

# 基于 Tobit 回归的山区高速公路事故率分析模型

孟祥海<sup>1</sup>, 刘振博<sup>1,2</sup>

(1. 哈尔滨工业大学 交通科学与工程学院, 黑龙江 哈尔滨 150090; 2. 四川省国土空间规划研究院)

**摘要:**以事故率为因变量来建立事故分析模型或预测模型时,由于事故率是连续型随机变量,传统的泊松分布、负二项分布及零堆积类分布模型已不再适用。加之,当统计年限较短或路段划分较短时,统计样本中的事故率会出现较多“0”值的情况,此时,事故率就属于受限的连续型变量,即受限因变量。通过对比分析可适用于受限因变量建模的 Truncated 回归和 Tobit 回归方法,基于 Tobit 回归建立了山区高速公路事故率分析模型并计算了模型中自变量的边际效应。模型的因变量为路段上的亿车公里事故率,自变量为路段上的年平均日交通量、路段长度及几何线形指标变量。研究结果表明:针对事故率这样一个受限因变量,采用 Tobit 回归来建立事故率分析模型是适宜的,所建立的事故率分析模型较好地反映了年平均日交通量、路段长度、几何线形条件等对事故率的影响。模型标定结果及边际效应数值表明:纵坡类型对山区高速公路事故率的影响最大;其次是纵坡坡度、路段长度和年平均日交通量;最后是竖曲线曲率。若设竖曲线曲率对事故率的影响程度为 1.0,则年平均日交通量、路段长度、纵坡坡度、纵坡类型的影响程度分别为 3.1、4.1、9.1、和 9.7。

**关键词:** 交通工程; 山区高速公路; 事故率; Tobit 回归; 分析模型

## 1 前言

在传统的交通事故分析模型或预测模型中,使用的因变量多为事故次数、伤亡事故次数或事故伤亡人数等离散型的随机变量,并采用泊松分布、负二项分布、零堆积泊松分布或零堆积负二项分布等对离散型事故数据进行拟合。马壮林等针对交通事故多发路段,建立了事故次数的泊松回归模型、负二项回归模型及零堆积回归模型;孟祥海等建立了线形与交通状态组合条件下的追尾事故次数负二项分布模型;依据负二项分布标定了高速公路基本路段事故次数预测模型,提出了基于负二项分布的路段安全性评价方法;基于统计及假设检验,深入分析了事故次数、伤亡事故数、事故死亡人数等离散型事故数据的统计分布特征。

然而在实际应用中,由于交通事故数据的隐私性和敏感性,往往很难直接获得事故次数、事故伤亡人数等数据,尤其是发生在具体道路上的上述数据。相反,在许多交通事故分析报告、交通安全评价报告甚至是交通安全研究报告中,亿车公里事故率、百万辆车事故

率等事故率指标经常被用来作为评价指标,用以描述道路交通安全状况。也就是说,在有些没有事故次数、伤亡人数等事故绝对指标的情况下,事故率即成为可获取的事故指标。因此,在此背景下,基于事故率来建立事故分析模型或预测模型就具有了一定的现实意义。另外,事故率本身就包含了交通量、路段长度等事故关联因素信息,是非常有效和客观的交通安全评价指标。

从统计学原理上看,事故次数作为因变量,只能取非负整数,属于离散型随机变量,因此,采用泊松回归或负二项回归来建立事故分析模型是合理的。当统计期短或路段划分较短时,事故次数统计数据中可能会出现大量的“0”值,此时,还可考虑采用零堆积泊松回归或零堆积负二项回归来建立事故分析模型。不同于事故次数,事故率属于连续型的随机变量,泊松、负二项及零堆积类回归方法已不再适用,需要有新的回归方法。同理,当统计期短或路段划分较短时,事故率统计数据中仍可能出现大量的“0”值,此时,事故率取值是受限的,属于受限因变量。

一般而言,连续型受限因变量的回归方法常采用

收稿日期:2019-10-23(修改稿)

基金项目:国家自然科学基金资助项目(编号:7170010718);广东省交通运输厅科技项目(编号:2012-01-001-02)

