

# 구조방정식모형을 이용한 고속도로 교통사고 심각도 분석

## Analysis of Traffic Accident Severity for Korean Highway Using Structural Equations Models

이 주 연

(연세대학교 도시공학과  
박사과정)

정 진 혁

(연세대학교 도시공학과  
부교수)

손 봉 수

(연세대학교 도시공학과  
교수)

### 목 차

#### I. 서론

1. 연구의 배경 및 목적
2. 연구의 방법 및 범위

#### II. 관련 연구 고찰

#### III. 구조방정식 모형의 기법

#### IV. 모형의 추정

1. 모형의 구조 및 변수 설정
2. 모형의 추정 및 결과

#### V. 결론 및 향후 연구과제

#### 참고문헌

## I. 서론

### 1. 연구의 배경 및 목적

교통사고를 감소시키기 위해, 사고에 영향을 미치는 요인들을 분석하여 교통사고를 예측하는 모형은 지속적으로 개발되어 왔다. 우리나라의 고속국도 총연장은 약 3,000km에 이르며, 이는 전 세계에서 10위 안에 드는 수치이다. 그러나 고속국도 1km당 사고 건수는 다른 나라들에 비하여 현격히 높은 실정인데, 1980년대 이래로 빠르게 증가한 교통수요와 교통관련 인프라가 이러한 높은 사고율에 영향을 미친 것으로 보인다.

사고율과 함께 중요하게 인식되는 지표는 사고의 심각도이며, 사고의 심각도는 도로의 기하구조나 운전자 행태, 차종, 날씨 등 많은 요인들에 의해 직접적인 영향을 받을 것으로 판단된다. 이 밖에도 여러 요인들이 복합적으로 작용하여 사고를 일으키고, 사고의 심각도에 영향을 미칠 것으로 보인다. 구조방정식(Structural Equations Model)은 이처럼 여러 가지 변수들 간의 복잡한 관계를 규명하는데 적합한 모형으로, 사고 심각도에 영향을 미치는 요인들을 도로 요인 및 운전자 요인, 환경 요인 등으로 구분하고 구조방정식을 이용하여 각 변수들간의 관계를 분석하였다.

### 2. 연구의 범위 및 방법

본 연구에서는 사고의 심각도에 영향을 미치는 요인을 크게 '도로 요인' 및 '운전자 요인', '환경 요인' 등으로 구분하고 각 요인들을 대표할 수 있는 변수들을 선정하여 각각의 변수들이 사고 심각도에 미치는 영향을 분석하였다. 모형은 구조방정식을 적용하여 추정하였고, LISREL 8.51의 PRELIS 및 SIMPLS 프로그램을 이용하였다. 모형 추정에 이용한 자료는 한국도로공사관내의 2005년 전국 고속도로사고정보 데이터이며, 총 2,880개의 사고데이터가 분석에 이용되었다. 분석대상 고속도로는 경부선, 경인선, 제2경인선, 영동선, 서울외곽순환선, 중부선, 제2중부선, 서해안선 등 총 24개 전국의 고속도로이다.

## II. 관련 연구 고찰

교통사고를 분석하고 예측하기 위한 연구들은 오래전부터 지속적으로 진행되어 왔다. 고속도로 사고를 대상으로 한 연구의 경우 대부분 트럼펫 인터체인지에서 사고를 분석하였는데, 이는 우리나라 고속도로 IC의 대부분이 트럼펫 형식으로 구성되어 있어 자료의 구득이 쉽고 연구결과의 활용도가 높기 때문인 것으로 판단된다.

박호신(2007)은 고속도로의 트럼펫 인터체인 지상에서 일어나는 교통사고와 도로 기하구조 및 교통량 등의 교통사고 요인들과의 관계를 분석하였으며, 음이항분포를 적용하여 연결로의 형식별(직결, 준직결, 루프연결로) 음이항회귀모형을 개발하였다. 모형에 이용된 변수는 교통량과 중차량 비율, 도로의 곡률 및 곡선장, 유출입여부 등이며 이러한 변수들은 사고를 설명하는데 유의한 변수들임을 보였다.

