

■ 論 文 ■

AFT 생존분석 기법을 이용한 고속도로 교통사고 지속시간 예측모형

A Prediction Model on Freeway Accident Duration using AFT Survival Analysis

정 연 식

(한국교통연구원 책임연구원)

송 상 규

(한국도로공사 SmartWay사업단)

최 기 주

(아주대학교 환경건설교통공학부 교수)

목 차

- I. 서론
 - II. 생존 분석
 - 1. 개요
 - 2. 생존 분석 기법
 - III. 사고 자료
 - 1. 개요
 - 2. 변수 설정
 - IV. 교통사고 지속시간 예측 모형 개발
 - 1. 모수적 방법을 이용한 예측 모형 개발
 - 2. 모형의 선정
 - 3. 모형의 해석
 - V. 결론
- 참고문헌

Key Words : 사고 지속시간, 생존분석, AFT 모형, 분할 다항식, AIC(Akaike's Information Criterion) Accident Duration, Survival Analysis, AFT Model, Fractional Polynomials, AIC(Akaike'S Information Criterion)

요 약

교통사고의 특성과 사고에 대한 지속시간 사이의 관계에 대한 이해는 사고의 효과적인 대응과 사고에 의한 혼잡을 감소시키는데 핵심 요소가 된다. 때문에 본 연구의 목적은 AFT metric 모형을 적용한 사고 지속시간을 분석하는 것이다. 비록 로그 로지스틱 및 로그 정규 AFT 모형이 통계적 이론과 기존 연구 사례를 기반으로 선정되었으나, 로그 로지스틱 모형이 보다 우수하게 추정되었다. AFT 모형은 예측 목적으로도 널리 사용되기 때문에, 추정된 모형은 사고 발생시 사고 관련 기본 정보 접수 즉시 고속도에서의 사고 지속시간 예측에 사용될 수 있다. 결과적으로, 예측된 사고 지속시간 정보는 사고를 처리하기 위한 제반 의사 결정에 도움을 줄 뿐 아니라 교통 혼잡의 감소 및 추가 사상자의 감소로 그 효과가 이어질 것으로 판단된다.

Understanding the relation between characteristics of an accident and its duration is crucial for the efficient response of accidents and the reduction of total delay caused by accidents. Thus the objective of this study is to model accident duration using an AFT metric model. Although the log-logistic and log-normal AFT models were selected based on the previous studies and statistical theory, the log-logistic model was better fitted. Since the AFT model is commonly used for the purpose of prediction, the estimated model can be also used for the prediction of duration on freeways as soon as the base accident information is reported. Therefore, the predicted information will be directly useful to make some decisions regarding the resources needed to clear accident and dispatch crews as well as will lead to less traffic congestion and much saving the injured.

1. 서론

고속도로에서의 혼잡 현상은 다양한 원인들에 의해 발생된다. 이러한 혼잡 현상은 두 가지 주요 원인으로 구분할 수 있으며, 통행량이 도로의 용량을 초과한 경우와 도로에 교통의 통행을 방해하는 사건이 발생한 경우이다. 특히, 두 번째 원인을 돌발 상황(incident)이라 하며, 교통사고, 작업, 차량의 고장, 갑작스런 기상 변화 등이 여기에 해당된다. 이러한 돌발 상황에 대한 효과적인 고속도로 운영관리를 위해 돌발상황관리 시스템(incident management system)과 이에 적용될 다양한 교통 모형들이 개발되어 왔다. 특히, 돌발 상황의 지속시간(duration) 예측 및 돌발상황 지속시간에 영향을 주는 인자들의 특성 파악은 돌발 상황 발생 후 적절한 대응을 위해 매우 중요한 요소들이다. 이로 인해 다양한 연구가 이 분야에 시도되어 왔으며, 이러한 모형들은 다음과 크게 3가지로 분류할 수 있다.

- ① Descriptive statistics: 이 기법은 자료 수집 기술이나 관리 기법이 발달되지 못했던 초창기에 주로 사용되었으며, CCTV (De Rose Jr., 1964), 경찰 조사자료 (Goolsby et al., 1971), time laps camera (Juge et al., 1974) 등을 이용하여 수집된 돌발상황의 지속시간에 대한 기술 통계량 (즉, 평균 및 분산)에 의존하였다.
- ② Analytical model: 돌발 상황 자료 수집 기술과 관리 기법이 발달함에 따라 다양한 종류의 자료가 수집되었다. 즉, 다변량 (multivariate) 자료를 분석하기 위해 다양한 분석적 모델이 적용되었다 (Golob et al., 1987; Giuliano, 1989; Jones et al., 1991; Khatkhatk et al., 1995; Garib et al., 1997; Nam and Mannering, 2000; Stathopoulos and Karlaftis, 2002).
- ③ Heuristic method: 이 방식은 다양한 유형의 유고상황에 대하여 지속시간을 예측하기 위하여, 지식기반 전문가 시스템을 적용하였다. (Ozbay and Kachroo, 1999; Smith and Smith, 2001).

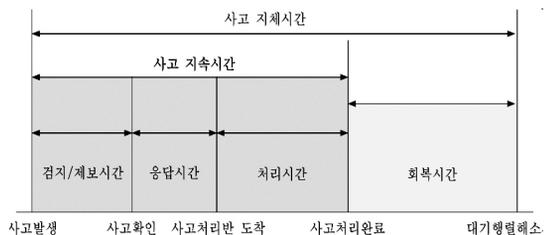
열거된 대부분의 모형들은 돌발 상황 자료를 기반으로 구축되었다. 그러나, 우리나라 고속도로의 경우 돌발 상황의 원인은 대부분 작업과 교통사고에 의한 것으로 나타났으며, 혼잡을 발생시키는 돌발 상황의 대부분은

교통사고인 것으로 나타났다 (백승걸·박재범, 2004). 따라서, 전체 돌발 상황 중에서 교통사고에 대한 지속시간의 연구는 더욱 중요할 것으로 기대된다.

교통사고에 대한 지속시간은 돌발상황에 대한 지속시간의 정의(Garib et al., 1997; Nam and Mannering, 2000; Smith and Smith, 2001)와 마찬가지로, 사고 발생 시점과 사고처리 완료 시점 간의 시차로 정의 할 수 있다. 또한, Highway Capacity Manual (TRB, 1994)에서 정의한 것과 같이, 사고는 4가지 진행요소로 구분할 수 있다: (1) 사고 검지 및 제보 시간, (2) 대응시간, (3) 처리시간, (4) 정상 회복 시간. <그림 1>에 나타난 것처럼, 사고에 대한 지속시간은 사고 발생 후 처리 시간까지의 첫 3 요소만을 포함한다. 즉, 사고 지속시간은 사고에 대한 검지 및 제보시간과 대응 및 처리시간에 의해 결정된다. 최근에는 휴대 전화 보급률 증가로 인해 대부분의 사고는 신속하게 검지된다. 이로 인해, 사고에 대한 신속한 대응과 처리가 사고 지속시간의 최소화는 물론 사고로 인한 혼잡효과의 최소화에 직접적으로 영향을 미치게 된다.

교통사고에 대한 신속한 대응과 처리를 위해서는 구체적인 방안과 각 부처간의 협조가 필요하며, 이러한 작업들은 사고에 대한 특성뿐 아니라 예측된 사고 지속시간을 기반으로 진행될 수 있다. 이러한 예측된 정보는 사고 처리에 대한 구체적인 계획에 도움을 줄 뿐 아니라, 운전자들에게는 유용한 교통정보로 활용될 수 있다. 즉, 예측된 정보로 인한 교통의 분산효과, 돌발적인 혼잡에 대한 예측으로 인한 2차 사고 예방 효과 등으로 이어지게 된다.

따라서, 본 연구의 목적은 우리나라 고속도로에서 발생한 사고 자료를 토대로 교통사고 지속시간에 대한 예측 모형을 개발하는 것이다. 연구를 위해 2004년 한해 24개 주요 고속도로에서 발생한 3,144건의 교통사고 중에서, 자료수집이 완전하게 이루어진 2,940건의 자료가 적용되었다. 또한, 사고 지속시간에 대한 유의한 인자들의



<그림 1> 사고처리에 대한 진행과정

해석 뿐 아니라 지속시간에 대한 예측을 위해 모수적 생존기법(parametric survival analysis)을 적용하였다.

