

# 删失数据下事件持续时间多因素生存分析模型

蒋宏, 方守恩, 陈雨人

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

**摘要:** 交通事件持续时间是事故管理研究中最重要指数之一, 应用半参数比例风险模型和参数加速失效模型分析了多因素事件持续时间及其危险因子之间的关系. 利用浙江省某高速公路3年内采集的事件数据, 通过对交通事件各状态概率分布函数的显著性分析, 比较了比例风险模型和加速失效模型对协变量选择及生存率对共同协变量的敏感性. 基于Cox比例风险模型和对数罗吉斯蒂加速失效模型的参数估计表明, 有6个显著性的协变量入选各自估计的生存函数, 包括报警时段、事故类型、报警方式、占用车道数、受伤人数和当场亡人. 对于最显著示性变量(当场亡人), 持续时间生存概率估计曲线表明, 加速失效模型对于该变量更敏感, 且Cox比例风险模型更适合于短时持续时间的预测. 此外, 生存模型可以基于事故报告信息预测持续时间大小的概率, 有益于事故预后措施的实施及为紧急救援提供决策参考.

**关键词:** 交通事件; 持续时间; 生存分析; 比例风险模型; 加速失效模型

**中图分类号:** U491.3

**文献标识码:** A

## Multivariate Survival Analysis Models for Incident Duration With Censored Data

JIANG Hong, FANG Shouen, CHEN Yuren

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

**Abstract:** Traffic incident duration is one of the most important indexes in accident management studies. The semi-parametric proportional hazards model (Cox model) and parametric accelerated failure time model (AFT model) were employed for analysis of the association between multivariate incident durations and risk factors. With the three-year traffic incidents data collected from some freeways in Zhejiang Province, the Cox model and AFT model with covariates selected to remain in the model and survival probability of sensitivity to common covariates were illustrated and compared. Based on the Cox and the Log-logistic AFT

models, the parameter estimates showed that six significant covariates were selected to remain in the estimated survival functions, including accident time, accident type, number of lanes occupied, number of the injured and accident of fatality. With regard to the most significant indicator variable (accident of fatality), the estimates of the survival probability curves for duration demonstrate that the AFT model give a higher sensitivity than the Cox model, and the Cox model is more appropriate for short time incident duration prediction. Moreover, survival analysis models can predict duration based on incident report and benefit rescue countermeasure and emergency aid decision after accidents.

**Key words:** traffic incident; duration; survival analysis; proportional hazards model; accelerated failure time model

交通事件的持续时间对于道路管理人员实施紧急疏散策略、制定交通管理措施和交通流诱导等具有十分重要的意义. 交通事件持续时间则受到天气、事故类型、到场时间、占用车道数、涉及车辆数、伤亡人数、救援车辆数等随机因素的交互影响.

交通事件持续时间预测方法主要有: 基于统计的预测方法、回归模型、时间序列模型、决策树、非参数回归法和模糊逻辑法等<sup>[1]</sup>. 其中, 决策树模型是比较简单有效的方法, 例如姬杨蓓蓓等<sup>[2]</sup>采用基于贝叶斯决策树的算法, 利用数据建立交通事件持续时间的预测模型, 且具有很好的鲁棒性; 刘伟铭等<sup>[3]</sup>在对事件数据进行显著性分析后, 建立了高速公路交通事件持续时间预测决策树. 而目前国外较多采用概率模型对事件持续时间特性进行研究. 多数研究表明交通事件持续时间服从对数正态分布<sup>[4-5]</sup>, 而丛浩哲等<sup>[6]</sup>对浙江省某条高速公路交通事件持续时间数据检验表明数据不服从正态分布; Doohee等<sup>[7]</sup>的研究表明事故的检测时间、报告时间和响应时间服

收稿日期: 2011-10-13

基金项目: 国家“八六三”高技术研究发展计划(2007AA112233)

第一作者: 蒋宏(1981—), 男, 工程师, 博士生, 主要研究方向为道路交通安全、交通规划设计. E-mail: yamatuo@hotmail.com

通讯作者: 方守恩(1961—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为道路交通安全、道路规划与计算机辅助设计等.

