

DOI:10.20035/j.issn.1671-2668.2022.06.008

基于 NAIS 数据的交叉口车-车事故严重程度影响因素分析

肖乐^{1,2}

(1.西华大学 汽车与交通学院, 四川 成都 610039;2.汽车测控与安全四川省重点实验室, 四川 成都 610039)

摘要:为研究道路交叉口车-车事故严重程度影响因素,分析人、车、路、环境与交叉口事故严重程度之间的关系,以国家车辆事故深度调查体系(NAIS)中 471 起交叉口事故数据为样本,将伤亡程度作为因变量,通过聚类压缩、平行性检验选取 13 个因素作为自变量,构建有序 Logistic 模型分析对事故严重程度影响显著的因素。结果表明,车辆类型、事故发生时段、是否有路灯、路口是否有信号灯、驾驶员主要过失均显著影响交通事故严重程度;驾驶员存在超速驾驶违法行为、涉及非机动车的交叉口交通事故最严重;交通信号控制设施越完善、照明条件越好事故严重性越低。

关键词:交通安全;交叉口;车-车事故严重程度;有序 Logistic 模型;影响因素

中图分类号:U491.31

文献标志码:A

文章编号:1671-2668(2022)06-0037-04

道路交叉口是多种交通流交汇与冲突的关键区域,是道路交通事故的多发地,研究道路交叉口事故严重程度影响因素对事故预防具有重要意义。牛志鹏等从道路条件、交通环境、安全设施管理三方面对道路平面交叉路口的交通安全进行了定性和定量分析。赵晓华等基于驾驶行为数据构建交叉口进口道安全性结构方程,挖掘风险因素,结果表明交叉口各方向车道总数、违法监控设备数量等对交叉口安全性具有显著影响。吕通通等构建互信息贝叶斯网络模型,分析了各影响因素变化与事故严重程度的定量关系。赵丹等构建双变量 Probit 模型,以事故形态和事故严重程度为因变量,分析了农村交叉口交通事故的风险因素。Islam S.等建立随机参数 Logistic 模型,对比分析了农村和城市道路中摩托车事故严重程度影响因素的差异性。温惠英等以英国单车事故数据为样本构建多项式 Logit 模型,分析了交叉口单车事故严重程度影响因素。Ariana Vorko-Jovic 等研究了城市道路交通事故的发生及严重程度与人、道路、环境、道路信号灯设计、交通管理部门管理等因素之间的关系。上述研究多采用国外开源数据,难以反映具有中国特色的道路交通事故特征,且部分研究以国内某一城市的事故统计数据为研究对象,难以表征道路交通事故的共性。鉴于此,本文统计、分析国家车辆事故深度调查体系(National Automobile Accident In-Depth Investi-

gation System, NAIS)中的交叉口车-车事故数据,研究道路交叉口事故严重程度影响因素,为交叉口事故预防提供理论支撑。

1 数据概述及变量描述

1.1 数据采集

NAIS 由国家市场监督管理总局缺陷产品管理中心联合 8 所具有较深事故研究背景的高校、事故鉴定机构及科研机构共同建立,其目的是采集具有中国道路交通事故特征的深度数据,并建立汽车主被动安全研究基础数据库。2011—2019 年已收集 4 000 多起道路交通事故案例,覆盖全国 7 个地区,包括东北、华南、西南、华北、华东等地。数据主要由交通事故采集人通过到事故现场复勘和交警部门获得,包括事故照片、编码的数据、警方资料、事故现场视频、事故重构文件、CAD 事故现场图及事故分析报告等,数据完整、详细,对交通事故研究具有较高价值。本文筛选交叉口(三枝、四枝)车-车事故数据共计 556 条,剔除并清洗信息记录不全数据后,得到有效数据 471 条。

1.2 变量选取及说明

1.2.1 因变量

根据初始数据将事故严重程度分为 4 个等级,分别为轻微事故(仅轻伤)、一般事故(重伤,未死亡)、重大事故(死亡一两人)、特大事故(死亡 3 人及

以上),由于特大事故仅占0.002%,将其与重大事故归为一类,最终将人员伤亡事故严重程度由低到高划分为3个等级,分别为轻微事故(仅轻伤)、一般事故(重伤,无死亡)、重大事故(死亡1人及以上),频数分别为40、210、221。因变量编码及描述见表1。