II. 생존 분석

1. 개요

사고의 지속 시간처럼, 특정 사건의 발생을 시간 경과에 따라 관찰 및 분석하는 기법을 생존 분석이라 하며, 주로 의학 분야와 산업 공학에서 사용되어 왔다. 즉, 관찰 대상이 관찰 시점부터 관찰 종료 시점에 도달했을 때, 최종적으로 그 종료 시점까지 얼마나 지속되었는가를 나타내는 생존 기간(survival time)이 주 관심 대상인 것이다. 예를 들면, 고장 전까지 기계의 가동 시간, 환자의 생존 시간, 근로자의 실직 기간 등이 생존 시간에 해당된다. 따라서, 의학 분야에서는 새로이 개발된 치료제와 치료법에 대한 효과를 기존의 약제 및 치료법과 비교 분석하기 위한 임상 시험을 위해, 그리고, 공학 분야에서는 제품의 수명에 대한 통계적인 분석을 위해 주로 사용되어 왔다. 이러한 이유로 인해, 생존 분석은 그 응용 분야에 따라 고장 시간 자료 분석 (failure time data analysis), 지속 시간 분석(duration analysis)으로도 불리지만, 본 연구에서는 사고의 지속 시간(accident duration)과의 혼동을 피하기 위해 생존 분석으로 통칭하여 사용하였다.

2. 생존 분석 기법

생존 분석 기법은 크게 모수적 (parametric) 방법과 비모수적(nonparametric) 방법으로 나누며, 모수적 방법과 비모수적 방법의 중간 형태 (semiparametric 혹은 Cox 모형)를 분류하여 3가지로 분류하기도 한다. 비모수적 방법은 누적 생존률의 산출에 의한 Kaplan Meier 방법과 생명표 (life table) 법, 그리고 두 생존 곡선의 비교를 위한 로그순위검정 (log-rank test)과 윌콕슨 (Wilcoxon), 타론 웨어 (Tarone Ware), 페토 프랜티스 (Peto Prentice) 등의 기법이 있다. 이러한 기법은 위험 함수에 대한 특별한 유형을 지정하지 않기 때문에 비모수적 방법으로 불리우며, 생존 시간에 대한 둘 혹은 그 이상의 그룹간의 비교에 적합하다(Collett, 2003).

그러나, 대부분의 생존 자료에 대한 연구에서는 이러한 단일 변수의 효과뿐만 아니라, 교락요인 (가설상의 요인 외에 처리 효과에 관여하는 제 3의 요인)을 고려한 다변량 분석을 해야하는 경우가 많다. 즉, 생존 시간과 다양한 예측 인자들과의 관계를 알아보는 다변량 분석기법¹⁾ 적용되며, 이때 위험 함수는 지수(exponential), 와이불(Weibull), 로그 정규(log-normal), 로그 로지스틱(log-logistic), 감마(gamma) 분포와 같이 특정 유형을 따르는 것으로 가정하여 모수적 기법으로 불린다. 이러한 모수적 방법은 그 모형의 구조에 따라, 비례적 위험함수 (proportional hazards: PH)의 가정을 적용하는 PH metric 모형과 가속화 시간(accelerated failure time: AFT)을 적용하는 AFT metric 모형으로 분류된다. 또한, PH모형은 위험 함수에 대한 설명변수의 효과를 모형화하지만, AFT모형은 생존 시간 자체에 대한 설명 변수의 효과를 모형화 하기 때문에, AFT 모형이 PH모형보다 예측 목적에 적합하다고 할 수 있다(Cantor, 2003).

마지막으로 Cox모형은 PH 모형과 그 구조는 동일하지만, Cox(1972)는 편우도 (partial likelihood: PL) 추정 방법을 적용하여 위험 함수의 분포적 가정 없이 모형의 계수를 추정하는 방법을 제안하였다. 따라서, Cox 모형을 준모수적(semiparametric) 방법이라 부르기도 한다. 비록 Cox모형이 위험 함수의 모수적 유형을 지정하지 않아도 된다는 장점을 가지고 있지만, 위험 함수 보다는 설명 변수의 효과를 추정에 그 주 목적이 있기 때문에, 모형의 결과를 생존 시간 추정에 사용하기에는 부적합 하다(Hosmer and Lemeshow, 1999)

전술한 바와 같이, 본 연구에서는 사고의 지속시간과 설명 변수와의 관계보다는 사고의 지속시간 자체의 예측을 목적으로 하기 때문에, Cox 모형 보다는 모수적 모형을 적용한다. 또한, 사고 지속 시간의 위험 함수에 대한 설명 변수의 효과를 모형화하는 PH모형보다는 사고 지속 시간 자체에 대한 설명 변수의 효과를 모형화하는 AFT모형을 적용한다. 따라서, AFT모형에 대하여 보다 자세히 설명하기로 한다.

1) AFT 모형의 구조

AFT모형은 로그 생존 시간 T 와 설명 변수 X 의 특성간의 관계를 다음과 같이 선형으로 가정한다:

$$\ln(T) = \beta'X + z \tag{1}$$

1) 기본적으로 다변량 생존분석 기법은 설명변수와 종속변수간의 관계를 알아보는 회귀모형이다.

여기에서 β' 모형의 계수에 대한 벡터를 나타내며, z 는 오차항 (error term)을 나타낸다. 따라서, 이식에서 오차항의 분포에 따라 모형은 지수, 와이블, 로그 정규, 로그 로지스틱, 감마 AFT모형으로 불리게 되며, 이 식은 다음과 같이 다시 표현할 수 있다.

$$Y = \mu + \sigma u \quad \text{or} \quad u = \frac{Y - \mu}{\sigma} \quad (2)$$

여기에서 $Y \equiv \ln(t)$, $\mu \equiv \beta'X$ 이며, $u = z/\sigma$ 는 밀도함수 $f(u)$ 의 오차항이며, σ 는 위험함수에 대한 형태모수 (shape parameter)에 관련된 척도 인자 (scale factor)를 나타낸다. 이 구조는 일반화 선형 모형 (generalized linear model: GLM) 혹은 로그 선형 모형의 구조로 참조되기도 한다.

2) 모형의 해석

ATF 모형은 $\ln(T) = \beta'X + z$ 와 같이 로그 선형 관계를 따르기 때문에, 만일 $\psi \equiv e^{-\beta'X} = e^{-\mu}$ 라고 가정하면,

$$\ln(T\psi) = Z \quad (3)$$

여기에서 ψ 항을 상수로 가정하면, 이 항은 시간 척도 인자와 같은 역할을 하게 된다. 즉, 이 항은 다음과 같은 두 가지의 경우를 형성하게 된다.

$\psi > 1$: 이 경우는 마치 시계의 클락 속도가 빨라지는 것을 의미한다. 따라서, 모형에서 생존 시간이 짧아짐을 의미한다.

$\psi < 1$: 이 경우는 마치 시계의 클락 속도가 느려지는 것을 의미한다. 따라서, 모형에서 생존 시간이 늘어남을 의미한다.

만일 생존 함수에서 이러한 시간 척도 특성을 고려하기 위해, 다음과 같이 생존 함수의 정의해 보자.

$$S(t; \beta, X) = \Pr [T > t | \beta, X] \quad (4)$$

혹은

$$S(t; \beta, X) = \Pr [Y > \ln(t) | \beta, X] \quad (5)$$

따라서,

$$S(t; \beta, X) = \Pr [\sigma u > \ln(t) - \mu] = \Pr [e^{\sigma u} > te^{-\mu}] \quad (6)$$

이 식에서, 만일 모든 설명 변수가 0일 때 (즉, $X=0$) 생존 함수를 기저 생존 함수 (baseline survivor function) $S_0(t)$ 라 한다면,

$$S_0(t) = \Pr [T > t | X=0] \quad (7)$$

혹은

$$S_0(t) = \Pr [e^{\sigma u} > t\psi_0] \quad (8)$$

따라서, 특정 시간 s 에 대하여 이 식은 $S_0(s) = \Pr [e^{\sigma u} > s\psi_0]$ 로 다시 표현할 수 있다. 만일, $s = t \frac{e^{-\mu}}{\psi_0}$ 라 하고, 이 항을 $S_0(s)$ 에 대입하게 되면, 생존 함수는 다음과 같이 표현할 수 있다.

$$S(t; \beta, X) = S_0[te^{-\mu}] = S_0[t\psi] \quad (9)$$

여기에서 $\psi \equiv e^{-\mu}$ 이기 때문에, $\psi > 1$ 일 때 $\mu < 0$ 를 의미하며, $\psi > 1$ 일 때 $\mu < 0$ 를 의미하는 것으로 해석할 수 있다. 따라서, 설명 변수의 효과는 상수 척도 인자 $\psi \equiv e^{-\mu}$ 에 따라 시간 척도가 변하는 것이다.