E-mail: fangsek@tongji.edu.cn

从 Weibull 模型,清除时间则服从 Log-logistic 模型.

本文分别应用生存分析模型中的比例风险模型和加速风险模型对交通事件持续时间及影响因素(协变量)进行建模,并讨论 2 个模型的适用性.

### 1 交通事件持续时间的特性

交通事件持续时间一般包括 4 个独立的阶段:事件发现阶段、事件响应阶段、事件清除阶段和交通恢复阶段<sup>[8]</sup>.事件发现阶段:从交通事件发生到交通管理者、警察或高速公路救援队发现交通事件的时间;事件响应阶段:从交通事件被确认到救援车辆到达现场的时间;事件清除阶段:救援队伍处理受伤人员、封闭车道直到移除车辆和碎片后离开现场即救援队伍将干扰交通运行的障碍清除所处的时段;事件恢复阶段:交通事件被清除后车辆排队开始消散直到交通流恢复到正常交通状态所处的时段.

目前国内对于交通事件持续时间的准确预测尚有一定的难度,其中比较重要的原因是缺少对事件状态全过程的记录.从公路交管部门对交通事件的记录信息来看,可用于交通持续时间统计分析的只有事件确认时刻、开展救援时刻以及处置完毕时刻.由于记录的缺失,导致事件持续时间数据出现 2 种类型(删失和完全),其中  $t_1$  和  $t_2$  为删失数据, $t_3$  为完全数据,如图 1.

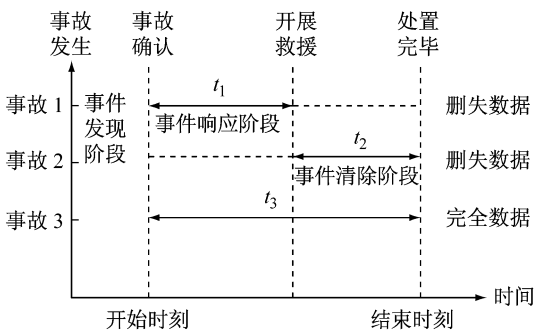


图 1 事件持续时间数据示例

Fig.1 Illustration of incident duration data

由于交通恢复阶段受交通流和交通管理等多种因素影响难以观测和记录,交通事件的发现时间相对其他阶段可以忽略不计,本文将事件确认到处置完毕的时间段作为研究的观测周期.

### 2 生存分析的基本函数

生存分析是研究多种影响因素与生存时间有无

联系以及联系程度大小的一种既考虑时间又考虑结果的统计方法,并可充分利用删失数据提供的不完全信息<sup>[9]</sup>.生存时间(事件状态的持续)指从某起点事件开始到被观测对象出现终点事件所经历的时间.

生存时间  $T$  是一个连续性的非负随机变量,当取  $t$  时具有累计分布函数  $F(t)$ ,又称为失效函数(failure function).由定义可知,事件持续时间的概率是随机变量  $T$  的函数,表示为

$$F(t) = P(T < t) = \int_0^t f(u)du, \quad 0 < t < \infty \tag{1}$$

式中: $P(T < t)$ 表示事件  $\{T < t\}$  发生的概率.由式(1)可得概率密度函数为

$$f(t) = \frac{\partial F(t)}{\partial t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \tag{2}$$

生存函数  $S(t)$  (survival function) 又称累计生存率,简称生存率,表示具有协变量  $X$  的观察对象其生存时间  $T$  大于时间  $t$  的概率,定义为

$$S(t) = P(T > t) = 1 - F(t), \quad t > 0 \tag{3}$$

还有一个在  $t$  时刻处(附近)对死亡发生的可能性进行度量的函数  $h(t)$ ,称为危险函数(hazard function),它是一  $\ln S(t)$  关于  $t$  的导数,定义如下:

$$h(t) = \frac{f(t)}{S(t)} = \frac{d}{dt} [-\ln S(t)] \tag{4}$$

累计危险函数为

$$H(t) = \int_0^t h(u)du \tag{5}$$

事件持续时间的危险函数、密度函数、累计频率函数和生存函数的关系如图 2.