作者简介:孟祥海,男,博士,教授,博士生导师。E-mail:mengxianghai100@126.com

Truncated 回归(断尾回归)和 Tobit 回归(截取回归)两种。其中,断尾回归适用于统计数据能取得全体样本,而只有在大于或小于某个常数才能被观测到的情况;Tobit 回归适用于当统计数据大于等于或小于等于某个常数时,所有的数据均被记录为这个常数的情况。Tobit 回归虽然可以取得全部的观测数据,但对于某些观测数据,因变量取值被压缩到一个点上了,此时,因变量的概率分布就变成了由一个离散点和一个连续分布所组成的混合分布。因此,当事故率属于受限因变量时,采用 Tobit 回归来建立事故分析模型是适合的。

鉴于中国尚未将 Tobit 模型应用于道路交通安全领域的实际情况,该文尝试应用 Tobit 回归来建立山区高速公路事故率与几何线形条件之间的关系模型,旨在验证 Tobit 模型的适用性问题,并据此分析几何线形条件对交通事故的影响。山区高速公路复杂的几何线形条件,是诱发交通事故的重要原因之一,从事事故率角度进一步分析这种影响关系也是十分有意义的。

## 2 事故率分析模型建模及求解方法

### 2.1 事故率分析模型形式

事故率分析模型是一个描述事故率指标与事故影响因素之间相关关系的多元线性回归模型,即:

$$Y_i = \beta X_i + \epsilon_i, i=1,2,\dots,N \quad (1)$$

式中: $Y_i$  为路段  $i$  上的事故率; $X_i$  为路段  $i$  上的事故影响因素; $\beta$  为待估计的参数变量; $\epsilon_i$  为残差项; $N$  为路段数(即样本数)。

### 2.2 回归方法选择

针对式(1)的多元线性回归模型,由于事故率是连续型受限因变量,因此,可采用断尾回归和 Tobit 回归两种方法来建立事故率分析模型。

对于断尾回归,设样本总体  $Y$  原来的概率密度为  $f(Y)$ ,则断尾后的概率密度  $f^*(Y)$  为:

$$f^*(Y) = \begin{cases} \frac{f(Y)}{P(Y>0)} & Y>0 \\ 0 & Y\leq 0 \end{cases} \quad (2)$$

式中: $P(Y>0)$  为事故率  $Y>0$  的概率。

当事故率为“0”时,断尾回归认为断尾处概率密度为“0”,即在断尾回归中事故率为“0”的情况不会出现,显然,这与事实是不符的。

对于 Tobit 回归,受限因变量的概率密度分布被变换成由一个离散点和一个连续分布所组成的混合分

布。Tobit 回归认为,在左侧受限处的概率密度不为“0”,即事故率为“0”的情况可能会出现,这更符合事故率数据的实际情况。事实上,此时应用 Tobit 回归,可看做求解事故率的最优解问题:在交通事故不发生的情况下,事故率的最优解即为边角解(Corner solution),即事故率为“0”;在交通事故发生的情况下,则事故率一定为正数。

应用 Tobit 回归建立事故率分析模型可表达为:

$$Y_i^* = \beta X_i + \epsilon_i, i=1,2,\dots,N \quad (3)$$

$$Y_i = \begin{cases} Y_i^* & Y_i^* > 0 \\ 0 & Y_i^* \leq 0 \end{cases} \quad (4)$$

式中: $Y_i^*$  为不可观测的潜在变量; $\epsilon_i$  仍为残差项,Tobit 回归中假定残差项服从均值为 0、标准差为  $\sigma$  ( $\sigma > 0$ ) 的正态分布,即  $\epsilon_i : N(0, \sigma^2)$ 。

### 2.3 事故率分析模型参数求解方法

应用 Tobit 回归时,如果用最小二乘法(OLS)进行估计,非线性项将被纳入残差项中,无论使用整体样本还是去掉离散点后的子样本,都不能得到一致的估计。

若采用去掉离散点后的子样本,即使用“ $Y_i > 0$ ”的子样本进行 OLS 参数估计时,可求得事故率均值  $E(Y_i)$  为:

$$E(Y_i | X_i; Y_i > 0) = \beta X_i + \sigma \cdot \lambda(-\beta X_i / \sigma) \quad (5)$$