결과적으로, 설명 변수의 계수를 해석하기 위해, 다음과 같이 로그 생존 시간의 미분항을 고려해 본다면:

$$\frac{\partial \ln(T)}{\partial X_k} = \beta_k \quad (10)$$

만일 AFT모형에서 다른 모든 특성이 변하지 않고, 특정 설명 변수만 한 단위 변화시켰을 때, AFT모형에서의 생존 시간 그 계수에 비례함 비례하여 변화함을 알 수 있다. 이러한 해석은 생존 시간의 변화보다는 위험률의 비례적 변화로 해석되는 PH모형과는 다르다.

III. 사고 자료

1. 개요

2004년 한해 동안 24개 주요 고속도로에서 총 3,144건의 교통 사고가 한국도로공사 교통정보센터에

〈표 1〉 고속도로 교통사고 변수

범주	변수		유형	구성비 (%)	범주	변수		유형	구성비 (%)	
사고 시간	계절	겨울	Dummy	28.0	사고 특성	사고후 상태	추락	Dummy	2.8	
		봄	Dummy	21.0			화재	Dummy	3.0	
		여름	Dummy	33.3			기타	Dummy	0.5	
		가을	Dummy	17.7			2상태 이상	Dummy	6.5	
	요일	공휴일	Dummy	2.01		차량 피해 구분	완과	Dummy	4.8	
		일요일	Dummy	5.3			반과	Dummy	15.9	
		월요일	Dummy	12.9			1/3과	Dummy	15.6	
		화요일	Dummy	12.6			경미	Dummy	41.6	
		수요일	Dummy	12.9			2가지 이상	Dummy	22.1	
	시간	목요일	Dummy	13.4		사고 환경	기상 상태	맑음	Dummy	58.9
		금요일	Dummy	14.8				흐림	Dummy	16.7
		토요일	Dummy	16.1				비	Dummy	20.9
시간	야간	Dummy	40.0	노면 상태	눈		Dummy	2.9		
	오전점두	Dummy	9.4		안개		Dummy	0.6		
	한낮	Dummy	37.0		강풍		Dummy	0.1		
사고 특성	사고 위치	오후점두	Dummy	13.6	갓길 추정차		Yes	Dummy	2.3	
		본선	Dummy	85.8			No	Dummy	97.7	
		터널	Dummy	2.0			작업 구간	Yes	Dummy	5.9
		진출입부 교량	Dummy	1.6				No	Dummy	94.1
	도로 점유 상태	램프	Dummy	7.4	이하 구조 및 도로 시설물 특성		포장 구분	콘크리트	Dummy	50.1
		기타*	Dummy	1.1				평면 선형	아스팔트	Dummy
		본선	Dummy	32.7		적선	좌커브 500미만		Dummy	67.1
		중앙분리대	Dummy	1.8			우커브 500이상	우커브 500미만	Dummy	2.6
	갓길	Dummy	37.7	우커브 500이상		우커브 500이상		Dummy	13.5	
	도로외 반대방향	Dummy	7.8			중단 구배	평탄	Dummy	52.2	
	진출입부	Dummy	4.3	오르막 3%미만			오르막 3%미만	Dummy	18.2	
	2상태 이상	Dummy	12.7			내리막 3%미만	오르막 3%이상	Dummy	2.9	
사고 원인	졸음	Dummy	24.8	내리막 3%이상			내리막 3%미만	Dummy	23.0	
	과속	Dummy	21.9			내리막 3%이상	Dummy	3.7		
	주시태만	Dummy	12.7	포장 상태		양호	Dummy	99.0		
	안전거리미확보	Dummy	4.7			불량	Dummy	10		
	핸들과대조작	Dummy	14.1	중앙 분리대 유형	없음	Dummy	8.3			
	기타 위반**	Dummy	4.7		고정방호벽(81cm)	고정방호벽(81cm)	Dummy	58.4		
	타이어파손	Dummy	8.9	고정방호벽(127cm)		고정방호벽(127cm)	Dummy	15.6		
차량 결함	Dummy	2.8	가드레일		가드레일	Dummy	8.3			
외부***	Dummy	2.1		이동방호벽	이동방호벽	Dummy	3.6			
기타	Dummy	3.3	녹지		녹지	Dummy	2.8			
사고 유형	단독	Dummy		18.0	기타	기타	Dummy	3.0		
	차-시설	Dummy	58.6	갓길 방책 시설		없음	Dummy	10.5		
	차-사람	Dummy	1.4		가드레일	가드레일	Dummy	65.9		
	추돌	Dummy	15.8	가드웬스/낙석방지책		가드웬스/낙석방지책	Dummy	2.9		
	충돌	Dummy	2.0		가드케이블/파이프	가드케이블/파이프	Dummy	0.3		
	화재	Dummy	2.0	콘크리트 방호벽		콘크리트 방호벽	Dummy	10.1		
	접촉	Dummy	0.6		기타	기타	Dummy	10.3		
기타	Dummy	1.6	운전자특성	성별		남성	Dummy	87.9		
차량 유형	승용(자가)	Dummy			37.7	연령	여성	Dummy	12.1	
	승용(SUV)	Dummy		11.0	19세 미만		19세 미만	Dummy	0.4	
	승합(소형)	Dummy		2.0		20~29	20~29	Dummy	18.9	
	승합(중형)	Dummy		6.1	30~39		30~39	Dummy	31.0	
	승합(대형)	Dummy		2.4		40~49	40~49	Dummy	33.2	
	화물(소형)	Dummy		1.8	50~59		50~59	Dummy	13.0	
	화물(중형)	Dummy		9.9		60~	60~	Dummy	3.5	
화물(대형)	Dummy	12.6		운영적 특성	제보 유형		고객	Dummy	17.9	
트레일러	Dummy	9.0				순환중	순환중	Dummy	29.9	
특수 차량	Dummy	5.8			유관기관		유관기관	Dummy	10.4	
승용(영업)	Dummy	1.6				정보센터	정보센터	Dummy	19.7	
사망자****	0	Dummy	91.7		영업소		영업소	Dummy	11.6	
	1+	Dummy	8.3			기타	기타	Dummy	10.5	
	부상자수	0: 1: 2+	정량적		100.		사고후 상태	정지	Dummy	73.7
사고차량	1: 2: 3: 4+	정량적	100.		전도	Dummy		8.9		
사고후 상태	정지	Dummy	73.7		전복	Dummy		4.6		

* 비상활주로, 휴게소, 정류장

** 음주, 추월불량, 적체불량, 통행위반, 중앙선침범 등

*** 노면잡물, 보행자, 동물

**** 사고후 72시간내 사망

접수되었다. 이 자료들은 고속도로상에서 발생하는 교통사고 발생시 현황 파악 목적으로 한국도로공사의 속도양식에 따라 입력된 것이다. 입력된 정보는 사고시간, 사고 특성, 환경조건, 운전자 특성, 도로 기하구조 및 포장 상태, 사고 접수 방법에 의한 6개의 항목으로 구분할 수 있다. 그러나, 몇몇 자료들은 입력과정시 오류로 인해 완전한 형식을 갖추지 못했기 때문에 연구에 제외되었다. 결과적으로 총 2,940건의 교통사고 자료가 분석에 적용되었다.

2. 변수 설정

일반적으로 변수는 자료의 형태에 따라 정량적 변수(quantitative variable)와 정성적 변수(qualitative or categorical variable)로, 인과관계에 따라 독립변수(independent variable)와 종속변수(dependent variable)로 나뉜다. 정량적 변수는 관심대상에 대한 속성을 수치로 표현할 수 있으며 그 값들은 의미가 존재한다. 반면에 정성적 변수는 관심대상이 범주(category)로 구분된다. 이러한 범주는 수치로 표현할 수는 있지만 그 값들은 무의미하다. 예를 들어, 시간대에 대한 변수를 4가지로 범주로 구분하여, AM peak=1, Midday=2, PM peak=3, Night=4와 같이 수치를 부여할 수 있다. 그러나 각 수치들은 범주에 대한 구분 외에 아무런 의미가 존재하지 않는다.