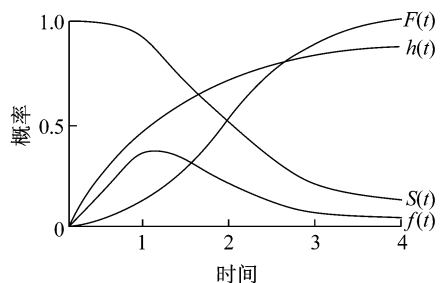


图 2 生存分析模型函数<sup>[10]</sup>

Fig.2 Survival analysis models functions<sup>[10]</sup>

### 3 比例风险模型

生存分析模型中最常用的多因素分析方法就是比例风险模型(proportional hazards model)<sup>[11]</sup>,该

模型是一种半参数模型,不需要假定生存时间的分布,但却可以通过一个模型来分析生存时间的分布规律以及危险因素(协变量)对生存时间的影响。其基本形式如下:

$$h(t | Z) = h_0(t) \exp\left(\sum_{k=1}^p \beta_k Z_k\right) \quad (6)$$

式中: $h(t | Z)$ 为具有  $p$  个协变量  $Z$  的个体  $j$  在时刻  $t$  的危险率; $h_0(t)$ 为基准风险函数,它是全部协变量都为零或标准状态下的风险函数,一般是未知的; $Z_k$  ( $k=1, \dots, p$ )为个体  $j$  不随时间改变的协变量; $\beta_k$  ( $k=1, \dots, p$ )为变量参数,若  $\beta_k > 0$ ,表明该协变量为危险因子,会增加风险函数值,对生存时间出现负作用, $\beta_k < 0$  表明该协变量为保护因子,会减少风险函数值,即延长生存时间, $\beta_k = 0$  表明该因素为无关因素。

之所以被称为比例风险模型是因为任意 2 个具有协变量  $Z$  和  $Z^*$  的个体,其危险率成比例

$$\frac{h(t | Z)}{h(t | Z^*)} = \exp\left[\sum_{k=1}^p \beta_k (Z_k - Z_k^*)\right] \quad (7)$$

协变量参数向量  $\beta = (\beta_1, \dots, \beta_k, \dots, \beta_p)'$  可根据偏似然估计方法得到。令  $t_1 < t_2 < \dots < t_D$  表示顺序事件时间,定义时间  $t_i$  时的风险集  $R(t_i)$  为在  $t_i$  之前仍处于研究过程的所有观测个体集合, $Z_{jk}$  为个体  $j$  的第  $k$  个协变量, $Z_{(i)k}$  是与在时间  $t_i$  时失效的个体相关的第  $k$  个协变量,定义危险函数的偏似然函数如下:

$$L(\beta) = \prod_{i=1}^D \frac{\exp\left[\sum_{k=1}^p \beta_k Z_{(i)k}\right]}{\sum_{j \in R(t_i)} \exp\left[\sum_{k=1}^p \beta_k Z_{jk}\right]} \quad (8)$$

为了评价比例风险模型的拟合程度,选用 Cox-Snell 残差对其进行检验。如果模型中  $\beta$  的估计值为  $b = (b_1, \dots, b_k, \dots, b_p)'$ ,则 Cox 残差定义为

$$r_j = \hat{H}_0(t_j) \exp\left(\sum_{k=1}^p Z_{jk} b_k\right), \quad j = 1, \dots, n \quad (9)$$

式中: $\hat{H}_0(t_j)$ 为第  $j$  个个体研究时间  $T_j$  下的基准累计危险率估计值,如果模型选用正确,那么  $r_j$  的累计风险估计  $\hat{H}_r(r_j)$  与  $r_j$  的平面图应是通过原点的  $45^\circ$  直线。

