841、1.860、2.314、2.204、1.823, $|B|$ 分别为 0.927、0.864、0.696、0.610、0.620、0.839、0.790、0.600。

由图 3 可知,事故严重度为中度受伤的显著性影响因素包括 0~25 岁的驾驶人、驾驶人未超速、参与方车辆类型为半挂车、服役时间为 25~36 个月的车辆、与灯杆碰撞、无障碍物、阴天,其优势比分别为 2.484、2.064、2.28、2.613、2.524、2.18、2.115, $|B|$ 分别为 0.91、0.725、0.824、0.96、0.926、0.779、0.749。

由图 4 可知,事故严重度为未受伤或轻微受伤显著性影响因素包括参与方车辆类型为小型轿车、事故时车速为 0~40 km/h、车辆状态为静止、道路灯点亮、无纵坡、晴天、具体时间段为 5:00~6:00,其优势比分别为 2.413、2.002、2.115、2.116、2.048、1.776、2.661、2.649, $|B|$ 分别为 0.881、0.694、0.749、0.75、0.717、0.569、0.979、0.974。

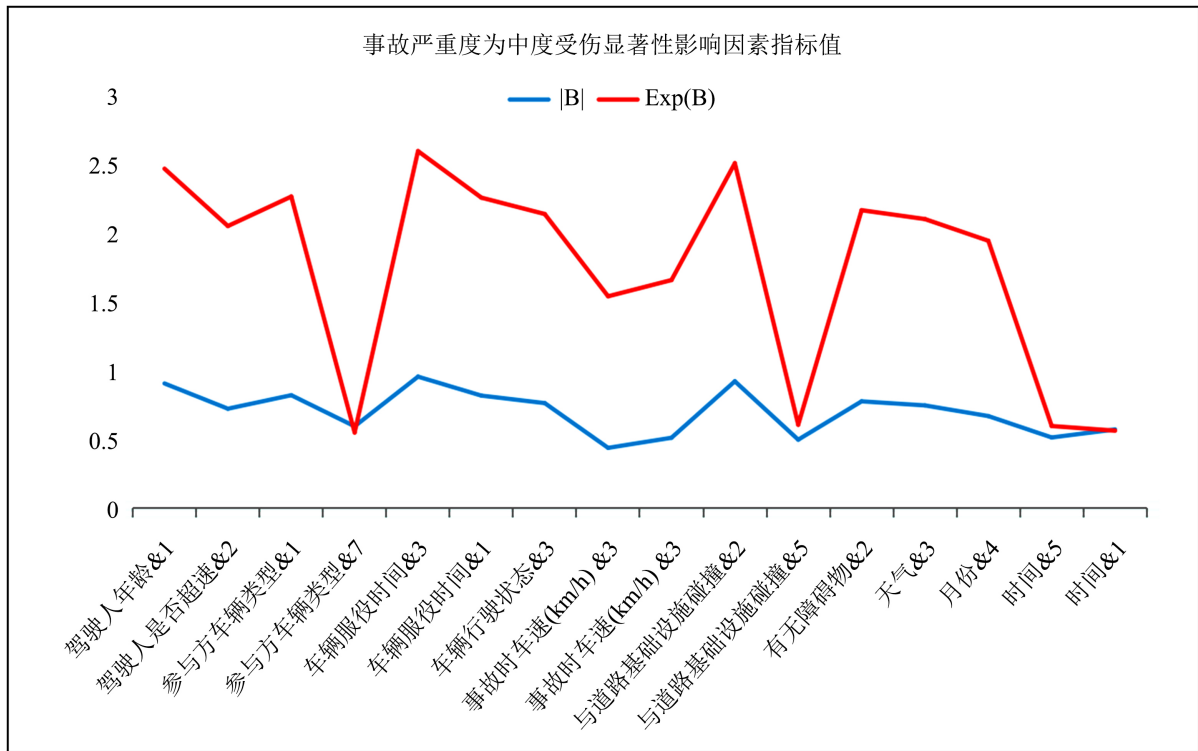


Figure 3. Significance influence factor index value causing moderate injuries
图 3. 事故严重度为中度受伤的显著性影响因素指标值