式中: $\lambda$  为逆米尔斯比率(IMR);对任意常数  $\alpha$ ,  $\lambda(\alpha) = \varphi(\alpha) / [1 - \Phi(\alpha)]$ ,其中  $\Phi(\alpha)$  为标准正态分布的分布函数, $\varphi(\alpha)$  为标准正态分布的概率密度函数。

在使用子样本进行回归时,由于忽略了非线性项  $\sigma \cdot \lambda(-\beta X_i / \sigma)$ ,该项被纳入到了残差项中,从而导致残差项与  $X_i$  相关,OLS 无法得到一致的估计。

若使用整体样本进行 OLS 参数估计时,可求得事故率均值  $E(Y_i)$  为:

$$E(Y_i | X_i) = E(Y_i | X_i; Y_i > 0) \cdot P(Y_i > 0 | X_i) = \Phi(\beta X_i / \sigma) [\beta X_i + \sigma \cdot \lambda(-\beta X_i / \sigma)] \quad (6)$$

因此,使用整体样本进行 OLS 参数估计时,也将非线性项纳入到残差项中,从而不能得到一致的估计。

为了解决这一问题,Tobin 于 1958 年提出使用最大似然估计法(MLE)。在使用 MLE 进行参数估计时,由于残差项  $\epsilon_i$  服从正态分布,即  $\epsilon_i : N(0, \sigma^2)$ ,则  $Y_i^* : N(\beta X_i, \sigma^2)$ ,由此可得:

$$P(Y_i = 0) = P(Y_i^* \leq 0) = P\left(\frac{Y_i^* - \beta X_i}{\sigma} \leq \frac{-\beta X_i}{\sigma}\right) = \Phi(-\beta X_i / \sigma) = 1 - \Phi(\beta X_i / \sigma) \quad (7)$$

$$P(Y_i | Y_i > 0) = P(Y_i^*) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{Y_i - \beta X_i}{2\sigma^2}} = \frac{1}{\sigma} \varphi \cdot \left( \frac{Y_i^* - \beta X_i}{\sigma} \right) \quad (8)$$

此时,式(4)中  $Y_i$  的概率密度函数为:

$$f(Y_i | X_i) = [1 - \Phi(\beta X_i / \sigma)]^{1(Y_i=0)} \left[ \frac{1}{\sigma} \varphi \cdot \left( \frac{Y_i^* - \beta X_i}{\sigma} \right) \right]^{1(Y_i>0)} \quad (9)$$

其中,  $1(\cdot)$  为示性函数,即如果括号里的表达为真,取值为 1;反之,取值为 0。得出样本整体的似然函数为:

$$L = \prod_0 [1 - \Phi(\beta X_i / \sigma)] \prod_1 \left[ \frac{1}{\sigma} \varphi \left( \frac{Y_i^* - \beta X_i}{\sigma} \right) \right] \quad (10)$$

$$\ln L = \sum_{Y_i>0} -\frac{1}{2} \left[ \ln 2\pi + 2 \ln \sigma + \left( \frac{Y_i^* - \beta X_i}{\sigma} \right)^2 \right] + \sum_{Y_i=0} \ln [1 - \Phi(\beta X_i / \sigma)] \quad (11)$$

对式(11)求令其最大化的解,即可得到模型参数估计结果。由于这是一个复杂的非线性问题,可考虑采用迭代法进行求解,例如采用牛顿迭代法等求解。

### 3 数据来源与数据处理

#### 3.1 数据来源

依托广东省交通运输厅科技项目“基于全社会成本的高速公路设计方案评价技术研究”(2012-01-001-02)科研课题,收集到了京珠高速公路粤北段 2006 年 1 月至 2009 年 6 月发生的 1 557 起事故数据、粤赣高速公路 2007 年 1 月至 2012 年 7 月发生的 1 678 起事故数据以及上述两条高速公路的几何线形数据。京珠高速公路粤北段为山岭重丘区高速公路,设计速度 100 km/h(局部路段 80 km/h),双向四车道,全长 109.29 km。粤赣高速公路为重丘区高速公路,双向四车道,设计速度 100 km/h,全长 136.1 km。