한편, 독립변수는 변수간의 관계에서 다른 변수에 영향을 주는 변수를 의미하며, 종속변수는 독립변수에 의해 그 영향이 결정되는 변수를 의미한다. 사고에 대한 지속시간을 분석할 경우, 사고에 대한 지속시간은 종속변수이며, 사고 지속시간에 영향을 줄 수 있는 사고의 유형, 환경조건, 운전자 특성, 도로의 기하구조, 사고 시간대 등의 변수는 독립변수에 해당된다.

이와 같이, 대부분의 사고관련 변수들은 명목적(nominal) 값으로 구분할 수 있는 범주형 변수들이다. 이러한 변수들은 더미변수(dummy variable)을 이용하여 모형에서 분석될 수 있다. 예를 들어, 전술한 4가지 범주의 교통사고 시간대에 대한 교통사고 지속시간의 관계를 분석하고자 할 때, 오전 침두시간대 교통사고에 대한 범주는, AM peak=1 (교통사고가 오전 침두시간대에서 발생한 경우) 혹은 AM peak=0 (교통사고가 그의 시간대에서 발생한 경우)와 같이 정의할 수 있다. 또한, 사고

발생 요일은 공휴일 변수를 추가하여 휴일과 일반 주중의 교통 특성을 분리하고자 하였다. <표 1>은 한국도로공사에의 교통사고 속도양식에 의해 기록된 교통사고관련 자료를 기반으로 구축된 기본 변수들이다. 이 표에서 계절에 대한 변수는 기상청에서 1996년부터 2005년까지 10년 평균 자료를 기반으로 발표된 자료를 근거로 겨울(11월 29일~03월08일), 봄(03월09일~06월 02일), 여름(06월03일~09월18일), 가을(09월19~11월28일)로 정의되었다. 또한, 시간대에 대한 구분은 야간(20:01~07:00), 오전침두(07:01~09:00), 한낮(09:01~16:30), 오후침두(16:31~20:00) 시간대에 의한 4가지로 구분하였다. 나머지 변수들은 입력된 양식에 기반으로 구분하였으며, 관련항목에 대한 자료의 수가 충분하지 않은 경우 유사한 유형으로 혹은 기타 항목으로 통합하여 정의하였다.

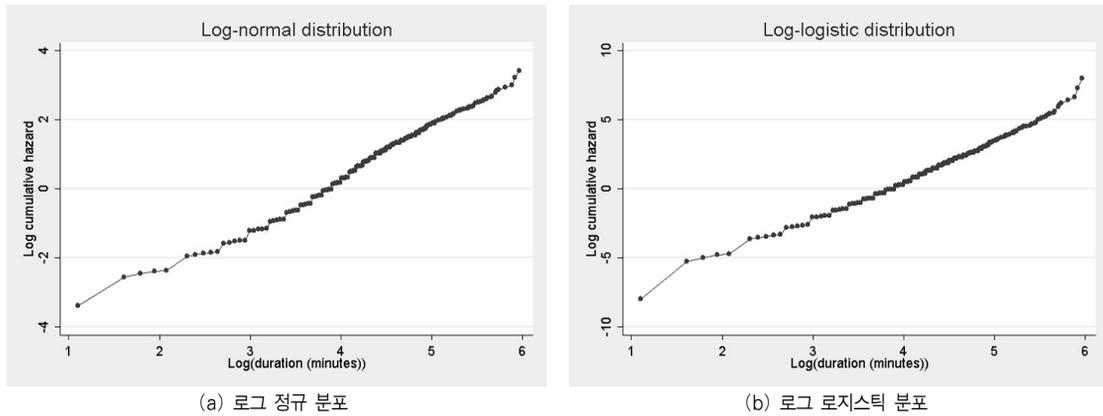
IV. 교통사고 지속시간 예측 모형 개발

1. 모수적 방법을 이용한 예측 모형 개발

1) 모형의 선정

본 연구에서처럼 사고 지속시간에 대한 여러 예측변수(covariate)가 있을 때, 이 변수들의 영향을 동시에 알아보는 다변량 분석(multivariate analysis)이 필요하다. 전술한 바와 같이 생존분석에서의 다변량 분석 기법은 모수적 방법과 Cox 모형이 있으나, Cox 모형의 주 목적은 생존시간의 추정이 아닌 설명 변수의 효과 추정이기 때문에 본 연구의 목적에 적합하지 않다. 또한, 사고 지속 시간의 위험 함수에 대한 설명 변수의 효과를 모형화하는 PH 모형 보다는 사고 지속 시간 자체에 대한 설명 변수의 효과를 모형화하는 AFT모형을 적용하였다. AFT 모형은 사고의 지속시간과 이에 관련된 여러 변수들의 영향을 동시에 알아볼 수 있을 뿐 아니라, 예측의 목적으로 사용될 수 있다.

그러나, 모수적 생존분석기법을 적용하기 위해서는 위험 함수에 대한 적절한 분포를 지정해야 한다. 모수적 생존분석기법을 적용하여 분석을 하고자 할 때 가장 많이 사용되는 분포로는 지수(exponential), 와이불(Weibull), 로그 정규(log-normal), 로그 로지스틱(log-logistic), 감마(gamma) 분포 등이 있으며, 가장 적합한 분포를 선



〈그림 2〉 로그 누적 위험도에 대한 그래프

정하기 위해 사용되는 일반적인 방법은 위험 함수와 선정된 분포와의 관계를 그래프에 의해 비교하는 것이다. 즉, 만일 선정된 분포가 적절하다면, 두 함수의 관계에 대한 그래프는 선형으로 나타나게 된다 (Cantor, 2003).

〈그림 2〉은 위험함수와 선정된 분포와의 관계 그래프에서 비교적 선형 추세를 나타낸 그래프를 나타낸 것이다. 로그 정규분포 (Golob et al., 1987; Giuliano, 1989)와 로그 로지스틱 분포 (Giuliano, 1989; Jones et al., 1991; Stathopoulos and Karlaftis, 2002)는 기존 연구에서 제시된 분포들이다. 결과적으로, 본 연구에서는 이 두 분포를 기반으로 사고 지속시간에 대한 예측 모형을 구축할 것이다.

2) 모형의 구축 과정

전술한 바와 같이 사고 지속 시간에 대한 AFT 모형의 구축을 위해 이미 적절한 두 분포가 선정되었기 때문에, 이후 모형에 포함될 변수 선정에 대한 일련의 과정은 〈그림 3〉에 나타난 것처럼, 기본적으로 일반 회귀 모형의 구축 과정과 유사하다. 즉, 〈표 1〉의 모든 변수들은 예후 변수의 후보로 사용되며, 비 모수적 방법인 로그 순위 검정을 통해 유의한 예후 변수를 우선 선정한다. 이때 유의 수준은 일반적으로 20%를 적용한다 (Fisher et al., 1991). 이렇게 선정된 예후 변수는 다변량 AFT모형에 포함시키기 위해 단계별 변수 선택법 (stepwise selection)을²⁾ 적용한다. 단계별 변수 선정시, P_E , 즉,

유의수준에 대한 기준 (criterion)은 5% (즉, $P_E = 0.05$)를 사용하였다.

그러나, 변수 선택 과정에서 초기 모형에서 제외된 변수라 할지라도, 초기 다변량 모형내에서는 유의한 인자가 될 수 있기 때문에 초기 모형 내에서 검증을 실시해야 한다. 즉, 유의하지 않아 제외되었던 각 예후인자를 초기 모형에 다시 적용하였을 경우, 새로운 모형의 우도비 (likelihood ratio)가 유의하게 변했다면 모형에서 다시 고려하게 되는 것이다. 또한, 새롭게 추정된 초기 모형 내에 연속형 예후 변수가 존재 한다면, 이 변수와 로그 위험 함수와의 관계가 선형인가를 검증해야 한다. Roystone과 Altman (1994)은 분할 다항식 (fractional polynomials)을 적용하여 이러한 선형 관계를 조사하는 방법을 개발하였다.

