

基于 Logistic 的山区连续长下坡路段交通事故严重程度分析

张娟, 朱文强, 王凯, 陶珂

(西安建筑科技大学 土木工程学院, 陕西 西安 710055)

摘要: 为研究山区连续长下坡路段交通事故影响因素, 以事故严重程度为因变量, 从人、车、路、环境4个方面归纳出天气状况、车辆类型、平面线形等11个因素对事故严重程度的影响, 应用分类变量方法将其定义为11种指标, 并进行变量编码。以陕西省2条山区连续长下坡路段三年共计169起交通事故为建模数据, 结合Logistic回归理论, 建立了各因素对山区连续长下坡路段交通事故严重程度的影响度量模型。结果表明: 坡度和事故地点是影响山区连续长下坡路段交通事故严重程度的主要因素, 其中, 非特殊路段发生事故伤亡的概率较小, 其发生伤亡的概率是特殊路段的32.4%。发生伤亡事故的概率随坡度增加呈递增趋势。

关键词: 交通工程; 山区连续长下坡路段; Logistic模型; 事故严重程度

中图分类号: U491

文献标志码: A

文章编号: 1006-7930(2017)06-0841-06

Analysis on traffic accident severity of continuous long downgrade section in the mountainous area based on Logistic

ZHANG Juan, ZHU Wenqiang, WANG Kai, TAO Ke

(School of Civil Engineering, Xi'an Univ. of Arch. & Tech., Xi'an 710055, China)

Abstract: In order to study the influence factors of traffic accidents on continuous long downgrade section in mountainous area, taking the accident severity as dependent variable, the influence of 11 factors including weather condition, vehicle type, road alignment and so on were induced from the human factors, vehicle, road, and environment concerned. These 11 indexes were then coded with own grade section in mountainous area in Shaanxi province were used for building the measurement model of traffic accident severity, which influenced by various factors at continuous long downgrade section in mountainous area based on logistic model theory. The results show that the accident position and slope are the important factors to affect the accident severity at the continuous long downgrade section in mountainous area. At the unrepresentative section, there is a low probability of injury and death, accounting for 32.4% of that at the representative section. The probability of occurrence of casualty accidents increases as the slope increases.

Key words: traffic engineering; continuous long downgrade section of mountainous area; logistic model; accident severity

连续长坡由于受地形地貌的限制, 部分路段线形的设计不可避免地采用极限值或接近极限值, 虽满足规范要求, 但相对于其他路段, 具有交通事故频发、重大及恶性交通事故多等特点^[1-2]。据统计, 与连续长坡路段有关的交通事故占山区公路交通事故总数的40%以上, 且下坡路段的事故率明显高于上坡路段^[3]。因此, 有必要研究山区连续长坡路段的交通事故成因, 为提高其安全性提供理论支持。

近年来, 越来越多的学者对连续长坡路段交

通事故展开了研究。陈斌等^[4]分析了连续长大下坡路段的急弯陡坡、连续陡坡和连续缓坡3种线形组合形式路段的交通事故特征和事故原因, 运用数理统计与回归分析方法, 研究了交通事故与道路纵断面参数之间的关系; 郭应时等^[5]研究了山区公路交通事故与平面线形的关系, 利用最小二乘法拟合事故率与角度变化率之间的曲线关系; 付锐等^[6]研究了连续下坡道路的纵坡坡度、坡长与事故率的关系, 发现事故率与平均坡度呈指数上升关系。但是, 这些研究均未从交通事故的严重程度考

虑来构建影响交通事故严重程度的度量模型，也不能分析出哪些因素对交通事故程度的影响较大？

由于交通事故的发生存在不确定性，即事故的发生与否是未知的，因此，许多学者提出利用 Logistic 回归模型来研究交通事故的严重程度与影响因素之间的关系。N. N. Sze 等^[7]研究了受伤地点、行人行为、道路类型等 17 个因素对行人受伤严重程度的影响；马壮林等^[8]分析了事故发生时段、碰撞类型、天气和日标准小客车交通量与 AADT 之比对公路隧道交通事故严重程度的影响；冯忠祥等^[9]研究了事故时间段、天气状况、能见度等 10 个因素对绕城高速交通事故严重程度的影响。以上的研究对象主要集中在城市道路、隧道、绕城高速，鲜有学者利用 Logistic 回归对山区连续长坡路段建立交通事故严重程度模型，基于此，本文以山区连续长下坡路段为研究对象，采用 Logistic 回归，构建交通事故严重程度模型，分析各因素对交通事故严重程度的影响。