#### 3.2 路段划分

由于该文建立的事故率分析模型主要是基于公路几何线形条件,因此,应基于几何线形条件对上述两条高速公路进行路段划分。平面线形划分为直线路段和平曲线路段两类。竖曲线的划分是以变坡点为界,划分为竖曲线前半段和后半段,若前半段或后半段的切线纵坡为上坡,则该半段竖曲线归类为竖曲线上坡段,

反之归类为竖曲线下坡段。将平纵线形进行组合,可得到以下 8 种路段:直线—上坡路段、直线—下坡路段、直线—竖曲线上坡路段、直线—竖曲线下坡路段、平曲线—上坡路段、平曲线—下坡路段、平曲线—竖曲线上坡路段、平曲线—竖曲线下坡路段,编号分别为路段类型 1~8。路段划分结果如表 1 所示,京珠高速公路粤北段共划分出 1 082 个路段,粤赣高速公路共划分出 1 242 个路段,这些路段就是建立事故率分析模型的样本。

表 1 路段划分结果

路段类型	路段数/个	
	京珠高速公路粤北段	粤赣高速公路
1	110	114
2	110	114
3	154	163
4	154	163
5	89	105
6	89	105
7	188	239
8	188	239
合计	1 082	1 242

#### 3.3 事故率计算与异常值处理

对划分后的路段,进行事故率指标计算。事故率选用亿车公里事故率,计算公式如下:

$$R_j = \left[ \sum_{i=1}^n A_{ij} / \left( \sum_{i=1}^n AADT_{ij} \cdot L_j \times 365 \right) \right] \times 10^8 \quad (12)$$

式中:  $R_j$  为路段  $j$  上的事故率[次/(亿车·km)];  $A_{ij}$  为路段  $j$  上第  $i$  年的事故次数(次);  $AAADT_{ij}$  为路段  $j$  上第  $i$  年的年平均日交通量(pcu/d);  $L_j$  为路段  $j$  的长度(km);  $n$  为统计年限,对于不足一整年的数据,须折算。

对因路段划分长短不同所导致的亿车公里事故率异常值(主要是因路段过短而出现的极高的事故率值),以及过长或过短的路段,应进行处理。为此,对路段长度及亿车公里事故率进行缩尾处理,剔除上下各 1% 的路段长度极端值和 1% 高的事故率极端值。最终得到了 2 239 个有效路段(表 2)。其中,京珠高速公路粤北段有效路段 1 033 个,粤赣高速公路 1 206 个。

由表 2 可知:事故率为“0”的路段数所占比例较大,达到了路段总数的 47.5%。这表明:亿车公里事故率是受限因变量(含“0”值较多的连续型变量),因

此,采用受限因变量的回归方法来建立事故率分析模型是合理的。

表2 有效路段及事故率

高速公路	有效路段数/个	事故率为“0”的有效路段数/个	事故率大于“0”的有效路段数/个	事故率为“0”的有效路段占比/%
京珠高速公路粤北段	1 033	482	551	46.7
粤赣高速公路	1 206	582	624	48.3
总体	2 239	1 064	1 175	47.5

### 3.4 数据分析

剔除异常值后,有效路段上的事故率及其线形数

表3 事故率及道路线形数据统计结果

指标	亿车公里事故率/ [次·(亿车·km) <sup>-1</sup> ]	年平均日交通量/(×10 <sup>3</sup> pcu)	路段长度/m	平曲线半径/km	竖曲线半径/km	纵坡坡度绝对值/%
样本量(2 239个)	2 239	2 239	1 198	1 471	2 239	
平均值	34.26	13.25	207.29	2.97	23.65	1.88
标准差	53.61	2.99	125.74	2.27	24.20	1.19
最小值	0	7.46	8.00	0.50	7.00	3.00