## 4 加速失效模型

参数加速失效模型 (accelerated failure time models) 可作为半参数比例模型的替代模型<sup>[12]</sup>,其对

事件发生时间  $X$  的自然对数  $Y = \ln(X)$  建模,参数模型的形式如下:

$$Y = \mu + \gamma'Z + \sigma W \quad (10)$$

式中: $\mu$  为截距; $\gamma' = (\gamma_1, \gamma_2, \dots, \gamma_p)$  为回归系数向量; $Z$  为协变量矩阵; $\sigma$  为未知的尺度参数; $W$  为误差项。通常根据假设误差项服从的不同分布来产生不同的回归模型<sup>[13]</sup>,其对应关系如表 1。

表 1 常用的参数加速失效模型

Tab. 1 Some commonly used accelerated failure time models for parametric duration models

W	Y
单参数极值分布	指数分布 (exponential)
双参数极值分布	威布尔分布 (Weibull)
对数分布	对数罗吉斯蒂分布 (Log-logistic)
正态分布	对数正态分布 (Log-normal)
对数-伽玛分布	伽玛分布 (Gamma)

如果令  $S_0(x)$  为随机向量  $\exp(\mu + \sigma W)$  的生存函数,对于所有的观测时刻  $x$ ,带有协变量  $Z$  的危险率与基本危险率  $h_0$  的关系可变换为

$$h(x | Z) = h_0[x \exp(-\gamma'Z)] \exp(-\gamma'Z) \quad (11)$$

模型中的参数可以通过极大似然估计法得到。假定有  $n$  个样本,定义  $f_j(t_j)$  和  $S_j(t_j)$  为观测个体  $j$  在时刻  $t_j$  的密度函数和生存函数,其示性函数为  $\delta_j$ ,构造似然函数为

$$L(\gamma, \mu, \sigma) = \prod_{j=1}^n \{f_j(t_j)\}^{\delta_j} \{S_j(t_j)\}^{1-\delta_j} \quad (12)$$

模型拟合程度评价采用概率图检验的方法;此外,在一定置信水平下的拟合模型并不具有唯一性,本文以赤池信息量准则<sup>[14]</sup> (Akaike information criterion, AIC) 判断相对优劣性来选择更优的模型, AIC 准则要求 AIC 的取值越小越好。模型 AIC 函数 ( $A_{IC}$ ) 的定义如下:

$$A_{IC} = -2l + 2(e + c) \quad (13)$$

式中: $l$  为对数似然函数; $e$  为模型中协变量的个数; $c$  为模型独立参数个数。

## 5 模型应用

### 5.1 数据准备

实例分析对象为浙江省某条双向 8 车道高速公路,全长约 78 km,设计车速  $120 \text{ km} \cdot \text{h}^{-1}$ 。当地交警部门采集的交通事件数据信息包括:星期、天气、报警时间、报警类型、到达现场时间、事件类型、占用车道数、涉及车辆数、受伤人数、死亡人数、最先到达现场的车辆、救援车辆数等。交通事件数据采集的时间跨度为 3 年,共得事件观测个体 1 455 个。对采集的

定量和定性数据进行编码和赋值:如将报警时间分为早高峰、白天、晚高峰和夜间 4 个水平;又如由散落物引起的事故涉及车辆数为 0;观测个体是否删失分别用 0 和 1 标识. 模型变量的分组和赋值如表 2.

表 2 变量说明及赋值一览

Tab.2 Summary of variables explanation and value assignment

变量	因素	赋值说明
Z <sub>1</sub>	星期及节假日	星期一至星期天,其他法定节假日:0~7
Z <sub>2</sub>	天气	晴天:0,阴天:1,雨天:2,雪天:3,雾天:4 早高峰(07:01~09:00):0, 白天(09:01~16:30):1, 晚高峰(16:31~20:00):2, 夜间(20:01~07:00):3
Z <sub>3</sub>	报警时段	驾驶员:0,交警部门:1, 高速管理:2,路政及其他:3
Z <sub>4</sub>	报警方式	追尾:0,翻车:1,撞中间护栏:2,抛锚:3,刮擦:4,炸胎:5,撞路侧护栏:6,散落物:7,碰撞:8,着火:9,其他:10
Z <sub>5</sub>	事故类型	0~5 条,5 条为单向路段全封闭(含紧急停车带)
Z <sub>6</sub>	占用车道数	单位:辆
Z <sub>7</sub>	涉及车辆数	单位:人
Z <sub>8</sub>	受伤人数	是:1,否:0
Z <sub>9</sub>	当场亡人	单位:min
T	持续时间	完全数据:0,删失数据:1
Y	删失变量	