表1 因变量编码

事故特性	变量名称	取值	比例/%
事故严重程度	轻微事故	1	8.49
	一般事故	2	44.59
	重大事故	3	46.92

1.2.2 自变量

调研国内外学者对交叉口事故影响因素选取情

况,结合原始数据事故特征,从驾驶员、车辆、道路、环境4个维度选取13个离散型变量作为自变量,分别为驾驶员主要过失、性别、天气、事故发生时段、参与车辆类型、道路行政等级、路段类型、路口是否有信号灯、路面状况、道路线形、路灯是否开启、是否有交通限速标志、对向机动车道隔离形式。由于道路交叉口流量大、冲突多,车型构成复杂,车辆安全状况参差不齐,选取8种车型共4类用于分析车辆类型的影响。结合历年交叉口交通违法行为统计数据,选择超速驾驶、酒后驾驶、未按规定让行和其他违法操作分析驾驶员主要过失对事故特性的影响(见表2)。

表2 自变量设置说明

变量名称	变量代码	变量类型	变量定义
路段类型	X_1	绝对值自变量	1=三枝路口,2*=四枝路口
天气	X_2	绝对值自变量	1=晴,2=阴,3=雨,4*=雪/雾
事故发生时段	X_3	绝对值自变量	1=日间,2=夜间,3*=晨昏
参与车辆类型	X_4	绝对值自变量	1=非机动二/三轮车,2=机动二/三轮车,3=货车,4*=乘用车
性别	X_5	绝对值自变量	1=女性,2*=男性
道路行政等级	X_6	绝对值自变量	1=公路,2=城市道路,3=厂矿道路,4*=乡村道路
路口是否有信号灯	X_7	绝对值自变量	0=否,1*=是
道路线形	X_8	绝对值自变量	1=弯/坡,2*=平直
路灯是否开启	X_9	绝对值自变量	0=否,1*=是
路面状况	X_{10}	绝对值自变量	1=湿/积雪,2*=干
是否有交通限速标志	X_{11}	绝对值自变量	0=否,1*=是
驾驶员主要过失	X_{12}	绝对值自变量	1=超速驾驶,2=酒后驾驶,3=未按规定让行,4=其他违法操作,5*=无过失
对向机动车道隔离形式	X_{13}	绝对值自变量	1=无隔离,2*=有隔离

注:*代表变量对照组。

2 有序 Logistic 回归模型构建

2.1 模型构建

有序 Logistic 回归是针对有序多分类变量的一种统计分析方法,而交通事故严重程度预测是一个有序多级别划分的非线性问题。设有有序多分类变量 $Y_i (i \in \{1, 2, \dots, n\})$ 有 j 个等级,取值为 $1, 2, \dots, j$, X 为 m 个自变量 x_1, x_2, \dots, x_m , 则有序 Logistic 回归模型表达式为:

$$\ln\left(\frac{P(Y_i \leq j | X)}{1 - P(Y_i \leq j | X)}\right) = \alpha_j + \sum_{m=1}^M \beta_{jm} x_m \quad (1)$$

式中: $P(Y_i \leq j | X)$ 为累积概率, $P(Y_i \leq j | X) = 1$; α_j 为常数项; M 为自变量数量; β_{jm} 为回归系数。

有序 Logistic 模型共有 $j-1$ 个。有序 Logistic

的概率模型表示为:

$$P(Y_i \leq j | X) = \frac{\exp(\alpha_j + \sum_{m=1}^M \beta_{jm} x_m)}{1 + \exp(\alpha_j + \sum_{m=1}^M \beta_{jm} x_m)} \quad (2)$$

2.2 模型适用性检验

2.2.1 变量的多重共线性检验

在进行模型拟合之前,采用方差膨胀因子对13个变量之间的共线性情况进行诊断。方差膨胀因子是容忍度的倒数,用于判断是否存在多重共线性,其值为0~10表示可以接受。结果表明13个自变量之间无多重共线性(见表3)。

2.2.2 平行性检验

平行性检验结果见表4。卡方值为24.636,显著性 P 为 $0.369 > 0.05$,接受原假设,有序Logistic回归模型合适。

表3 变量的多重共线性诊断

变量	容差	方差膨胀因子	变量	容差	方差膨胀因子
X_1	0.825	1.212	X_8	0.963	1.038
X_2	0.570	1.754	X_9	0.660	1.516
X_3	0.681	1.469	X_{10}	0.587	1.704
X_4	0.912	1.096	X_{11}	0.816	1.225
X_5	0.973	1.028	X_{12}	0.932	1.073
X_6	0.911	1.097	X_{13}	0.841	1.189
X_7	0.664	1.506			