表4 不同类型路段事故率数据统计

路段类型	样本量/个	事故率/[次·(亿车·km) <sup>-1</sup> ]			
		平均值	标准差	最小值	最大值
1	208	25.71	38.80	0	246.420
2	213	45.72	62.03	0	350.185
3	310	27.39	41.06	0	222.440
4	310	44.83	71.08	0	374.040
5	177	25.25	42.79	0	284.230
6	171	33.94	52.11	0	339.837
7	426	29.40	46.43	0	369.960
8	424	38.76	57.48	0	347.480
总体	2 239	34.26	53.61	0	374.040

其中,纵坡类型变量取“0”或“1”值,属于虚拟变量,用以区分上下坡路段,上坡路段取“0”,下坡路段取“1”。选择平曲线曲率和竖曲线曲率的原因是,可方便变量的赋值。路段类型1~4均为平面直线与纵断面线形组合路段,此时平曲线半径为无限大,而平曲线曲率则可取为“0”值。同理,在纵坡与平面线形组合的路段中,纵坡路段的竖曲线半径为无限大,而竖曲线曲率

据、交通量数据统计结果见表3。不同类型路段上的事故率统计结果见表4。由表4可知:下坡路段事故率整体高于上坡路段,事故率平均值较大的是直线一下坡路段和直线一竖曲线下坡段,事故率平均值较小的是平曲线一上坡路段和直线一上坡路段。

## 4 事故率分析模型自变量及其赋值

山区高速公路事故率分析模型中选取的因变量为亿车公里事故率( $R$ );选取的自变量为年平均日交通量( $AADT$ )、路段长度( $L$ )、平曲线曲率( $D_H$ )、竖曲线曲率( $D_V$ )、纵坡类型( $ID$ )和纵坡坡度( $i$ ),单位分别为pcu/d、m、km<sup>-1</sup>、km<sup>-1</sup>、无量纲和%。

则可取为“0”。由于在路段划分中,上、下坡路段已分属不同的路段类型,因此,纵坡坡度均取绝对值。线形指标原始数据与赋值后的数据示例,见表5。

## 5 事故率分析模型参数标定与分析

### 5.1 模型参数标定

为横向比较不同自变量对事故率的影响程度,需要消除各自变量特征尺度的潜在影响,因此,需要对自变量数据进行归一化处理。归一化公式为:

$$M_i^* = \frac{M_i - M_{i\min}}{M_{i\max} - M_{i\min}}, i = 1, 2, \dots, n \quad (13)$$

式中, $M_i^*$ 为第*i*类变量归一化后的取值; $M_i$ 为第*i*类变量原始值; $M_{i\min}$ 为第*i*类变量原始值中的最小值; $M_{i\max}$ 为第*i*类变量原始值中的最大值; $n$ 为变量种类数(此处为6)。

模型的因变量,即亿车公里事故率,也应进行归一化处理,公式同式(13)。

对归一化后的事故率及事故影响因素数据,应用Stata统计分析软件进行参数标定,结果见表6。

表 5 线形指标原始数据与赋值后数据对比示例

样本 编号	路段 类型	原始数据				重新赋值后				
		路段长度	平曲线半	竖曲线半	纵坡坡	路段长	平曲线曲率	竖曲线曲率	纵坡类	纵坡坡
		<i>L</i> /m	径 <i>R<sub>H</sub></i> /km	径 <i>R<sub>V</sub></i> /km	度 <i>i</i> /%	度 <i>L</i> /m	<i>D<sub>H</sub></i> /(km <sup>-1</sup> )	<i>D<sub>V</sub></i> /(km <sup>-1</sup> )	型 <i>ID</i>	度 <i>i</i> /%
1	1	226	—	—	1.5	226	0	0	0	1.5
2	2	218	—	—	-1.0	218	0	0	1	1.0
3	3	218	—	7.0	2.8	218	0	0.14	0	2.8
4	4	247	—	20.0	-1.0	247	0	0.05	1	1.0
5	5	117	8.0	—	0.4	117	0.13	0	0	0.4
6	6	329	0.9	—	-2.8	329	1.11	0	1	2.8
7	7	275	8.0	20.3	0.4	275	0.13	0.05	0	0.4
8	8	260	0.9	14.6	-2.4	260	1.11	0.07	1	2.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2 239	2	117	—	—	-1.4	117	0	0	1	1.4