1 山区连续长下坡路段事故严重性二分类 Logistic 模型

1.1 Logistic 模型原理

Logistic 回归为概率型非线性回归模型，是研究因变量与自变量之间关系的一种多变量分析方法。该方法通常先将取值在实数范围内的值通过

Logit 变换转化为目标概率值，然后进行回归分析。Logistic 回归参数的估计通常采用最大似然法。Logistic 模型的数学表达式为

$$\ln \frac{P}{1-P} = \alpha + X\beta + \epsilon \quad (1)$$

式中， P 为事件发生的概率； $\alpha = (\alpha_1 \ \alpha_2 \ L \ \alpha_n)^T$ 为模型的截距项； $\beta = (\beta_1 \ \beta_2 \ L \ \beta_n)^T$ 为待估计参

数； $X = \begin{pmatrix} x_{11} & x_{12} & L & x_{1k} \\ x_{21} & x_{22} & L & x_{2k} \\ M & M & O & M \\ x_{n1} & x_{n2} & L & x_{nk} \end{pmatrix}$ 为解释变量； $\epsilon = (\epsilon_1 \ L \ \epsilon_n)^T$ 为误差项。

1.2 自变量选取

交通事故的发生涉及到人、车、路、环境四大因素系统，是各系统之间共同作用产生的结果。综合以前学者的研究及连续长下坡路段交通事故特性^[10-14]，本文选取以下因素作为具体变量，并进行编码，详见表 1。

前 9 个影响因素编码后均为分类变量，但具体数值并不代表实际值，且碰撞类型、路表情形、车辆类型、过失违章类型和车辆故障 5 个因素虽有数值但非等距变化，因此，应按照虚拟变量进行设置。在实际模型计算过程中会对虚拟变量进行编码，如果虚拟变量有 k 个分类，则虚拟变量转换为 k 个变量。如不同碰撞类型分类编码如表 2 所示。

表 1 交通事故影响因素的变量编码及含义

Tab. 1 The influence factors of traffic accident of variable code and implication

序号	影响因素	变量	变量编码和含义
1	天气状况	x_1	晴=0，不利天气=1
2	事故时间段	x_2	白天(6: 00~18: 00)=0，晚上(18: 00~6: 00)=1
3	平面线形	x_3	直线段=0，曲线段=1
4	碰撞类型	x_4	正面碰撞=1，侧面碰撞=2，尾随碰撞=3，其他(翻车，坠车等)=4
5	路表情形	x_5	干燥=1，潮湿=2，冰雪=3
6	车辆类型	x_6	小汽车=1，客车=2，货车=3，其他=4
7	过失违章类型	x_7	违法违规=1，制动不当=2，其他操作不当=3，超速行驶=4，其他=5
8	事故地点	x_8	特殊路段(隧道、桥梁、村庄、峡谷等)=0，非特殊路段=1
9	车辆故障	x_9	制动失效=1，爆胎=2，其他=3
10	坡度	x_{10}	实际值
11	坡长	x_{11}	实际值

表 2 虚拟变量碰撞类型分类编码

Tab. 2 Categorical variable code of collisions type

碰撞类型	不同碰撞类型分类编码		
	碰撞类型 1	碰撞类型 2	碰撞类型 3
正面碰撞	0	0	0
侧面碰撞	1	0	0
尾随碰撞	0	1	0
其他	0	0	1

路表情形、车辆类型、过失违章类型和车辆故障 4 个因素，均属于多分类变量，处理方法与车辆类型相同；而天气状况、事故时间段、平面线形、事故地点为二分类变量，编码为 0 和 1；坡度和坡长为连续性变量，采用实际值。