그들이 제시한 m 차 분할 다항식은 다음과 같이 정의된다.

$$\phi_m(\mathbf{X}; \boldsymbol{\zeta}, \mathbf{p}) = \zeta_0 + \sum_{j=1}^m \zeta_j X^{(p_j)} \quad (11)$$

여기에서 m 은 양의 정수를 나타내며, \mathbf{P} 는 (P_1, P_2, \dots, P_m) 를 나타내는 벡터이며, 이 값들은 $P_1 < \dots < P_m$ 를 의미하는 실수(real number)이다³⁾. 또한, $\boldsymbol{\zeta}$ 는 $(\zeta_0, \zeta_1, \dots, \zeta_m)$ 를 나타내는 벡터이며, $X^{(p_j)}$ 는 Box-Tidwell 변환 (Box-Cox 변환이 아님)을 나타낸다. 즉,

2) 일반 회귀 모형에서의 사용되는 방식과 동일한 단계별 선택법이 사용되었으며, 보다 자세한 내용은 Hosmer와 Lemeshow (2000)의 서적에 기술되어 있다.

3) Roystone과 Altman은 $\rho = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ 내에서 적절한 \mathbf{P} 를 발견할 것을 제안하였다.

$$X^{(p_j)} = \begin{cases} X^{(p_j)} & \text{if } p_j \neq 0 \\ \ln X & \text{if } p_j = 0 \end{cases} \quad (12)$$

지금까지의 과정을 통해 수립된 모형은 일반 회귀 모형에서처럼 모형의 평가과정을 거치게 된다. 본 연구에서는 이상치(outlier) 혹은 영향 관측치(influential observation)에 대한 평가 방법과 모형의 전반적인 적합도(overall fit)에 대한 평가 방법이 적용되었다. 첫 번째로, 모형에서 이상치와 영향 관측치를 측정하기 위해 스코어 잔차(score residuals)와 우도 이동(likelihood displacement)⁴⁾ 통계량이 적용되었다. 스코어 잔차는 추정된 각 변수에 대한 이상치 혹은 영향 관측치를 파악하기 위해 사용되는 방법이며, 우도 이동 통계량은 모형 전체에 대한 영향 관측치를 파악하기 위해 사용되는 방법이다. 만일, 이 두 방법에 의해 이상치 혹은 영향 관측치가 발견된 경우에는 관련 관측자료를 제거한 후, 모형을 다시 추정하게 된다. 또한, 모형 전반적인 적합도에 대한 평가를 위해 Cox Snell 잔차를 적용하였다.

마지막으로 관측되지 않은 인자들에 대한 영향력(frailty 혹은 unobserved heterogeneity)을 알아보는 단계가 필요하다. 즉, 지금까지 구축된 다변량 모형은 모든 개개의 관측치 사이에서 발생하는 차이점은 관측된 설명변수에 의해 설명되어진다고 가정하였다. 그러나,

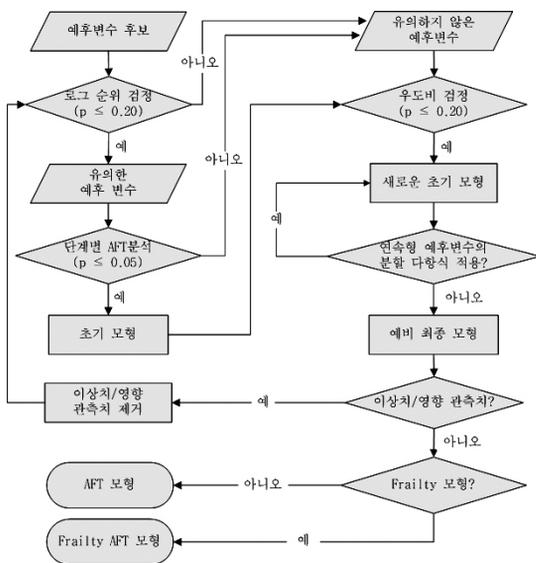
관측되지는 않았지만 모형에 영향을 줄 수 있는 랜덤 효과(radom effect) 혹은 프레이리티(frailty)를 설명하기 위한 모형이 평가되었다. 이와 같은 모형을 프레이리티 모형이라 하며 (Vaupel et al., 1979), 측정되지 않은 새로운 예후 변수를 위험 함수에 추가함으로써 이루어진다. 또한, 새로이 추가된 변수에 대한 위험도는 음의 값을 가질수 없기 때문에, 양의 값을 가지는 분포인 감마(gamma) 분포가 가장 많이 사용된다.

3) 로그 정규 AFT 모형

〈표 2〉은 전술한 모형의 구축 과정을 통해 추정된 사고 지속시간에 대한 로그 정규 AFT모형을 나타낸다. 이 모형은 21개의 인자들이 유의수준 5%에서 유의하게 추정되었다. 만일 모형의 계수가 양의 값을 나타낸다면 이 변수는 사고지속시간에 더 많은 영향을 주는 변수이며, 그 반대의 경우 지속시간에 적은 영향을 미치는 변수로 해석된다. 예를 들어, 고속도로 순찰대가 순찰중에 발견된 사고는 추정계수의 값이 0.322로, 사고의 지속시간은 다른 매체에 의해 접수된 사고보다 그 지속시간이 짧음을 알 수 있다.

전술한 바와 같이 추정된 모형에 대한 평가는 일반 회귀 모형과 매우 유사하다. 첫째로, 생존분석에서 모형에 대한 영향 관측치를 알아보기 위해, 스코어 잔차 혹은 우도 이동 통계량(likelihood displacement statistic 혹은 Cook's distance)이 주로 사용된다. 스코어 잔차는 각 변수에 대한 영향관측을 알아볼 때 용이하게 사용되며, 〈표 2〉의 모형에서는 특별한 영향 관측치를 발견하지 못했다. 또한, 모델 전체에 대한 영향관측을 알아볼 때 우도 이동 통계량이 주로 사용된다. 우도 이동 통계량과 마팅게일 잔차(Martingale residuals)에 대한 관계 그래프를 그렸을 경우, 만일 모든 점들이 컵 모양으로 분포되어 있으며 서로 근접해 있다면, 모형에 특별한 영향관측은 없는 것으로 판단된다. 〈그림 4〉의 (a)는 우도 이동 통계량과 마팅게일 잔차와의 관계를 나타내며, 이 그림에서 두 점이 다른 점들에 비해 떨어져 있는 것으로 보이나, 이 두 점(사고 자료)을 삭제한 후, 동일한 모형을 추정한 결과 각 변수에 대한 계수들에 대한 변화는 20%를 넘지 않았다. 즉, 이 두 사고 자료는 모델에 영향을 주는 관측치가 아님을 증명하는 것이다.

두 번째로, 위험함수에 대해 설정된 분포의 적합성과



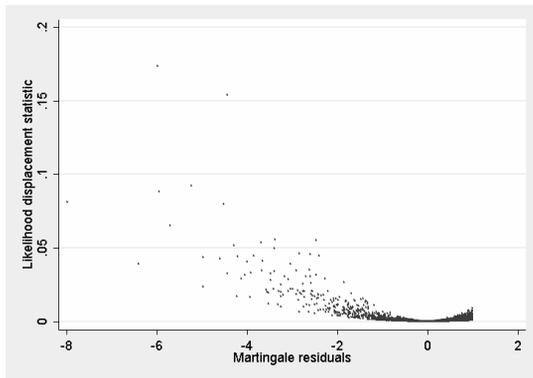
〈그림 3〉 AFT 모형의 구축 과정

4) 혹은 쿡의 거리 (Cook's distance)라고도 불린다.