김태영(2007)은 트럼펫 인터체인지 램프에서의 사고자료를 분석하여 회전방향 및 도로 기하구조 요소와 교통사고와의 관계를 분석하였다. 분석결과 우측으로 회전하는 램프가 좌측으로 회전하는 램프에 비해 사고건수가 높은 것으로 나타났으며, 회전방향에 종단경사와 곡선반경을 조합하여 사고건수를 비교한 결과 우회전-상향 램프와 좌회전-곡선반경 200 이상의 램프가 다른 조건의 램프에 비해 사고 잠재력이 더 높은 것으로 분석되었다.

윤병조(2006)는 트럼펫 인터체인의 사고를 대상으로 분석을 수행하여 연결로의 유입·유출 방향 및 연결로의 형식에 따라 사고율에 차이가 있음을 보였다. 연결로 형식별 사고건수는 루프, 준직결, 직결의 순으로 높게 나타났으며, 유출연결로의 사고건수는 전체 연결로 사고의 약 74%를 차지하고 사고율은 유입연결로에 비해 2.84% 높은 것으로 나타났다.

박정순(2007)은 4지 신호교차로를 대상으로 도로환경 요인과 교통사고와의 상관관계를 분석하고 사고추정모형을 개발하였다. 다중회귀모형의 추정결과 ADT, 주도로의 평균차로폭원, 횡단보도수, 종단경사, 제한속도차이 등이 교통사고건수에 많은 영향을 미치는 것으로 나타났으며, 교통사고의 98% 이상이 어느 한 가지 요인에 의해 발생하는 것이 아니라 인적요인, 차량요인, 도로환경요인 중 두 가지 이상 요인들의 복합적인 작용에 의해서 발생하는 것으로 나타났다.

구조방정식 모형을 이용하여 수행된 기존 연구 역시 매우 다양하다.

추상호(2007)는 통행수요, 교통시설공급, 통행비용, 통신수요, 통신시설공급, 통신비용, 토지이용, 경제활동 등의 8개 내생변수와 사회인구를 1개 외생변수를 설정하여 통신이 통행에 미치는 영향을 분석하였고, 신동호(2002)는 통행행태(통행수/연계통행)와 통행시간수요, 가정내활동 및 가정외활동을 내생변수로, 사회경제적특

성을 외생변수로 설정하여 구조모형을 수립하였다.

Xuedong Lu(1999) 역시 나이, 성별, 수입, 소유차량 대수 등의 지표가 가정내와 가정외에서의 목적별 활동에 미치는 영향을 구조방정식을 이용하여 분석하였다.

### III. 구조방정식 모형의 기법

구조방정식 모형은 사회학 및 심리학에서 개발된 측정이론에 토대를 둔 확인적 요인분석과 계량경제학에서 개발된 연립방정식 모형에 토대를 둔 다중회귀분석 및 경로분석 등이 결합된 성격을 갖는 방법론이라고 할 수 있다. 구조방정식 모형에서는 관측변수와 잠재변수를 구분하여 분석하는데, 이는 관측변수는 측정오차에 의해 오염되어 있지만 잠재변수는 측정오차에 의해 오염되어 있지 않은 순수한 성격을 가지고 있어 이들 간의 관계를 분석하는 것이 올바른 계수를 추정할 수 있기 때문이다. 구조방정식 모형은 공분산행렬(상관행렬)이 분석의 대상이 되며, 일련의 변수사이의 공분산행렬(상관행렬)을 토대로 연구자가 수립한 모형이 가능한 많은 분산을 설명해 줄 수 있는가를 확인하고자 한다. 즉, 개별 사례에 초점을 맞춘 통계 기법이 아니라, 자료에서 얻어진 표본공분산행렬과 모델예측행렬 간의 차이를 가능한 한 적게 하고자 하는 분석법이다(배병렬, 2005). 구조방정식 모형은 연구자가 설정한 인과관계에 대한 모형을 검증하기에 가장 적합한 분석기법으로 기존의 회귀분석, 분산분석 또는 경로분석과는 달리 모델 내에 내재되어 있는 측정오차를 알 수 있으며, 이론변수와 측정변수 사이의 관계를 검증가능한 것이 가장 큰 장점이라 할 수 있다(조선배, 1996).