所有的自变量需要通过检验筛选之后才可进入模型，一般分为建模前检验和建模后检验。建模

前检验称为比分检验(Score Test), 以未包含某一个或几个参数的模型为基础, 保留模型中参数的估计值, 并假设新增加的参数为 0, 计算似然函数的一阶偏导及信息矩阵, 两者相乘即为比分检验统计量 S . 当样本量较大时, S 也服从卡方分布. 该检验是在建模前根据变量之间的结构关系判断自变量与因变量之间的密切程度, 对自变量进行筛选. 当 Score 值达到显著性水平, 表明该自变量与因变量有显著性的关系. 若 Score 值相对应的 p 值小于给定的显著性水平 α , 则该自变量进入模型; 否则, 该自变量将被剔除.

另一种变量筛选方法为建模后检验法, 称为 Wald 检验. Wald 检验的原假设 β_j 为零, Wald 统计量的数学表达式为

$$Wald_i = \left(\frac{\hat{\beta}_j}{S_{\hat{\beta}_j}} \right) \quad (2)$$

式中: $\hat{\beta}_j$ 为参数估计值; $S_{\hat{\beta}_j}$ 为参数估计值的标准误差.

Wald 统计量在原假设得到满足的条件服从一个自由度为 1 的卡方分布. 如果 Wald 值对应的显著性检验值大于给定的显著性水平 α , 则认为因变量与该自变量的关系不显著, 该自变量予以删除; 否则, 该自变量进入模型.

1.3 模型构建

以交通事故严重程度作为因变量, 将交通事故分为两类: 伤亡事故和非伤亡事故. 伤亡事故表示在事故统计有效期内有人员伤亡的事故, 即当第 i 起山区连续长下坡路段事故有人员伤亡, $y_i=1$; 非伤亡事故表示在事故统计有效期内无人员伤亡的事故, 即当第 i 起山区连续长下坡路段事故无人员伤亡, $y_i=0$.

假设因变量为 y_i , n 个自变量分别为 x_1 , x_2 , L , x_n , 依据 Logistic 原理, 发生伤亡事故的概率为:

$$p = p(y_i=1 | x_1, x_2, L, x_n) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)} \quad (3)$$

其中, β_0 为常数项; $\beta_1 \sim \beta_n$ 为回归系数.

2 Logistic 回归模型检验

Logistic 模型构建后, 需要对模型进行检验和评价, 分为对回归方程整体的显著性检验、回归系数的显著性检验和模型拟合度检验. 在自变量筛

选过程中, 已经采用回归系数的显著性检验, 因此, 下面不再对回归系数的检验进行赘述.

2.1 回归方程整体的显著性检验

二元 Logistic 回归方程的显著性检验用于检验因变量的对数与所有变量之间的显性线性关系是否显著, 检验的原假设是各回归系数同时为零, 备择假设为各回归系数不同时为零. 常用的检验方法有对数似然比检验和 Hosmer-Lemeshow 检验. 其中, 对数似然比检验假设没有引入任何解释变量的回归方程的似然函数值为 L_0 , 引入解释变量之后回归方程的似然函数为 L_1 , 则似然比为 L_0/L_1 . 构造似然比卡方统计量, 借助卡方分布进行显著性检验. 当卡方值达到给定的显著性水平, 表明回归方程整体显著.

2.2 模型拟合度检验

模型的拟合优度检验可用 Person 拟合度检验、Deviance 拟合度检验、似然比检验和 Hosmer-Lemeshow 检验. 其中, Hosmer-Lemeshow 检验是根据模型预测概率的大小将所有观察值十等分, 然后根据每一组中因变量各种取值的实测值与理论值计算 Person 卡方 χ^2_{HL} ^[15]. 在原假设成立的条件下, χ^2_{HL} 渐近服从 $m-2$ 的 χ^2 分布(m 为组数, m 且 ≤ 10), χ^2_{HL} 当检验不显著表明模型拟合效果较好.

3 实例分析

本文以陕西省内两条山区连续长下坡路段三年内发生的 169 起交通事故作为建模数据.

3.1 自变量选择

考虑到自变量之间可能存在一定的相关性, 采用因子分析对自变量之间进行分析. 由于数据中天气状况、路表情况均只有一个变量, 所以对天气状况和路表情况未进行因子分析.