### 5.2 模型拟合与检验

采用逐步回归法对变量进行筛选及最大似然估计,比例风险模型中通过显著性检验的协变量按照对持续时间影响程度大小排列为 Z<sub>9</sub>,Z<sub>6</sub>,Z<sub>3</sub>,Z<sub>8</sub> 和 Z<sub>5</sub> 这 5 个变量. 此外,通过变量方差分析发现,Z<sub>1</sub>,Z<sub>2</sub>,Z<sub>4</sub> 和 Z<sub>7</sub> 这 4 个协变量对于生存率没有显著影响,这些变量被剔除在最终的回归模型外;从比例分析模型协变量参数估计系数的符号来看,回归方程系数除 Z<sub>5</sub> 以外全部小于零,说明报警时段和占用车道数等全部为保护因子,即协变量每增加 1 个等级交通事件的持续时间就会延长,危险度降低. 对于加速风险模型,通过显著性检验的协变量按照对持续时间影响程度大小排列为 Z<sub>9</sub>,Z<sub>6</sub>,Z<sub>3</sub>,Z<sub>8</sub>,Z<sub>5</sub>,Z<sub>4</sub> 这 6 个变量. 此外,通过变量方差分析发现,Z<sub>1</sub>,Z<sub>2</sub>,Z<sub>7</sub> 这 3 个协变量对于生存率没有显著影响,这些变量被剔除在最终的回归模型外. 2 个模型协变量最终的筛选及参数估计结果如表 3.

虽然有些协变量被剔除在模型外,但并不能排除这些因素对持续时间的影 响,预选影响因子之间存在的强相关性会导致该现象的产生. 例如,当显著性相关变量中有一方进入回归模型中,另一方变量

表 3 模型的参数估计

Tab.3 Estimated parameters for the models

模型	协变量	因素	自由度	参数估计	标准差	Wald 卡方	P 值
比例风险模型	Z <sub>3</sub>	报警时段	1	-0.232	0.027	73.695	<0.000 1
	Z <sub>5</sub>	事故类型	1	0.027	0.011	6.300	0.012 0
	Z <sub>6</sub>	占用车道数	1	-0.435	0.043	104.862	<0.000 1
	Z <sub>8</sub>	受伤人数	1	-0.115	0.030	14.787	<0.000 1
	Z <sub>9</sub>	当场亡人	1	-0.704	0.194	13.239	<0.000 1
加速失效模型(Log-logistic)	μ	截距	1	3.110	0.083	1 417.200	<0.000 1
	Z <sub>3</sub>	报警时段	1	0.156	0.018	71.930	<0.000 1
	Z <sub>4</sub>	报警方式	1	-0.044	0.019	5.520	0.018 8
	Z <sub>5</sub>	事故类型	1	-0.055	0.010	32.260	<0.000 1
	Z <sub>6</sub>	占用车道数	1	0.323	0.027	146.760	<0.000 1
	Z <sub>8</sub>	受伤人数	1	0.085	0.020	18.300	<0.000 1
	Z <sub>9</sub>	当场亡人	1	0.640	0.124	26.750	<0.000 1
	σ	尺度	1	0.381	0.009		

则有可能会排除在模型外,如当事故类型进入模型后,Z<sub>7</sub> 被排除在模型外,但并不能就此说明 Z<sub>7</sub> 对交通事件的持续时间没有影响,通过对协变量进行皮尔逊相关性检验后发现,2 个变量间的皮尔逊相关系数为-0.550 8,说明进入模型的事故类型这个协变量部分反映了涉及车辆数的信息.

根据 5 个协变量拟合的比例风险模型绘制残差和残差累计危险率的 Cox-Snell 残差图(图 3),图中的直线大致呈 45°,说明模型拟合程度较好.