구조방정식을 추정하는 프로그램에는 LISREL, AMOS, EQS 등이 있으나, 본 연구에서는 LISREL을 이용하여 모형을 추정하였다. 일반적인 구조모형은 측정모델과 구조모델로 구성되어 있다. 측정모델(measurement model)은 관측변수와 잠재변수들 간의 관계를 나타내는 것으로 각각의 관측변수가 특정 잠재변수에 대해 어떻게 적재되어 있는가를 설정한 것이며, 구조모델(structural model)은 잠재변수들 간의 관계를 설정한 것으로 잠재변수간 직·간접적인 영향을 설명한다.

일반적인 구조방정식 모델은 다음의 세 가지 방정식으로 구성할 수 있으며, <그림 1>과 같은 형태로 나타낼 수 있다. <표 1>은 구조방정식의 구성요소를 설명한 것이다.<sup>1)</sup>

**x 측정모델**

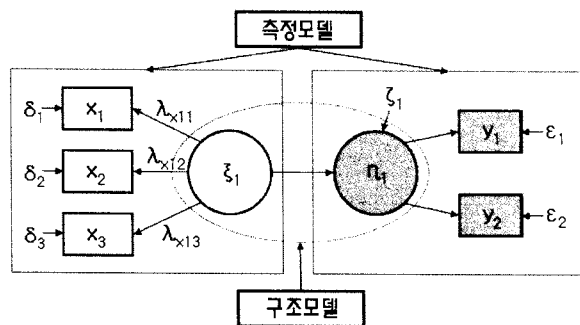
$$x_{(q \times 1)} = \lambda_{x(q \times n)} \times \xi_{(n \times 1)} + \delta_{(q \times 1)}$$

**y 측정모델**

$$y_{(p \times 1)} = \lambda_{y(p \times m)} \times \eta_{(m \times 1)} + \varepsilon_{(p \times 1)}$$

**구조방정식모델**

$$\eta_{(m \times 1)} = B_{(m \times m)} \times \eta_{(m \times 1)} + \Gamma_{(m \times n)} \times \xi_{(n \times 1)} + \zeta_{(m \times 1)}$$



<그림 1> 구조방정식 모델

<표 1> 구조방정식의 구성요소

측정 모델	x	외생관측변수들의 q×1 벡터
	y	내생관측변수들의 p×1 벡터
	ξ	잠재외생변수들의 n×1 벡터
	η	잠재내생변수들의 m×1 벡터
	δ	x 측정오차들의 q×1 벡터
	ε	y 측정오차들의 p×1 벡터
구조 모델	λ <sub>x</sub>	n 외생잠재변수를 측정하도록 설계된 q 관측변수와 관련된 q×n 회귀행렬
	λ <sub>y</sub>	m 내생잠재변수를 측정하도록 설계된 p 관측변수와 관련된 p×m 회귀행렬
	Γ	n 외생잠재변수와 m 내생잠재변수를 연결해주는 m×n 행렬계수
	B	m 내생잠재변수를 연결해주는 m×m 행렬계수
	ζ	방정식 오차를 의미하는 m×1 벡터

측정모델은 잠재변수 및 측정오차의 평균은 0이고, ε은 η, ξ 및 δ와 상관이 없으며, δ는 ξ, η 및 ε과 상관이 없다는 통계적 가정을 하고

1) 구조방정식에 관한 설명은 「LISREL 구조방정식 모델, 배병렬 저, 2005년」의 내용을 바탕으로 작성되었음.