从表 3、表 4 可以看出, 各自变量之间的相关系数较小, 说明各自变量之间相关性不显著.

使用 SPSS22.0 软件进行二元 Logistic 回归, 采用正向逐步选择法, 选取显著水平 $\alpha=0.05$, 将上述 11 个因素代入模型, 分别选择正面碰撞、小汽车、违法违规、制动失效作为参照变量. 先利用比分检验法筛选自变量, 依据 Score 值对应的 p 值是否满足给定的显著性水平, 选择出满足要求的变量, 检验结果见表 5. 鉴于建模数据中天气状况均为晴天, 路表情形均为干燥, 未涉及其他情况, 所以天气状况和路表情形不会作为影响事故严重

程度的因素。

按照比分检验结果,选择显著性 p 值小于 0.05 的变量进入模型,事故地点、车辆故障、坡度和坡长 4 个变量满足要求。

通过迭代计算后,按照 Wald 值检验要求,获得模型最终参数结果见表 6。

根据表 6,说明坡度和事故地点与事故严重程度有较强的显著关系。

表 3 相关性结果

Tab. 3 Correlation results

	事故时间段	平面线形	车辆类型	碰撞类型	过失违章类型	车辆故障	坡度	事故地点	坡长
事故时间段	1.000	-0.005	0.183	0.159	0.058	-0.116	0.009	-0.022	0.088
平面线形	-0.005	1.000	-0.030	0.014	-0.150	-0.004	0.043	0.038	0.042
车辆类型	0.183	-0.030	1.000	0.043	0.080	-0.220	-0.124	-0.034	-0.031
碰撞类型	0.159	0.014	0.043	1.000	0.016	-0.054	0.038	0.067	-0.107
过失违章类型	0.058	-0.150	0.080	0.016	1.000	-0.280	-0.113	-0.086	-0.228
车辆故障	-0.116	-0.004	-0.220	-0.054	-0.280	1.000	0.027	-0.029	0.148
坡度	0.009	0.043	-0.124	0.038	-0.113	0.027	1.000	0.048	0.233
事故地点	-0.022	0.038	-0.034	0.067	-0.086	-0.029	0.048	1.000	-0.297
坡长	0.088	0.042	-0.031	-0.107	-0.228	0.148	0.233	-0.297	1.000

表 4 显著性结果

Tab. 4 Significant results

	事故时间段	平面线形	车辆类型	碰撞类型	过失违章类型	车辆故障	坡度	事故地点	坡长
事故时间段		0.478	0.012	0.025	0.239	0.077	0.455	0.392	0.142
平面线形	0.478		0.357	0.432	0.032	0.478	0.297	0.323	0.306
车辆类型	0.012	0.357		0.299	0.163	0.003	0.063	0.337	0.352
碰撞类型	0.025	0.432	0.299		0.421	0.256	0.323	0.205	0.094
过失违章类型	0.239	0.032	0.163	0.421		0.000	0.084	0.147	0.002
车辆故障	0.077	0.478	0.003	0.256	0.000		0.369	0.363	0.034
坡度	0.455	0.297	0.063	0.323	0.084	0.369		0.277	0.002
事故地点	0.392	0.323	0.337	0.205	0.147	0.363	0.277		0.000
坡长	0.142	0.306	0.352	0.094	0.002	0.034	0.002	0.000	

表 5 Score 检验结果

Tab. 5 Score Test Results

影响因素	Score 值	自由度	p 值	影响因素	Score 值	自由度	p 值
事故时间段	1.082	1	0.298	过失违章类型 1	1.274	1	0.259
平面线形	1.847	1	0.174	过失违章类型 2	0.230	1	0.632
碰撞类型	2.372	3	0.499	过失违章类型 3	0.237	1	0.626
碰撞类型 1	0.162	1	0.687	过失违章类型 4	0.048	1	0.827
碰撞类型 2	1.454	1	0.228	事故地点	10.661	1	0.001
碰撞类型 3	2.285	1	0.131	车辆故障	7.805	2	0.020
车辆类型	3.954	2	0.139	车辆故障 1	2.267	1	0.132
车辆类型 1	2.407	1	0.121	车辆故障 2	4.670	1	0.031
车辆类型 2	3.839	1	0.050	坡度	15.478	1	0.004
过失违章类型	1.912	4	0.752	坡长	4.640	1	0.031