比例风险模型生存函数模型形式为

$$S(t | Z) = S_0(t) \exp(-0.232Z_3 + 0.027Z_5 - 0.435Z_6 - 0.115Z_8 - 0.704Z_9) \tag{14}$$

式中,S<sub>0</sub>(t)为基准生存函数. 对数据分别用各种假设回归模型进行拟合,得到对数罗吉斯蒂模型的 A<sub>IC</sub> 值最小(如表 4),因此最终采用对数罗吉斯蒂进行建模,加速失效模型其生存函数形式为

$$S_T(t | Z) = [1 + t^{\frac{1}{\sigma}} \exp(\hat{\eta}_i)]^{-1} = [1 + t^{\frac{1}{0.381}} \exp(\hat{\eta}_i)]^{-1} \tag{15}$$

$$\hat{n}_i = \frac{-\mu - \gamma/Z}{\sigma} = -(3.110 + 0.156Z_3 - 0.044Z_4 - 0.055Z_5 + 0.323Z_6 + 0.085Z_8 + 0.640Z_9)/0.381 \quad (16)$$

模型拟合程度的概率图检验如图 4.

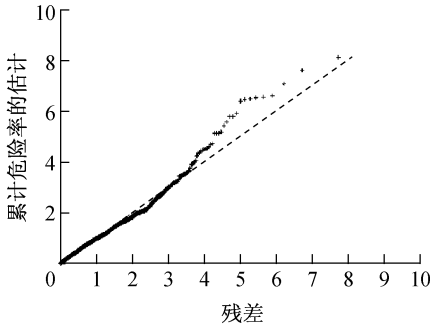


图 3 Cox-Snell 残差

Fig.3 Cumulative hazard of Cox-Snell residuals

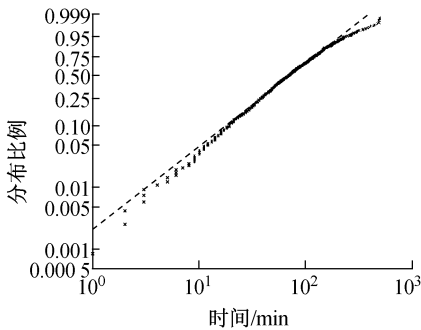


图 4 对数罗吉斯蒂模型概率图

Fig.4 Probability plot for Log-logistic model

为了比较 2 个模型对协变量的敏感程度,假定如下情景:在控制协变量  $Z_3=1, Z_4=1, Z_5=1, Z_6=1$  和  $Z_8=1$  的情况下,比较对模型影响程度最大的协变量  $Z_9$  对生存函数的影响.从图 6a 可见对于是否发生亡人事故,持续时间大于 25 min 的概率差别甚微,且生存时间的概率估计值均可达到 80% 以上;但随着持续时间的增加,发生亡人事故对于事件结束的概率影响较大;从图 6b 可见,对于亡人事故,概率持续时间大于 50 min 的概率约为 40%,而对于非亡人事故概率估计值约为 10%.

为了更进一步分析协变量对生存概率的影响,借鉴方差分析的思想比较 2 个模型对协变量当场亡人( $Z_9=0, 1$ )的敏感程度,考察任意观测时间下的生存函数概率绝对差  $|S_T(t|Z_9=1) - S_T(t|Z_9=0)|$  的变化情况,生存函数概率绝对差曲线如图 7.