있다. 또한 구조모델은 잠재변수 및 잔차의 평균은 0이며, ζ는 ξ와 상관이 없고, B행렬은 대각에 0의 값을 가지며 (I-B)는 정칙(non-singular) 행렬이라는 통계적 가정을 하고 있다.

구조방정식 모형의 모수를 추정하는 기법에는 최우추정법(maximum likelihood), 일반화최소자승법(generalized least square), 가중최소자승법(weighted least square), 2단계최소자승법(two-stage least square) 등이 있으며, 본 연구에서는 최우추정법을 이용하였다. 최우추정법에 이용되는 목적함수는 다음과 같다.

$$F_{ML} = \log \left\| \hat{\Sigma} \right\| + \text{tr}(S \hat{\Sigma}^{-1}) - \log \| S \| - (p+q)$$

여기서  $\|A\|$ 는 행렬 A의 determinant를 의미하며,  $\text{tr}(A)$ 는 행렬 A의 트레이스(trace)로 주대각원소의 합을 나타낸다. p와 q는 각각 내생잠재변수와 외생잠재변수의 개수를 의미한다.

## IV. 모형의 추정

### 1. 모형의 구조 및 변수 설정

본 연구에서는 사고데이터의 관측변수들을 크게 사고 심각도와 도로 요인, 운전자 요인, 환경 요인 등으로 분류하고, 이를 잠재변수로 설정하여 구조모형을 추정하였다. 도로 요인의 경우 평면선형과 종단구배를 관측변수로 설정하였으며, 평면선형과 종단구배가 커질수록 사고 발생확률이 증가하고 사고의 심각도가 증가하는 것으로 가정하였다. 운전자 요인의 경우 사고의 원인차 차종과 운전자 연령대를 관측변수로 설정하였으며, 원인차 차량의 크기가 커질수록, 운전자의 연령대가 높을수록 사고의 심각도가 증가할 것이라 가정하였다. 환경 요인은 노면상태와 날씨, 주·야간을 관측변수로 설정하였으며, 노면상태는 건조할 때보다는 습윤할 때가, 날씨는 맑음, 흐림, 비, 안개, 눈의 순으로, 주간보다는 야간에 사고심각도가 증가할 것으로 가정하여 분석을 실시하였다. 사고 심각도 잠재변수를 나타내는 관측변수로는 사고처리시간과 차량과손대수, 사망자수, 부상자수, 사고차량수 등 5개 변수를 이용하였다. 4개의 잠재변수 가운데 사고 심각도는 내생변수이며, 도로 요인과 운전자 요인, 환경 요인 등 3개 변수는 외생변수로 설정하여 모형을 구축하였다.

<표 2> 모형에 적용된 변수

구분	관측변수	설명 (괄호 안은 단위)	
내생변수	사고처리시간	발생한 사고가 처리되기까지 소요시간 (분)	
	차량파손대수	사고로 인해 파손된 차량대수 (대)	
	사망자수	사고로 인한 사망자수 (인)	
	부상자수	사고로 인한 부상자수 (인)	
	사고차량수	사고에 연관된 차량대수 (대)	
외생변수	도로요인	평면선형	- 직선, 커브 500m 이상, 커브 500m 이하
		종단구배	- 평탄, 오르막, 내리막
	운전자요인	원인차차종	- 승용, 승합, 화물, 트레일러
		운전자연령대	- 20세 미만, 20대, 30대, 40대, 50대, 60세 이상
	환경요인	노면상태	- 건조, 습윤
		날씨	- 맑음, 흐림, 비, 안개, 눈
		주야	- 주간, 야간