表4 平行性检验结果

模型	-2对数似然值	卡方	自由度	显著性 P
零假设	615.664			
广义	591.028	24.636	23	0.369

2.2.3 Pearson X^2 检验和偏差统计量

Pearson X^2 检验通过频数检验模型成立的假设,其标准 X^2 统计量计算公式为:

$$X^2 = \sum_k^K \frac{(O_k - E_k)^2}{E_k} \quad (K = 1, 2, \dots, k) \quad (3)$$

式中: O_k 为第 k 类协变类型的实际频数; E_k 为第 k 类协变类型的期望频数; K 为协变类型的种类数。

X^2 的自由度为协变类型数目与参数数目之间的差距, X^2 统计量的值越小,模型拟合效果越好。

在Logistic回归模型中,将偏差统计量 D 视为拟合优度统计量,计算公式为:

$$D = -2(\ln \bar{L}_s - \ln \bar{L}_f) \quad (4)$$

式中: \bar{L}_s 为拟合模型估计的最大似然值; \bar{L}_f 为饱和模型估计的最大似然值。

\bar{L}_s 的值和 \bar{L}_f 的值越接近, D 值越小,模型的拟合优度越好。

Pearson X^2 检验结果见表5。由表5可知:Pearson X^2 统计量的显著性 P 为0.518,偏差统计量的 P 值为1.000,模型的拟合效果好。

表5 Pearson X^2 检验结果

拟合优度指标	卡方	自由度	显著性 P
Pearson	690.679	693	0.518
偏差统计量	556.931	693	1.000

2.2.4 数据的单变量观察

利用SPSS描述性统计中的交叉表,依次对13

个指标与事故严重程度之间的关系进行卡方检验,检验结果见表6。由表6可知:路段类型、性别、路口是否有信号灯、道路线形、路面状况、对向机动车道隔离形式与事故严重程度没有显著关系。自变量筛选过程中,结合模型的Wald检验结果,逐步剔除无统计学意义的自变量,剔除顺序依次为路段类型、性别、道路线形、路面状况、对向机动车道隔离形式。剔除过程中发现路口是否有信号灯的显著性接近0.05,暂不予以剔除。

表6 单变量显著性检验结果

变量	显著性 P (双侧)	变量	显著性 P (双侧)
X_1	0.897	X_8	0.882
X_2	0.039	X_9	0.001
X_3	0.007	X_{10}	0.230
X_4	0.000	X_{11}	0.002
X_5	0.395	X_{12}	0.081
X_6	0.025	X_{13}	0.732
X_7	0.269		

3 模型计算结果分析

有序Logistic模型计算结果见表7。拟合优度检验中 $P < 0.001$,拒绝原假设,有序Logistic模型整体有意义。

表7 有序Logistic模型估计结果

自变量	偏回归系数 β	优势比 $\exp(\beta)$	显著性 P
晴	0.315	1.370	0.639
阴	0.639	1.895	0.355
雨	0.139	1.149	0.843
雪/雾	0.000	—	—
日间	-0.917*	0.400	0.013
夜间	-0.681	0.506	0.091
晨昏	0.000	—	—
超速驾驶	1.053*	2.866	0.024
酒后驾驶	0.321	1.379	0.624
未按规定让行	0.964*	2.622	0.022
其他违法操作	0.313	1.368	0.433
驾驶员无过失	0.000	—	—
非机动二/三轮车	2.743**	15.534	0.001
机动二/三轮车	1.667	5.296	0.079
货车	2.603**	13.504	0.001
乘用车	0.000	—	—
公路	1.359	3.892	0.358
城市道路	0.825	2.282	0.577

续表 7

自变量	偏回归 系数 β	优势比 $\exp(\beta)$	显著性 P
厂矿道路	0.661	1.937	0.661
乡村道路	0.000	—	—
路口无信号灯	0.463	1.589	0.062
路口有信号灯	0.000	—	—
无路灯	0.732*	2.079	0.043
有路灯	0.000	—	—
无交通限速标志	-0.410	0.664	0.075
有交通限速标志	0.000	—	—