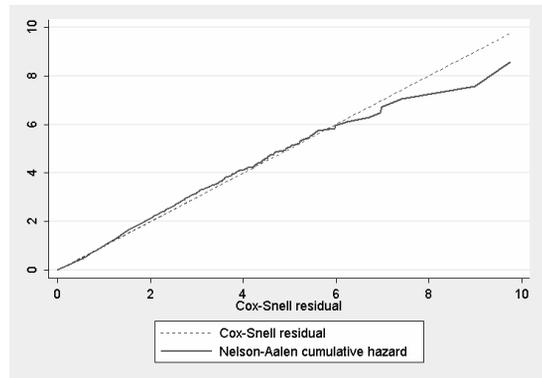
〈표 2〉 사고 지속시간에 대한 로그 정규 AFT 모형

변수	계수 (β)	표준 편차	χ^2	p-value	95% CI (β)	
					하위	상위
제보유형: 순찰중	-0.322	0.022	207.936	0.000	-0.366	-0.278
차량유형: 트레일러	0.491	0.037	171.872	0.000	0.418	0.565
차량유형: 특수차량	0.532	0.045	140.186	0.000	0.444	0.620
차량유형: 대형트럭	0.310	0.032	93.509	0.000	0.247	0.373
차량유형: 승용(영업)	0.562	0.081	48.303	0.000	0.403	0.720
사망자: 1+	0.204	0.039	27.773	0.000	0.128	0.280
사고후: 정지	-0.140	0.027	27.248	0.000	-0.193	-0.088
차량피해구분: 완파	0.261	0.051	26.626	0.000	0.162	0.360
차량유형: 소형트럭	0.334	0.077	18.923	0.000	0.183	0.484
작업구간: yes	0.191	0.044	18.923	0.000	0.105	0.277
차량피해구분: 반파	0.120	0.031	15.210	0.000	0.060	0.180
사고후 상태: 전복	-0.179	0.053	11.156	0.001	-0.284	-0.074
사고유형: 차-시설	-0.071	0.022	10.368	0.001	-0.115	-0.028
갓길 방책시설: 가드웬스/낙석방지책	0.186	0.060	9.548	0.002	0.068	0.304
계절: 가을	-0.078	0.027	8.526	0.003	-0.131	-0.026
사고차량수	0.044	0.016	7.728	0.005	0.013	0.076
차량피해구분: 1/3파	-0.078	0.030	6.760	0.009	-0.136	-0.019
도로점유상태: 갓길	-0.055	0.022	6.101	0.013	-0.099	-0.011
부상자수	0.040	0.017	5.664	0.017	0.007	0.074
차량유형: 중형트럭	0.083	0.035	5.570	0.018	0.014	0.152
시간대: 야간	0.048	0.021	5.198	0.023	0.007	0.089
상수	3.841	0.043	8031.744	0.000	3.757	3.925
$1/\ln(\sigma)$	-0.598	0.013	2105.892	0.000	-0.624	-0.573
σ	0.550	0.007			0.536	0.564

-2 log likelihood = 4,824.196



(a) 우도 이동 통계량



(b) 로그 정규 분포

〈그림 4〉 로그 정규 AFT 모형 평가에 대한 그래프

추정된 모형의 적합성을 평가할 때, Cox Snell 잔차를 이용한 그래프가 주로 사용된다. 만일 추정된 모형이 적합한 경우, Cox Snell 잔차에 대한 추정된 누적 위험 추정치는 원점을 지나며 기울기가 1인 그래프를 따라 형성된다(Hosmer and Lemeshow, 1999). 〈그림 4〉의 (b)는 Cox Snell 잔차의 누적 위험함수를 나타낸다.

비록 그래프의 마지막 부분에서 두 그래프의 편차가 생겼지만, 그 방향이 수렴하는 방향이므로, 사고 지속시간에 대한 모형이 잘못 추정되었다고 말할 수 없다. 마지막으로 프레이티를 고려한 로그 정규 AFT모형을 평가하였으나, 그 효과가 유의하지 않았기 때문에 프레이티 모형은 생략되었다.

4) 로그 로지스틱 AFT 모형

〈표 3〉는 로그 정규 AFT 모형의 추정과정과 동일한 과정을 통해 추정된 사고 지속시간에 대한 로그 로지스틱 AFT모형을 나타낸다. 이 모형은 로그 정규 모형과 매우 유사하지만, 사고 원인에 대한 변수로 기타 위반에 대한 변수가 더 추가되었다. 표 하단에 표시된 바와 같이

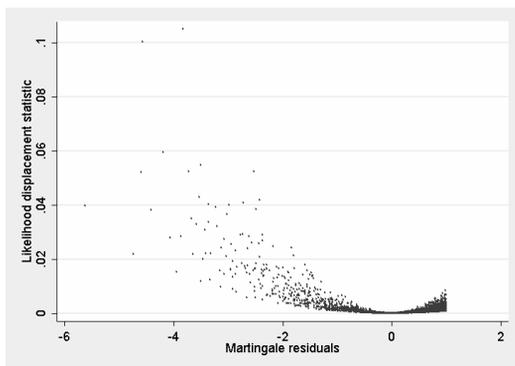
-2 로그 우도 (log likelihood) 값은 로그 정규에서 4,824.196 보다 4,733.875로 향상되었다.

로그 정규 AFT모형에서와 마찬가지로, 스코어 잔차 및 우도 이동 통계량을 적용하여 영향관측을 평가하였다. 〈그림 5〉에 나타난 것처럼 추정된 모형은 특별한 영향관측이 발견되지 않았으며, 특히, 전체적으로 우도 이동 통계량은 0.12보다 작으며, 로그 정규 모형보다 낮은

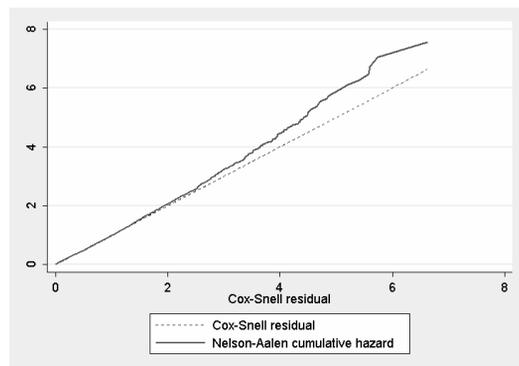
〈표 3〉 사고 지속시간에 대한 로그 로지스틱 AFT 모형

변수	계수 (β)	표준 편차	χ^2	p-value	95% CI (β)	
					하위	상위
제보유형: 순찰중	-0.301	0.022	189.338	0.000	-0.343	-0.258
차량유형: 트레일러	0.488	0.036	180.365	0.000	0.417	0.559
차량유형: 특수차량	0.532	0.045	138.298	0.000	0.443	0.620
차량유형: 대형트럭	0.303	0.031	95.063	0.000	0.242	0.364
차량유형: 승용(영업)	0.541	0.077	49.844	0.000	0.391	0.691
차량피해구분: 완파	0.271	0.048	31.360	0.000	0.176	0.365
사망자: 1+	0.191	0.036	28.730	0.000	0.121	0.261
사고후: 정지	-0.129	0.026	25.301	0.000	-0.180	-0.079
작업구간: yes	0.198	0.041	23.136	0.000	0.118	0.279
차량유형: 소형트럭	0.311	0.073	18.318	0.000	0.169	0.454
차량피해구분: 반파	0.118	0.029	16.810	0.000	0.062	0.174
사고후 상태: 전복	-0.155	0.049	9.797	0.002	-0.252	-0.058
사고유형: 차-시설	-0.066	0.021	9.797	0.002	-0.107	-0.025
계절: 가을	-0.078	0.025	9.425	0.002	-0.128	-0.028
사고차량수	0.046	0.015	9.120	0.003	0.016	0.076
갓길 방책시설: 가드웬스/낙석방지책	0.164	0.059	7.673	0.006	0.048	0.280
부상자수	0.044	0.016	7.618	0.006	0.013	0.075
차량피해구분: 1/3파	-0.072	0.029	6.150	0.013	-0.129	-0.015
차량유형: 중형트럭	0.082	0.033	6.052	0.014	0.017	0.147
시간대: 야간	0.048	0.020	5.664	0.017	0.008	0.087
사고원인: 기타 위반	-0.107	0.045	5.570	0.018	-0.196	-0.018
도로점유상태: 갓길	-0.049	0.021	5.198	0.023	-0.090	-0.007
상수	3.840	0.041	8952.944	0.000	3.760	3.919
1/ln(σ)	-1.196	0.015	5996.954	0.000	-1.226	-1.166
σ	0.302	0.005			0.293	0.312

-2 log likelihood = 4,733.875



(a) 우도 이동 통계량



(b) 로그 로지스틱 분포

〈그림 5〉 로그 로지스틱 AFT 모형 평가에 대한 그래프

값을 형성하고 있다. 또한, Cox Snell 잔차의 누적 위험함수는 로그 로지스틱 모형의 타당성을 보여준다. 마지막으로 프레일티를 고려한 로그 로지스틱 AFT모형을 평가하였으나, 그 효과가 유의하지 않았기 때문에 프레일티 모형은 고려되지 않았다.