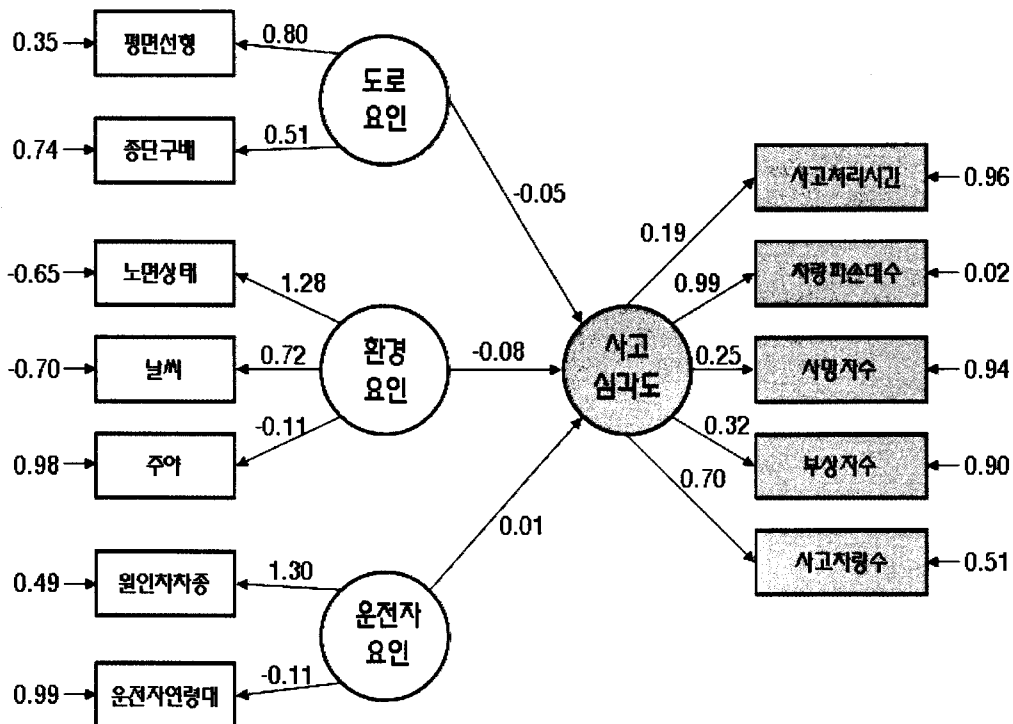
2. 모형의 추정 및 결과

모형의 추정을 위해 본 연구에서는 각각의 관측변수들의 상관행렬(correlation matrix)을 이용하였으며, 총 2,880개의 샘플을 가지고 분

석을 실시하였다. 추정된 모형은 다음의 <그림 2>와 같다.

본 연구에서는 최종적으로 도로 요인과 환경 요인, 운전자 요인을 외생잠재변수로, 사고 심각도를 내생잠재변수로 설정하여 구조방정식 모형을 구축하였다. 모형의 구조상 잠재변수들 간의 효과는 직접효과만이 도출되었으며, 간접 효과는 산출되지 않았다.

추정된 모형에 의하면, 도로 요인이 사고 심각도에 미치는 총효과는 -0.05(표준오차 0.024, t-value -2.174)로 계수의 부호가 음(-)의 값을 가지며, 이는 도로 요인이 증가할수록 사고의 심각도는 낮아지는 것을 의미한다. 또한 환경 요인이 사고 심각도에 미치는 총효과는 -0.08(표준오차 0.015, t-value -5.240)으로 환경 요인이 증가할수록 사고의 심각도는 낮아지는 것으로 분석되었다. 운전자 요인의 경우에는 사고 심각도에 미치는 총효과가 0.015(표준오차 0.020, t-value 0.742)로 운전자 요인이 증가할수록 사고의 심각도가 높아지는 것으로 나타났다. 추정된 모수들은 모두 표준화 해(standardized solution)이며, 이를 통해 외생잠재변수들이 내생잠재변수에 미치는 상대적 영향력을 파악할 수 있다. 본 연구의 모형에서는 환경 요인이 도로 요인이나 운전자 요인에 비해 사고 심각도에 더 많은 영향을 미치는 것으로 분석되었다.



<그림 2> 추정된 구조방정식 모형

모형에 의해 추정된 모수들 및 t-value를 정리하면 다음과 같다.