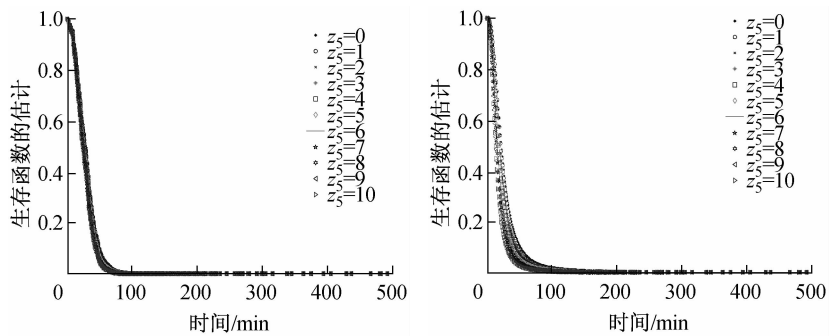
表 4 加速失效模型的参数

Tab.4 Analysis for parameters of accelerated failure time models

模型	对数似然估计值	e	c	A <sub>IC</sub>
指数分布	-1 627.44	6	1	3 268.88
威布尔分布	-1 497.79	6	2	3 011.58
对数罗吉斯蒂分布	-1 348.89	6	2	2 713.78
对数正态分布	-1 381.82	6	2	2 779.64
伽玛分布	-1 380.04	6	3	2 778.08

### 5.3 协变量敏感性分析

生存函数每种协变量组合都会生成不同的生存函数概率曲线.图 5 给出了不同事故类型下协变量  $Z_5$  的生存函数概率,可以看出,事故类型对于生存函数的概率影响较小,没有显著性不同,这是由协变量的参数估计值决定的,对于比例风险模型和加速失效模型其系数分别为 0.027 和 0.055.由此可见,对于生存概率影响的大小是由参数估计值的贡献决定的,因此,本文选择了贡献率最大的因子  $Z_9$  进行深入分析.



a 比例风险模型

b 加速失效模型

图 5 事故类型下的生存函数

Fig.5 Survivorship function with accident types

变量敏感性对比分析表明:在观测区间内 2 个模型对亡人事故的整体敏感性接近.但从图 7 可见:比例风险模型敏感性曲线的曲率变化率较小,而加速失效模型的曲率变化率较大,表明该模型对于协变量当场亡人的敏感性较强,在持续时间约为 30 min 时达到极值;此外,持续时间约为 60 min 的绝对差成为 2 个模型对协变量当场亡人敏感性的分水岭.

## 6 结论

在分析交通事件状态及持续时间规律的基础上,以某高速公路的 1 455 起交通事件持续时间及影响因素为分析对象,应用生存分析理论相关模型对交通事件持续时间分别用比例风险模型和加速失效

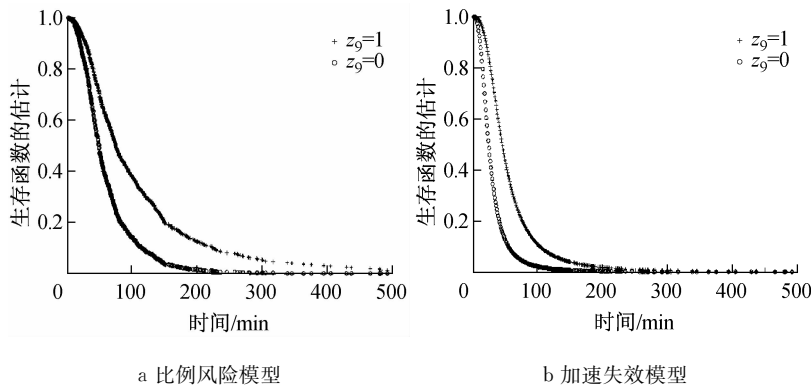


图 6 当场亡人下的生存函数

Fig.6 Survivorship function with accident of fatality

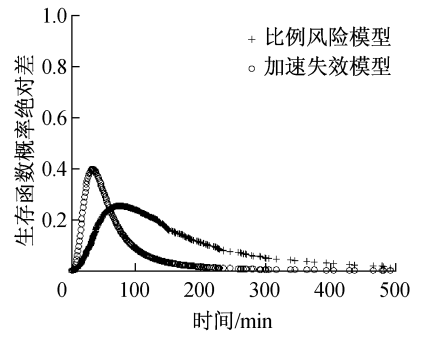


图 7 生存函数概率绝对差

Fig.7 Survival function probability absolute of error curve

模型对协变量进行了筛选并建立了相应的生存函数模型;并对持续时间的生存率贡献最大的协变量当场亡人进行了敏感性分析,比较了协变量在不同水平下的生存率,在持续时间小于 30 min 时加速失效模型对于当场亡人事故的敏感性较强,不利于短时的建模预测,且对于短时预测采用比例风险模型更符合实际操作.生存模型可以基于事故报告信息预测持续时间大小的概率,为事故预后措施的实施及紧急救援提供决策参考.