2. 모형의 선정

위험함수에 대한 가정된 두 분포에 의해 사고 지속시간에 대한 두 모형이 추정되었다. 비록 로그 로지스틱 모형이 로그 우도값에서 보다 우수한 모형으로 보이지만, 두 모형의 파라미터는 서로 다르므로 두 모형의 비교는 로그 우도값에 의한 방법보다 다른 방법을 적용해야 한다. Akaike(1974)는 모형의 복잡성(complexity)과 모델의 로그 우도를 동시에 고려한 모형 선정 방법을 제시하였다. 이 방법은 AIC (Akaike's Information Criterion)라 불리우며, 낮은 AIC값을 형성하는 모형이 우수한 모형으로 선정된다.

모형의 AIC는 다음과 같이 정의 된다.

$$AIC = -2LL + 2(k + c + 1) \tag{13}$$

여기에서 LL은 모형의 로그 우도값을 나타내며, k는 변수의 수를 나타낸다. 또한, c는 설정된 분포의 파라미터를 나타내며, 예를 들어, 로그 정규와 로그 로지스틱 분포의 경우 이 값은 1이 된다. 결론적으로, AIC는 모형에서의 파라미터 수와 로그 우도 값의 편차에 대한 두 배 값을 나타낸다. <표 4>에 나타난 것처럼, 비록 로그 정규 모형이 모형의 복잡성에서는 약간 우수할 수 있으나, 로그 우도 값 및 AIC 값에 의해 로그 로지스틱 모형을 최종적으로 선정하였다.

<표 4> 로그 정규 및 로그 로지스틱 AFT모형에 대한 AIC

모형	-2LL	k	c	AIC
로그 정규	4,824.196	21	1	4,870.196
로그 로지스틱	4,733.875	22	1	4,781.875

3. 모형의 해석

각 변수에 대한 해석은 추정된 계수에 대하여 지수함

수를 취함으로써 시간의 척도로 이루어진다. 예를 들어, <표 3>에서 고속도로 순찰대의 순찰중 제보된 교통사고의 경우, 추정된 계수값은 0.301이다. 때문에, 이 값에 지수함수를 취하면 $\exp(-0.301) = 0.740$ 이 되며, 마찬가지로, 이 변수의 95% 신뢰 구간은 $\exp(-0.343) = 0.710$ 과 $\exp(-0.258) = 0.773$ 이 된다. 따라서, 순찰중에 제보된 교통사고는 다른 유형에 의해 접수된 교통사고에 비해 그 지속시간은 26%감소하며, 95% 신뢰구간 하에 29%와 22.7%사이에서 사고 지속시간의 감소가 형성된다. 다른 변수들도 동일한 방식으로 해석할 수 있으며, <표 5>은 추정된 로그 로지스틱 AFT모형의 시간척도에 의한 해석을 나타낸 것이다. 이 표에서 계수의 값이 양의 값이거나 이 값에 대한 시간 비율(time ratio)이 1보다 큰 변수는 사고 지속시간에 더 많은 영향을 미친다. 예상한 바와 같이, 트럭 및 특수 차량에 의한 사고는 소형 차량에 의한 사고보다 그 처리 시간이 오래 걸리는 것으로 나타났으며, 특히, 트레일러와 특수 차량에 의한 사고는 다른 유형의 차량에 의한 사고 보다 그 지속시간 60% 이상 오래 걸리는 것으로 나타났다. 또한, 비록 승용차량이지만 택시와 같은 영업용 승용차량에 의한 교통사고는 그 지속시간이 다른 사고에 비해 약 72% 오래 지속되는 것으로 나타났다. 차량의 파손은 그 파손 정도에 따라 사고의 지속시간이 오래 걸리는 것으로 나타났으며, 특히, 차량의 파손이 반파 미만인 경우 사고의 지속시간은 줄어드는 것으로 나타났다. 사고후 차량의 상태에 따른 사고의 지속시간에 대한 효과 또한 모형에서 유의한 것으로 나타났으며, 사고후 차량의 상태는 전복 혹은 정지하였을 경우 다른 상태, 즉, 전도, 추락, 화재가 발생한 경우, 혹은 차량의 상태가 두가지 이상의 형태(예를 들면, 차량 한대는 정지와 다른 차량은 전복)로 타난 경우 그 지속시간은 약 12~14% 낮은 것으로 나타났다.

교통 사고에 대한 지속시간은 사망자가 발생한 경우와 그렇지 않은 경우로 나뉘어 분석되었으면, 그 결과 예상대로 사망자가 발생한 사고의 경우 그 지속시간이 그렇지 않은 사고에 비해 21% 정도 오래 지속되는 것으로 나타났다. 또한, 교통 사고 발생시 부상자의 수가 1명 증가 할 때 마다, 그리고, 사고관련 차량의 수가 1대 증가 할 때 마다 그 지속시간은 약 5% 정도 증가하는 것으로 나타났다. 교통사고 발생 시간대에 대한 지속시간 또한 유의하게 나타났으며, 교통사고가 가을에 발생한 경우는 다른 계절에 발생한 사고보다 그 지속시간이 약 7% 적게 나타났으며, 야간 시간대(20:01~07:00)에 발생한

사고는 다른 시간대에 발생한 사고보다 그 지속시간 5% 증가하는 것으로 나타났다.

사고의 원인의 경우, 기타 경미한 위반에 의한 사고는 졸음, 운전미숙, 차량결함 등에 의한 사고보다 그 지속시간이 10% 정도 낮은 것으로 나타났다. 사고의 유형은 차량과 도로의 시설물 간에 발생한 사고가 기타 차량간 사고 혹은 차량과 사람과의 사고보다 그 지속시간이 6% 낮게 형성되었다. 또한, 사고 후 차량이 갓길을 점유하고 있는 사고는 본선 혹은 중앙 분리대를 점유한 사고보다 약 5% 그 지속시간이 낮은 것으로 나타났으며, 만일 사고 구간의 갓길 방책 시설이 가드웬스 혹은 낙석 방지책이 설치된 도로 구간의 경우 방책시설이 없는 평지이거나, 가드레일, 혹은 가드 케이블 혹은 콘크리트 방호벽보다 그 지속시간이 약 18% 높게 형성되는 것으로 나타났다. 마지막으로, 작업 구간에서 발생한 사고는 일반 구간에서 발생한 사고보다 그 지속시간이 22% 정도 오래 지속되는 것으로 나타났다.

그러나, 사고의 지속시간은 요일별 다르게 형성되지 않았으며, 발생한 사고의 위치 또한, 사고의 지속시간에는 아무런 영향을 미치지 않는 것으로 나타났다. 또한, 기상 상태와 노면의 상태 역시 사고의 지속시간에는 영향을 주지 않는 것으로 나타났으며, 도로의 포장 재질,

평면 선형, 종단 구배, 도로의 포장 상태, 중앙분리대의 유형과 같은 도로의 기하구조 및 시설물의 특성은 사고의 지속시간에 영향을 주지 않는 것으로 나타났다. 운전자의 특성 또한 성별, 연령별 그 특성을 분석하였으나, 사고의 지속시간에는 아무런 영향을 미치지 않는 것으로 분석되었다.

V. 결론

본 논문은 고속도로에서 발생한 교통사고 지속시간에 대한 영향인자를 비모수적 방법을 이용하여 분석하고, 이를 토대로 지속시간을 예측할 수 있는 모수적 모형을 제시하였다. 연구에 적용된 생존분석 모형은 주로 의학과 산업공학 분야에서 사람의 생존율 혹은 기계의 고장에 대한 분석 목적으로 각주로 사용되어 왔다. 비록 교통 분야에서는 생존분석 용어가 다소 생소하지만, 통행 발생의 주기 혹은 본 연구에서와 같이 사고 지속시간에 대한 분석 목적으로 몇몇 연구가 시도된 바 있다.