表6 参数估计

Tab. 6 Parameters Estimation

影响因素	B	S.E	Wald	自由度	显著性	Exp(B)	Exp(B)的95%置信区间	
							上限	下限
坡度	0.468	0.151	9.596	1	0.002	1.596	1.187	2.145
事故地点	-1.128	0.362	9.689	1	0.002	0.324	0.159	0.659
常量	3.084	0.827	13.909	1	0.000	21.847		

3.2 模型检验

对模型整体进行检验,得到似然比卡方值为21.360,自由度为2,选取显著性水平为0.05,查卡方临界值表可知: $\chi^2(2)=5.991<26.177$, $p\leqslant 0.01$,说明回归方程整体通过检验。

对模型进行Hosmer和Lemeshow检验,得到卡方值为10.013,自由度为7,显著性水平取0.05,查卡方临界值表知: $\chi^2(7)=14.067>9.648$, $p=0.209>0.05$,表明模型拟合度较好。

3.3 结果分析

根据上述模型计算结果,坡度和事故地点对事故严重程度具有显著性影响。由式(3)可得,发生伤亡事故的概率为

$$p(y=1|x) = \frac{\exp(3.084 + 0.468x_{10} - 1.128x_{11})}{1 + \exp(3.084 + 0.468x_{10} - 1.128x_{11})} \quad (4)$$

其中: x_{10} 为坡度, x_{11} 为事故地点。

由式(4)可知,坡度的回归系数为0.561,说明坡度越大,发生伤亡事故的概率越大;而事故地点的回归系数为负值,表明特殊路段发生伤亡的概率较非特殊路段大,这和事故统计结果隧道、河流、桥梁、峡谷等特殊路段均属于事故多发地点相吻合,这些路段之所以成为事故多发路段主要是因为不利于驾驶员行车安全的因素较多,增加了驾驶员的操作难度,易酿成重大伤亡事故。

4 结语

(1)以山区连续长下坡路段为研究对象,事故的严重程度为因变量,从人、车、路和环境因素4个方面选择11个候选自变量,采用正向逐步选择法分析自变量与因变量是否存在显著相关关系,研究发现,坡度和事故地点是影响山区连续长下坡路段事故严重程度的主要因素。

(2)针对事故的严重程度为分类变量及某些自变量与因变量并非为确定的线性关系,依据Logistic回归原理,构建了山区连续长下坡路段交通事故严重程度度量模型。

(3)本文的研究样本局限为陕西省内的两条山区连续长下坡路段,且没有分析天气状况和路表情形对事故严重程度的影响,后续可增加对其他地区的山区连续长下坡路段的交通事故分析和建模。

参考文献 References

- [1] 张娟. 基于驾驶员因素的公路连续长坡路段线形研究[D]. 西安: 长安大学, 2012.
ZHANG Juan. Continuous long slope road alignment research based on driver's factor[D]. Xi'an: Chang'an University, 2012.
- [2] 袁伟, 付锐, 郭应时, 等. 考虑坡长因素的纵坡坡度对交通事故的影响分析[J]. 公路交通科技, 2008, 25(5): 130-135.
YUAN Wei, FU Rui, GUO Yingshi, et al. Influences of longitudinal gradient on traffic accident rate considering length of downgrade[J]. Journal of Highway and Transportation Research and Development, 2008, 25(5): 130-135.
- [3] 张小东, 高建平, 孔令旗. 高速公路连续长下坡路段行车安全分析[J]. 山东交通科技, 2005,(1):17-19.
ZHANG Xiaodong, GAO Jianping, KONG Lingqi. Safety analysis of continuous long downgrade section in expressway while driving[J]. Shandong Jiaotong Keji, 2005,(1):17-19.
- [4] 陈斌, 袁伟, 付锐, 等. 连续长大下坡路段交通事故特征分析[J]. 交通运输工程学报, 2009, 9(4): 75-84.
CHEN Bin, YUAN Wei, FU Rui, et al. Analysis of traffic accident characteristic on continuous long downgrade section[J]. Journal of Traffic and Transportation Engineering, 2009, 9(4): 75-84.
- [5] 袁伟, 郭应时, 付锐, 等. 考虑坡长因素的纵坡坡度对交通事故的影响分析[J]. 公路交通科技, 2008, 5(5): 130-135.
YUAN Wei, GUO Yingshi, FU Rui, et al. Influence of longitudinal gradient on traffic accident rate consid-