此外,交通事件持续时间的预测评价技术还需要从以下几个方面深入:①完善事件持续时间 4 个阶段的统计数据并建立相关的数据库,通过对事件持续时间分布规律的研究提高预测模型的精度;②预测模型的参数估计值的时间稳定性<sup>[15]</sup>也需要大量的数据来验证;③本文仅利用半参数加速失效模型和全参数加速失效模型对持续时间的生存函数进行了建模和比较分析,其他生存分析模型对于交通事件持续时间的适用性有待于进一步研究.

参考文献:

[1] 姬杨蓓蓓,张小宁,孙立军. 交通事件持续时间预测方法综述[J]. 公路, 2008, 33(3): 72.  
JIYANG Beibei, ZHANG Xiaoning, SUN Lijun. A review of the traffic incident duration prediction methods [J]. Highway Engineering, 2008, 33(3): 72.

[2] 姬杨蓓蓓,张小宁,孙立军. 基于贝叶斯决策树的交通事件持续时间预测[J]. 同济大学学报: 自然科学版, 2008, 36(3): 319.  
JIYANG Beibei, ZHANG Xiaoning, SUN Lijun. Traffic incident duration prediction grounded on Bayesian decision method-based tree algorithm[J]. Journal of Tongji University: Natural Science, 2008, 36(3): 319.

[3] 刘伟铭,管丽萍,尹湘源. 基于决策树的高速公路时间持续时间预测[J]. 中国公路学报, 2005, 18(1): 99.

LIU Weiming, GUAN Liping, YIN Xiangyuan. Prediction of freeway incident duration based on decision tree[J]. China Journal of Highway and Transport, 2005, 18(1): 99.

[4] Golob T F, Reeker W W, Leonard J D. An analysis of the severity and incident duration of truck-involved freeway accidents[J]. Accident Analysis and Prevention, 1987, 19(4): 375.

[5] Garib A, Radwan A E, Al-Deek H. Estimating magnitude and duration of incident delays [J]. Journal of Transportation Engineering, 1997, 123(6): 459.

[6] 丛浩哲,方守恩,王俊骅. 交通事件持续时间影响因素分析及其回归模型[J]. 交通信息与安全, 2010, 28(3): 80.  
CONG Haozhe, FANG Shouen, WANG Junhua. Factors analysis of freeway incident duration and regression modeling [J]. Computer and Communication, 2010, 28(3): 80.

[7] Doohee Nam, Fred Mannering. An exploratory hazard-based analysis of highway incident duration [J]. Transportation Research Part A, 2000, 34:85.

[8] 王建军,邓亚娟. 路网环境下高速公路交通事故影响传播分析与控制[M]. 北京:科学出版社, 2010.  
WANG Jianjun, DENG Yajuan. Traffic accident impact analysis and control of expressway under road network[M]. Beijing: Science Press, 2010.

[9] 彭非,王伟. 生存分析[M]. 北京:中国人民大学出版社, 2004.  
PENG Fei, WANG Wei. Survival analysis[M]. Beijing: China Renmin University Press Co. Ltd., 2004.

[10] Washington S P, Karlaftis M G, Mannering F L. Statistical and econometric methods for transportation data analysis[M]. Boca Raton: CRC Press LLC, 2003.

[11] Cox D R. Regression models and life tables[J]. Journal of Royal Statistic Society, 1972, 34(B): 187.

[12] Wei L J. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis[J]. Statistics in Medicine, 1992, 11: 1871.

[13] Allison P D. Survival analysis using SAS: a practical guide [M]. 2nd ed. Cary: SAS Institute Inc, 2010.

[14] Akaike H. A new look at the statistical model identification[J]. IEEE Transactions on Automatic Control, 1974, 19(6): 716.

[15] Chung Y. Development of an accident duration prediction model on the korean freeway systems [J]. Accident Analysis and Prevention, 2010, 42: 282.