注：“—”为原始数据空值项。

表 6 模型参数标定结果

自变量	系数	标准误差	<i>t</i>	<i>P</i> >  <i>t</i>	95%置信水平
AADT	-0.212	0.023	-9.05	0.000	(-0.258, -0.166)
<i>L</i>	0.282	0.037	7.59	0.000	(0.209, 0.354)
<i>D<sub>H</sub></i>	0.005	0.027	0.02	0.986	(-0.052, 0.053)
<i>D<sub>V</sub></i>	0.070	0.021	3.29	0.001	(0.028, 0.112)
<i>ID</i>	0.067	0.010	6.47	0.000	(0.047, 0.088)
<i>i</i>	0.063	0.021	2.97	0.003	(0.021, 0.105)
常数项 <i>a</i>	-0.038	0.020	-1.86	0.063	(-0.078, 0.020)

事实上,由于基于 Tobit 回归的山区高速公路事故率分析模型属于广义线性模型,模型的回归系数并不能直接代表自变量对因变量的影响程度大小,因此还需对模型自变量求其各自的边际效应。边际效应的计算结果如表 7 所示。

表 7 自变量边际效应计算结果

自变量	$\partial R/\partial x$	标准误差	<i>z</i>	<i>P</i> >  <i>t</i>	95%置信水平
AADT	-0.107	0.012	-9.02	0.000	(-0.130, -0.083)
<i>L</i>	0.142	0.018	7.70	0.000	(0.106, 0.178)
<i>D<sub>H</sub></i>	0.000 2	0.013	0.02	0.986	(-0.026, 0.027)
<i>D<sub>V</sub></i>	0.035	0.011	3.29	0.001	(0.014, 0.056)
<i>ID</i>	0.341	0.005	6.47	0.000	(0.024, 0.044)
<i>i</i>	0.319	0.011	2.97	0.003	(0.011, 0.053)

### 5.2 模型结果分析

由表 6 可知:在标定模型的 2 239 组数据中,非受限数据 1 175 组,以“0”值作为左侧受限界限的左侧受限数据 1 064 组。除平曲线曲率显著性较差外,年平均日交通量、路段长度、竖曲线曲率、纵坡类型、纵坡坡度绝对值在 95%置信水平下均是显著的,概率值 *P* 均小于 0.05。

由表 7 可知:年平均日交通量的边际效应为负,说明随着年平均日交通量的增加,亿车公里事故率逐渐减小。这也隐含着说明:事故次数随交通量的增长不是线性增加的,即事故次数的增长率要低于交通量的增长率。路段长度的边际效应为正,说明随着路段长度的增加,亿车公里事故率在增加。平曲线曲率对事故率的影响不明显,边际效应极小。竖曲线曲率的边际效应为正,说明随着竖曲线曲率的增大,即竖曲线半径的减小,亿车公里事故率在增大,该回归结果与竖曲线半径越大交通安全状况越好的相关研究结论是相符的。纵坡类型及纵坡绝对值的边际效应均为正,说明下坡路段亿车公里事故率要高于上坡路段,且随着纵坡坡度绝对值的增大,亿车公里事故率也随之增加。这一回归结果也证实了下坡路段危险性高于上坡路段、坡度越大危险性越大的相关研究结论。

对事故率影响最大的因素是纵坡坡度,最小的是竖曲线曲率。记竖曲线曲率对亿车公里事故率的影响程度为 1.0,则年平均日交通量、路段长度、纵坡坡度、纵坡类型对亿车公里事故率的影响程度分别为 3.1、4.1、9.1、9.7。