注：* 表示 $P < 0.05$ (显著)；** 表示 $P < 0.01$ (极显著)。

由表 7 可知：

(1) 驾驶员。驾驶员存在过失会对事故严重程度产生显著正影响。以驾驶员无过失事故严重程度为参照，超速驾驶、未按规定让行的偏回归系数分别为 1.053、0.964，超速驾驶对事故严重程度影响最显著，其次是未按规定让行，分别使交叉口事故严重程度上升 286.6%、262.2%。

(2) 车辆类型。非机动车二/三轮车和货车对事故严重程度产生显著正影响。以乘用车为参照，参与车辆涉及非机动车二/三轮车的偏回归系数为 2.743，对事故严重程度影响最大；其次是货车和机动二/三轮车，偏回归系数分别为 2.603、1.667。

(3) 道路。路口是否有信号灯、路灯是否开启对交叉口事故严重程度产生显著正影响。以路口有信号灯为参照，路口无信号灯条件下交叉口事故严重程度上升 158.9%。以有路灯为参照，无路灯情况下事故严重程度上升 207.9%。可见，完善的交通信号控制设施对降低事故严重程度极为重要。

(4) 环境。事故发生时段对事故严重程度有显著影响。以事故发生时间为晨昏为参照，事故发生时间为日间的偏回归系数为 -0.917，与事故严重程度呈显著负相关。相比于晨昏，日间发生的交通事故严重程度更轻，事故严重程度降低 40%。

4 结论

以 NAIS 数据库中交叉口事故数据为样本，从人、车辆类型、道路、环境四方面选取 13 个变量构建有序 Logistic 回归模型，分析车—车事故严重程度的影响因素，得出如下结论：1) 驾驶员过失、车辆类型、路口是否有信号灯、路灯是否开启、事故发生时段显著影响事故严重程度。2) 超速驾驶、车辆类型

为非机动二/三轮车、货车对事故严重程度影响最大；路口无信号灯、无路灯、事故发生时段也会使事故严重程度加重。

随着 NAIS 体系中数据量的逐年增加，未来可以综合更多的道路交叉口事故数据进行分析，以更准确地掌握交叉口车—车事故严重程度的影响机理。

参考文献：

- [1] BAGLOEE S A, ASADI M. Crash analysis at intersections in the CBD: A survival analysis model[J]. Transportation Research Part A: Policy and Practice, 2016, 94: 558—572.
- [2] 潘福全, 陆键, 项乔君, 等. 公路信号平面交叉口安全服务水平研究[J]. 东南大学学报(自然科学版), 2008, 38(2): 298—303.
- [3] 牛志鹏, 刘文佳, 吴东玲, 等. 城市道路平面交叉口交通安全审计研究[J]. 大连交通大学学报, 2019, 40(3): 17—22+29.
- [4] 赵晓华, 姚莹, 丁阳, 等. 基于导航数据的交叉口进口道安全风险评估及诊断方法[J]. 同济大学学报(自然科学版), 2020, 48(12): 1733—1741.
- [5] 吕通通, 张湛, 陆林军, 等. 基于互信息贝叶斯网络的交通事故严重程度分析[J]. 交通信息与安全, 2021, 39(6): 36—43.
- [6] 赵丹, 马社强, 张雨萌, 等. 农村公路交叉口交通事故特征关联性与风险因素分析[J]. 中国安全科学学报, 2020, 30(7): 146—151.
- [7] ISLAM S, BROWN J. A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama[J]. Accident Analysis & Prevention, 2017, 108(11): 163—171.
- [8] 温惠英, 汤左淦. 道路交叉口单车事故严重程度影响因素分析[J]. 公路工程, 2019, 44(2): 55—61+102.
- [9] ARIANA Vorko-Jovic, JOSIPA Kern, ZRINKA Biloglav. Risk factors in urban road traffic accidents[J]. Journal of Safety Research, 2006, 37(1): 93—98.
- [10] 肖向良. 电动自行车道路交通安全事故严重性影响因素分析[J]. 公路与汽运, 2020(6): 32—36.
- [11] 范凡. 基于累积 Logistic 模型的交叉路口轿车事故严重程度影响因素分析[J]. 大众标准化, 2021(11): 21—22+26.
- [12] 丁柏群, 宋子龙. 基于交通冲突的城市交叉口风险影响因素分析[J]. 森林工程, 2019, 35(5): 98—105.

收稿日期: 2022-02-24