본 연구에서는 로그 정규 및 로그 로지스틱 분포 기반의 AFT 모형을 적용하여 사고 지속시간을 분석하였으나, AIC에 의해 로그 로지스틱 모형이 최종적으로 선정되었다. 추정된 모형에 의해 트레일러, 특수차량, 트럭,

〈표 5〉 로그 로지스틱 AFT모형의 시간 척도에 의한 해석

변수	계수 (β)	시간 비율 (e^{β})	95% CI (e^{β})	
			상위	하위
제보유형: 순찰중	-0.301	0.740	0.710	0.773
차량유형: 승용(영업)	0.541	1.718	1.478	1.996
차량유형: 특수차량	0.532	1.702	1.558	1.860
차량유형: 트레일러	0.488	1.629	1.517	1.749
차량유형: 소형트럭	0.311	1.365	1.184	1.575
차량유형: 대형트럭	0.303	1.354	1.274	1.439
차량유형: 중형트럭	0.082	1.085	1.017	1.159
차량피해구분: 완파	0.271	1.311	1.192	1.441
차량피해구분: 반파	0.118	1.125	1.064	1.190
차량피해구분: 1/3파	-0.072	0.931	0.879	0.985
사고후 상태: 전복	-0.155	0.857	0.777	0.944
사고후 상태: 정지	-0.129	0.879	0.835	0.924
사망자: 1+	0.191	1.211	1.129	1.298
부상자수	0.044	1.045	1.013	1.078
사고차량수	0.046	1.047	1.016	1.079
계절: 가을	-0.078	0.925	0.880	0.972
시간대: 야간	0.048	1.049	1.009	1.091
사고원인: 기타 위반	-0.107	0.898	0.822	0.982
사고유형: 차-시설	-0.066	0.936	0.898	0.976
사고 후 도로점유상태: 갓길	-0.049	0.953	0.914	0.993
갓길 방책시설: 가드웬스/낙석방지책	0.164	1.178	1.049	1.323
작업구간: yes	0.198	1.219	1.125	1.322

영업용 승용 차량에 의한 사고, 차량의 피해가 반과 이상인 사고, 사망자가 발생한 사고, 작업구간에서 발생한 사고, 방책시설이 가드웬스 혹은 낙석 방지책 구역에서 발생한 사고, 야간에 발생한 사고, 그리고, 사고 관련 차량 수와 부상자 수가 증가함에 따라 지속시간은 증가하는 것으로 해석되었다. 그러나, 고속도로 순찰대의 순찰 중 접수된 사고, 사고 후 차량이 정지된 사고, 차량 피해가 반과 미만인 사고, 사고 후 차량이 전복된 사고, 가을에 발생한 사고, 차량이 시설물을 부딪친 사고, 사고 후 갓길을 점유한 사고, 그리고, 기타 위반에 의한 사고는 지속시간에 적은 영향을 미치는 것으로 나타났다. 특히, 고속도로 순찰자에 의해 제보된 사고의 경우 그 지속시간이 26% 감소하는 것으로 나타났으며, 따라서, 빠른 제보와 정확한 사고 정보의 제보는 사고의 지속시간의 감소에 매우 밀접한 관계가 있음을 알 수 있다.

제시된 로그 로지스틱 AFT모형은 사고 지속시간에 대해 영향을 미치는 인자의 파악뿐 아니라, 사고 발생시 사고 관련 기본정보의 접수와 동시 그 지속시간의 예측이 가능하다. 따라서, 본 연구에서 제시된 모형은 고속도로 운영자에게 사고에 대한 적절한 대응 전략의 수립에 도움을 줄 뿐 아니라 운전자들에게 실시간 교통정보 제공 용도로서 활용될 수 있다. 특히, 실시간 교통정보는 예측된 정보에 의해 이루어지며, 교통사고와 같은 돌발상황이 발생되었을 경우 더욱 가치가 있다. 즉, 돌발상황이 없이 도로의 지체가 특정 패턴을 형성하는 경우 운전자는 도로의 소통상황을 예측하게 되지만, 돌발상황이 발생된 경우에는 교통정보 센터에서 예측된 정보를 제공할 수 있다. 이러한 예측된 교통사고 정보는 운전자들의 대안도로 선정에 의한 교통수요 분산 효과뿐 아니라, 운전자들의 정체 구간에 대한 불안감 해소 및 2차 사고 감소로 그 효과가 이어질 것으로 판단된다.

참고문헌

1. Akaike, H. (1974), "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, 19(6), pp.716~723.
2. Bendel, R. B. and A. A. Afifi (1977), "Comparison of Stopping Rules in Forward "Stepwise" Regression", *Journal of the American Statistical Association*, 72(357), pp.46~53.
3. Cantor, A. (2003), "SAS survival analysis

techniques for medical research", Cary, NC, SAS Institute.

4. Collett, D. (2003), "Modelling survival data in medical research", Boca Raton, FL, Chapman & Hall/CRC.
5. Costanza, M. C. and A. A. Afifi (1979), "Comparison of Stopping Rules in Forward Stepwise Discriminant Analysis" *Journal of the American Statistical Association*, 74(368), pp.777~785.
6. Cox, D. R. (1972), "Regression Models and Life Tables", *Journal of the Royal Statistical Society, Series B (Methodological)*, 34(2), pp.187~220.
7. De Rose Jr., F. (1964), "An analysis of random freeway traffic accidents and vehicle disabilities", *Highway Research Record*, 59, pp.53~65.
8. Fisher, B., S. Anderson, et al. (1991), "Significance of ipsilateral breast tumour recurrence after lumpectomy", *The Lancet*, 338(8763), pp.327~331.
9. Garib, A., A. E. Radwan, et al. (1997), "Estimating Magnitude and Duration of Incident Delays", *Journal of Transportation Engineering*, 123(6), pp.459~466.
10. Giuliano, G. (1989), "Incident characteristics, frequency, and duration on a high volume urban freeway", *Transportation Research Part A: General*, 23(5), pp.387~396.
11. Golob, T. F., W. W. Recker, et al. (1987), "An analysis of the severity and incident duration of truck involved freeway accidents", *Accident Analysis & Prevention*, 19(5), pp.375~395.
12. Goolsby, M. E., W. Smith, et al. (1971), "Influence of incidents on freeway quality of service", *Highway Research Record*, 349, pp.41~46.
13. Hosmer, D. W. and S. Lemeshow (1999), "Applied survival analysis: regression modeling of time to event data", New York, Wiley.
14. Jones, B., L. Janssen, et al. (1991), "Analysis of the frequency and duration of freeway

- accidents in Seattle", *Accident Analysis & Prevention*, 23(4), pp.239~255.
15. Juge, J., K. Kennedy, et al. (1974), "Early detection and rapid removal of disabled vehicles and other hazards from the freeway", California Department of Transportation and Department of California Highway Patrol.
 16. Khattak, A., J. Schofer, et al. (1995), "A Simple Procedure for Predicting Freeway Incident Duration", *IVHS Journal*, 2(2), pp.113~138.
 17. Nam, D. and F. Mannering (2000), "An exploratory hazard based analysis of highway incident duration", *Transportation Research Part A: Policy and Practice*, 34(2), pp.85~102.
 18. Ozbay, K. and P. Kachroo (1999), "Incident management in intelligent transportation systems", Boston, Artech House.
 19. Royston, P. and D. G. Altman (1994), "Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling", *Applied Statistics*, 43(3), pp.429~467.
 20. Smith, K. and B. L. Smith (2001), "Forecasting the Clearance Time of Freeway Accidents", Smart Travel Lab Report. Charlottesville, VA, Center for Transportation Studies (University of Virginia).
 21. Stathopoulos, A. and M. G. Karlaftis (2002), "Modeling Duration of Urban Traffic Congestion", *Journal of Transportation Engineering*, 128(6), pp.587~590.
 22. TRB (1994), "Highway capacity manual: special report 209", Washington, D. C., Transportation Research Board, National Research Council.
 23. Vaupel, J. W., K. G. Manton, et al. (1979), "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality", *Demography*, 16(3), pp.439~454.
 24. 백승걸 · 박재범 (2004), "고속도로 돌발상황으로 인한 교통영향 예측시스템 고찰", *도로교통*, 94(겨울), pp.30~40.

✉ 주 작성 자 : 정연식

✉ 교 신 저 자 : 정연식

✉ 논문투고일 : 2007. 3. 13

✉ 논문심사일 : 2007. 5. 26 (1차)

2007. 7. 30 (2차)

2007. 8. 7 (3차)

2007. 8. 27 (4차)

✉ 심사판정일 : 2007. 8. 27

✉ 반론접수기한 : 2008. 2. 29