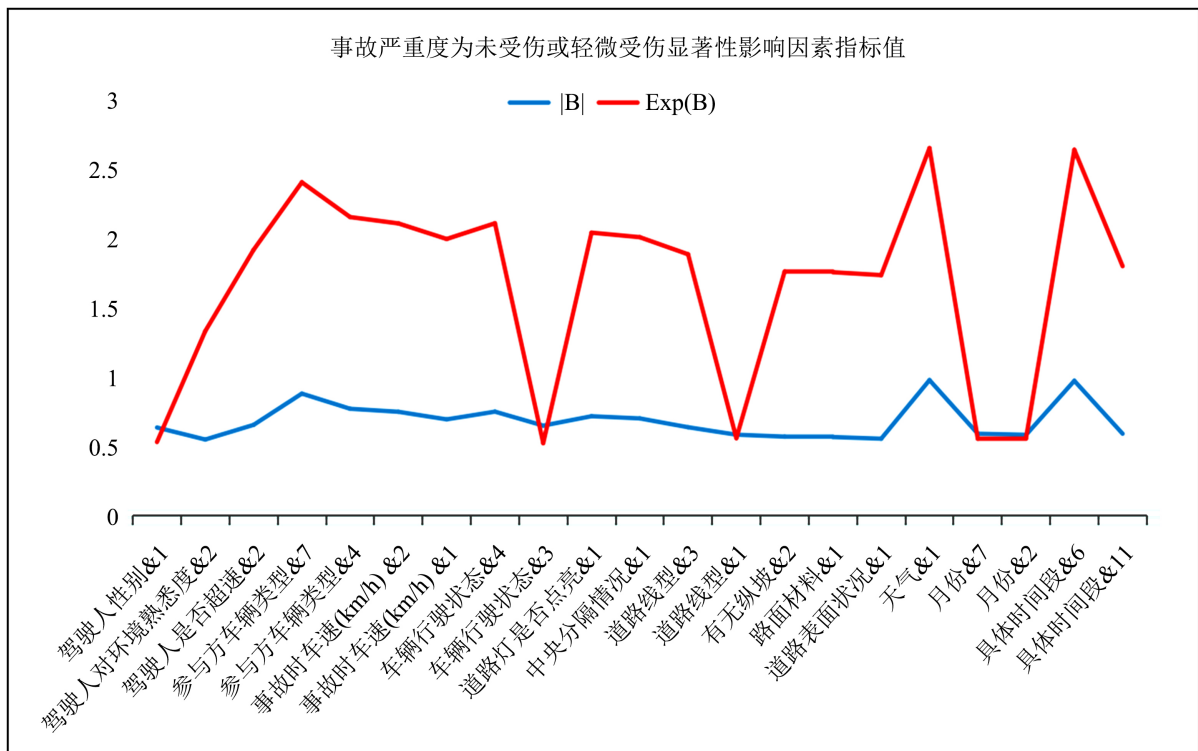


Figure 4. Significance influence factor index value causing no or minor injuries
图 4. 事故严重度为未受伤或轻微受伤的显著性影响因素指标值

5.2. 结果分析

根据筛选得到的不同水平严重度显著性影响因素, 分析认为导致事故严重度为死亡或重伤的原因有以下五个方面: 1) 驾龄为 49~60 个月的驾驶人自认为驾龄长而产生骄傲心理, 从而放松对行车风险的警惕性。2) 骑摩托车者通常在事故发生时暴露更多, 安全防护措施更少。3) 事故严重程度与行车速度密切相关, 事故车速越高造成人员伤亡概率更高。4) 在驾驶人对环境不熟悉且不按车道行驶的情况下, 车辆更易与路侧墙体、护栏端头、灯杆、树等道路基础设施发生碰撞, 导致严重事故。5) 夏季中的 0:00~3:00 为事故多发时段。夏季是疲劳驾驶的高发季且驾驶人衣着单薄, 事故中更易受到损伤; 0:00~3:00 时段为夜间, 驾驶人视野范围受限, 更易导致事故发生。