- ering length of downgrade[J]. Journal of Highway and Transportation Research and Development, 2009, 5 (5):130-135.
- [6] 付锐,郭应时,袁伟,等. 连续下坡道路事故率与纵断面参数关系研究[J]. 中国公路学报, 2009, 5 (3): 101-106.
FU Rui, GUO Yingshi, YUAN Wei, et al. Research on relation of traffic accident rate and longitudinal parameters in continuous downgrade road [J]. China Journal of Highway and Transport, 2009, 5 (3): 101-106.
- [7] SZE N N, WONG S C. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes[J]. Accident Analysis and Prevention, 2007, 39(6):1267-1278.
- [8] 马壮林,邵春福,李霞. 基于 Logistic 模型的公路隧道交通事故严重程度的影响因素[J]. 吉林大学学报(工学版), 2010, 3(2):423-426.
MA Zhuanglin, SHAO Chunfu, LI Xia. Analysis of factors affecting accident severity in highway tunnels based on logistic model[J]. Journal of Jilin University (Engineering and Technology Edition), 2010, 3 (2): 423-426.
- [9] 冯忠祥,雷叶维,张卫华,等. 道路环境对绕城高速公路交通事故严重程度影响分析[J]. 中国公路学报, 2016, 5(5):116-123.
FENG Zhongxiang, LEI Yewei, ZHANG Weihua, et al. Analysis on traffic accident severity influenced by road environment on circular highway[J]. China Journal of Highway and Transport, 2016, 5(5):116-123.
- [10] 赵金宝,邓卫,王建. 基于贝叶斯网络的城市道路交通事故分析[J]. 东南大学学报(自然科学版), 2011, 11 (6):1300-1306.
ZHAO Jinbao, DENG Wei, WANG Jian. Bayesian network-based urban road traffic accidents analysis [J]. Journal of Southeast University(Natural Science Edition), 2010, 11(6):1300-1306.
- [11] 张丽霞,刘涛,潘福全,等. 驾驶员因素对道路交通事故指标的影响分析[J]. 中国安全科学学报, 2014, 5 (5):79-84.
ZHANG Lixia, LIU Tao, PAN Fuquan, et al. Analysis of effect of driver factors on road traffic accident indexes[J]. China Safety Science Journal, 2014, 5 (5): 79-84.
- [12] 孔令铮. 交通事故致因中的人为因素分析[J]. 中国安全科学学报, 2013, 1(1):28-34.
KONG Lingzheng. Human factors in causation of traffic accidents[J]. China Safety Science Journal, 2013, 1 (1):28-34.
- [13] 张长生,马荣国. 山区高速公路交通事故分析及多发路段鉴别[J]. 长安大学学报(自然科学版)2010, 11 (6):76-80.
ZHANG Changsheng, Ma Rongguo. Identifying method of traffic accident black spots on mountain freeway[J]. Journal of Chang'an University (Natural Science Edition), 2010, 11(6): 76-80.
- [14] 潘娅英,陈武. 引发公路交通事故的气象条件分析[J]. 气象科技, 2006, 12(6):778-782.
PAN Yaying, CHEN Wu. Analysis of meteorological conditions for traffic accidents[J]. Meteorological Science and Technology, 2006, 12(6):778-782.
- [15] 张文彤,董伟. 统计分析高级教程,第二版[M]. 北京: 高等教育出版社,2013;173.
ZHANG Wentong, DONG Wei. Advanced course of statistical analysis (2nd ed)[M]. Beijing: Higher Education Press, 2013;173.

(编辑 吴海西)