<표 3> LAMBDA-Y

관측변수	$\lambda_y$ (t-value)
y1 (사고처리시간)	0.193
y2 (차량파손대수)	0.988 (9.749)
y3 (사망자수)	0.250 (8.308)
y4 (부상자수)	0.317 (8.996)
y5 (사고차량수)	0.699 (10.237)

<표 4> LAMBDA-X

구분	$\lambda_x$ (t-value)		
	도로요인	환경요인	운전자요인
x1 (평면선형)	0.805 (16.053)	-	-
x2 (종단구배)	0.507 (14.595)	-	-
x3 (노면상태)	-	1.283 (28.340)	-
x4 (날씨)	-	0.717 (24.198)	-
x5 (주야)	-	-0.114 (-7.844)	-
x6 (원인차차종)	-	-	1.304 (1.130)
x7 (운전자연령대)	-	-	0.146 (1.119)

<표 5> GAMMA (괄호 안은 t-value)

구분	도로요인	환경요인	운전자요인
사고심각도	-0.052 (-2.174)	-0.081 (-5.240)	0.015 (0.742)

<표 6> THETA-DELTA

관측변수	$\delta$ (t-value)
x1 (평면선형)	0.353 (4.558)
x2 (종단구배)	0.743 (20.497)
x3 (노면상태)	-0.647 (-5.591)
x4 (날씨)	0.486 (12.791)
x5 (주야)	0.987 (38.137)
x6 (원인차차종)	-0.701 (-0.233)
x7 (운전자연령대)	0.979 (21.321)

<표 7> THETA-EPSILON

관측변수	$\epsilon$ (t-value)
y1 (사고처리시간)	0.963 (37.851)
y2 (차량파손대수)	0.024 (0.653)
y3 (사망자수)	0.937 (37.726)
y4 (부상자수)	0.900 (37.415)
y5 (사고차량수)	0.511 (22.375)

본 연구모형이 주어진 자료에 전반적으로 잘 부합하는지를 판단하기 위한 적합도 지수는 다음의 <표 8>에 제시하였다. 카이제곱 통계량은 표본 공분산행렬(S)과 모형에 의해 표현되는 추정 공분산행렬( $\Sigma$ )와의 불일치정도를 평가하기 위한 지표로 본 모형에서는 649.831(P=0.00)으로 나타나 두 행렬간의 차이가 없다는 귀무가설을 기각하게 된다. 카이제곱 통계량은 관측변수들의 다변량 정규분포를 가정하고 있으며 표본 크기에 민감하게 반응하는 경향이 있다고 알려져 있다. 따라서 카이제곱 통계량 이외에 적합지수(GFI)나 조정적합지수(AGFI), 비교적합지수(CFI), 등의 적합도지수를 함께 고려하여 모형의 적합여부를 판단하는 것이 바람직한 것으로 알려져 있다.

<표 8> 사고심각도 모형의 적합도 지수

적합도 지수	값
카이제곱 통계량	649.831 (p=0.00)
적합지수 (GFI)	0.966 (0.9 이상)
조정적합지수 (AGFI)	0.944 (0.9 이상)
비교적합지수 (CFI)	0.936 (0.9 이상)
잔차평균자승 이중근 (RMR)	0.050 (0.05 이하)
근사오차평균자승의 이중근 (RMSEA)	0.0639 (0.05 이하)
Critical N (CN)	327.444 (200 이상)

주) 카이제곱 통계량을 제외하고 나머지 지수들의 권장 수 용치는 괄호 안에 표시하였음

적합지수(GFI)는 일반적으로 0~1사이의 값을 가지며 모형이 표본공분산행렬을 설명하는 비율을 나타내는 지표로 회귀분석에서의 R<sup>2</sup>값과 비슷하게 해석된다. 본 연구모형의 적합지수는 0.966으로 나타나, 표본의 대부분이 개발된 모형에 의해 설명될 수 있는 것으로 판단된다.