## 6 结论及有待进一步研究的问题

### 6.1 结论

(1) 由于事故率是连续型随机变量,加之会出现大量“0”值的情况,因此,采用Tobit回归来建立事故率分析模型是适宜的。

(2) 基于Tobit回归建立的山区高速公路事故率分析模型,较好地反映出了年平均日交通量、几何线形条件等对事故率的影响。模型标定结果及边际效应数值表明,纵坡类型对事故率的影响最大;其次依次是纵坡坡度、路段长度、年平均日交通量和竖曲线曲率。

### 6.2 有待进一步研究的问题

(1) 该文所建立的事故率分析模型,仅考虑了交通量、路段长度以及几何线形条件等对交通事故的影响,在后续的研究中,还应进一步考虑路面条件、天气条件以及交通状态等对交通事故的影响,从而提高分析模型的精度。

(2) 该文仅考虑了交通安全中“量”的问题,即事故率。在后续的研究中,还应考虑事故严重程度指标,如伤亡事故率等,这是交通安全中“质”的问题。

### 参考文献:

- [1] 马壮林,邵春福,胡大伟,等.高速公路交通事故起数时空分析模型[J].交通运输工程学报,2012(2).
- [2] 孟祥海,张晓明,郑来.基于线形与交通状态的山区高速公路追尾事故预测[J].中国公路学报,2012(4).
- [3] 孟祥海,李昕,郑来.基于事故数据与安全服务水平的高速公路路段安全性评价方法[J].交通运输工程学报,2014(6).
- [4] 孟祥海,覃薇.基于统计及假设检验的高速公路事故数据分布特性研究[J].交通运输工程学报,2018(1).
- [5] Amemiya T. Regression Analysis when the Dependent Variable is Truncated Normal[J]. Econometrica, 1973, 41(6):997-1016.
- [6] Tobin J. Estimation of Relationships for Limited Dependent Variables[J]. Econometrica, 1958, 26(1):24-36.

- [7] Pch A, Tarko A P, Mannering F L. Tobit Analysis of Vehicle Accident Rates on Interstate Highways[J]. Accid Anal Prev, 2008, 40(2):768-775.
- [8] Pch A, Mannering F L, Shankar V N, et al. A Study of Factors Affecting Highway Accident Rates Using the Random-Parameters Tobit Model[J]. Accident Analysis and Prevention, 2012, 45(2):628.
- [9] Anastasopoulos P C, Shankar V N, Haddock J E, et al. A Multivariate Tobit Analysis of Highway Accident-Injury-Severity Rates[J]. Accident Analysis & Prevention, 2012, 45(3):110-119.
- [10] Anastasopoulos P C. Random Parameters Multivariate Tobit and Zero-Inflated Count Data Models: Addressing Unobserved and Zero-State Heterogeneity in Accident Injury-Severity Rate and Frequency Analysis[J]. Analytic Methods in Accident Research, 2016, 11:17-32.
- [11] Zeng Q, Wen H, Huang H, et al. A Bayesian Spatial Random Parameters Tobit Model for Analyzing Crash Rates on Roadway Segments[J]. Accident Analysis & Prevention, 2017, 100:37-43.
- [12] Zeng Q, Wen H, Huang H, et al. A Multivariate Random-Parameters Tobit Model for Analyzing Highway Crash Rates by Injury Severity[J]. Accident Analysis & Prevention, 2017, 99(Part A):184-191.
- [13] Chen F, Ma X, Chen S. Refined-Scale Panel Data Crash Rate Analysis Using Random-Effectstobit Model[J]. Accident Analysis & Prevention, 2014, 73:323.
- [14] 孟祥海,关志强,郑来.基于几何线形指标的山区高速公路安全性评价[J].中国公路学报,2011(2).
- [15] 孟祥海,荆林朋,侯芹忠.高速公路平纵线形与事故率的关系及其安全性评价[J].交通信息与安全,2015(4).
- [16] 孟祥海,侯芹忠,史永义,等. IHSDM高速公路事故预测模型[J].交通运输工程学报,2016(1).
- [17] 付锐,郭应时,袁伟,等.连续下坡道路事故率与纵面参数关系研究[J].中国公路学报,2009(3).
- [18] 邓晓庆,孟祥海,郑来.基于BP神经网络的高速公路事故预测模型[J].交通信息与安全,2016(1).