导致事故严重度为未受伤或轻微受伤的原因有以下五个方面: 1) 小型轿车安全性能和安全系数高, 车辆各方面的安全防护装置较完备。2) 较低行车速度或车辆在静止状态下, 事故人员受伤程度较轻。3) 道路基础设施完善, 如道路灯点亮、道路无纵坡有利于行车安全。4) 晴天能见度高, 视野开阔, 驾驶人对周围环境的判断更准确。5) 在 5:00~6:00、10:00~11:00 等交通流非高峰期驾驶人受外界干扰少。

针对上述分析原因提出以下建议降低事故损失: 1) 加强对驾驶员尤其是有一定驾龄的驾驶员的培训力度, 提高驾驶人驾驶技术和综合素质; 加强“一盔一带”行动的宣传力度, 增强群众佩戴安全头盔、使用安全带的意识。2) 严格落实车辆的保养、年检制度, 出车前、行车中、收车后都要勤检勤查, 发现故障及时排除。3) 事故多发路段设置有效的限速装备, 严格按照交通标志规定行驶, 对违法者加大处罚力度。4) 保持完善的道路基础设施建设并进行定期养护, 不断优化道路线形, 为驾驶人提供一个畅通舒适的行车环境。

6. 结论

本文以 2012~2021 年十年间广西低等级公路交通事故数据为基础, 建立多因素 logistic 回归分析模型对广西低等级公路事故严重度的显著性影响因素进行了评估研究, 主要得到以下结论:

1) 导致事故严重度为死亡或重伤的显著性影响因素有: 驾驶人驾龄为 49~60 个月, 参与方车辆类型为半挂车、客车、摩托车, 发生事故时车速为 60 km/h 以上, 车辆与道路固定物碰撞, 事故发生月份在 4 月、5 月、6 月, 事故发生时间段在凌晨 0:00~3:00。

2) 事故严重度为未受伤或轻微受伤的显著性影响因素有: 参与方车辆类型为小型轿车、事故时车速为 0~40 km/h、车辆状态为静止、道路灯点亮、无纵坡、晴天、事故发生时间段在 5:00~6:00。

3) 从人-车-道路-环境四方面提出降低广西低等级公路交通事故损失相关建议: ① 增强驾驶人驾车技术、综合素质和安全意识。② 确保车辆安全性能、安全系数, 完备安全防护措施。③ 事故多发路段设置严格的限速装备, 对违法者加大处罚力度。④ 保持完善的道路基础设施建设并进行定期养护, 不断优化道路线形设计, 提高驾驶环境的安全舒适性。

基金项目

桂林电子科技大学大学生创新创业训练计划项目(S202210595280)资助。

参考文献

- [1] Xiaonan, C., Chen, W., Shengdi, C., et al. (2017) Model Development for Risk Assessment of Driving on Freeway under Rainy Weather Conditions. *PLOS ONE*, 12, e0149442. <https://doi.org/10.1371/journal.pone.0149442>
- [2] 张小玉. 低等级公路交通安全设施存在的问题研究[J]. 工程科技 II 辑, 2019(16): 12-13.
- [3] 周穆雄, 高岩, 王峻极, 尤志栋. 基于 Logistic 回归的低等级公路重大交通事故分析[C]//第十一届中国智能交通

年会论文集, 2016: 334-340.

- [4] 毛力增, 段里仁, 毛恩荣. 广西农村机动车驾驶入交通事故与影响因素分析[J]. 中国卫生统计, 2008, 25(3): 286-288.
- [5] 司远. 低等级公路交通安全预警研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2014.
- [6] Sun, M., Sun, X. and Shan, D. (2019) Pedestrian Crash Analysis with Latent Class Clustering Method. *Accident Analysis & Prevention*, **124**, 50-57. <https://doi.org/10.1016/j.aap.2018.12.016>
- [7] 杨文臣, 谢碧珊, 房锐, 秦雅琴. 山区双车道公路机动车碰撞事故严重度致因比较分析与预测[J]. 交通运输系统工程与信息, 2021, 21(1): 190-195.