자유도에 의해 적합지수를 수정한 조정적합지수 AGFI는 0.944로 권장 수용수준인 0.90 이상인 것으로 나타났으며, 귀무가설이 반드시 진실하다는 가정을 보완하기 위해 비중심 카이제곱 분포를 적용하여 산출하는 비교적합지수 CFI값도 0.936으로 나타났다. 또한 S와  $\Sigma$ 와의 차이를 이용하여 산출하는 RMR(root mean square residual) 값은 0.0501, RMSEA(root mean square error of approximation) 값은 0.0639로 비교적 좋은 적합도를 가지는 것으로 나타났다. RMSEA 지표는 표본의 크기가 상당히 큰 제안 모델을 기각시키는 카이제곱 통계량의 한계를 극복하기 위한 적합지수이다(Browne and Cudeck, 1993). CN(critical N)은 카이제곱 검정을 위해 주어진 유의수준에서 구조방정식 모형이 수용할 수 있는 적절한 표본크기를 나타내 주는 지수로, 200이상이면 모형이 표본을 잘 설명하는 것으로 해석할 수 있으며 본 모형의 CN 값은 327.444로 나타났다.

#### IV. 결론 및 향후 연구과제

본 연구에서는 고속도로의 사고심각도에 영향을 미치는 요인들을 분석하고자 사고심각도를 구조모형의 내생잠재변수로, 도로 요인 및 운전자 요인, 환경 요인을 외생잠재변수로 설정하여 모형을 추정하였다. 잠재변수들을 설명하기 위한 관측변수에는 사고처리시간, 차량파손대수, 사망자수, 부상자수, 사고차량수, 평면선형, 종단구배, 원인차 차종, 운전자 연령대, 노면상태, 날씨, 주야 등 총 12개 변수가 포함되었으며, 2005년의 고속도로 사고데이터 2,880개를 표본으로 이용하였다. 모형의 추정결과 도로 요인과 환경 요인은 사고 심각도와 음(-)의 관계에 있는 것으로 나타났으며, 운전자 요인은 사고 심각도와 양(+)의 관계에 있는 것으로 나타났다.

사고는 매우 다양한 요인들이 복합적으로 작용하여 발생하기 때문에 예측을 하는데 어려움이 있다. 본 연구에서 적용한 도로나 운전자, 환경 요인의 관측변수들을 더욱 다양화하고 새로운 외생 또는 내생잠재변수를 추가하여 모형의 구조를 발전시킨다면 변수들간의 영향관계를 명확히 하고 사고를 예측하는 데 도움이 될 것으로 생각된다.

#### 참고문헌

1. 김태영(2007), “트럼펫 IC 램프의 운전조건과 교통사고 분석”, 대한교통학회지 제25권 제1호, pp. 73~80
2. 박정순(2007), “도로환경요인과 교통사고의 상관분석 및 사고추정모형 개발”, 대한교통학회지 제25권 제2호, pp. 63~72
3. 박호신(2007), “고속도로 인터체인지 연결로에서의 교통사고 예측모형 개발”, 대한교통학회지 제25권 제3호, pp. 123~136
4. 배병렬(2005), “LISREL 구조방정식모델 - 이해, 활용 및 프로그래밍 -”, 도서출판 청람
5. 윤병조(2006), “트럼펫 IC형식 연결로 교통사고 특성분석에 관한 연구”, 대한교통학회지 제24권 제7호, pp. 41~51
6. 조선배(1996), “LISREL 구조방정식모델”, 영지문화사
7. 신동호(2002), “구조방정식모형을 이용한 활동 및 통행패턴분석에 관한 연구”, 한국지역개발학회지 제14권 제1호, pp. 125~146
8. 추상호(2006), “구조방정식모형을 이용한 통신이 통행에 미치는 영향 분석”, 대한교통학회지 제24권 제3호, pp. 157~165
9. Xuedong Lu(1999), “Socio-demographics, activity participation and travel behavior”, Transportation Research Part A 33, pp. 1~18
10. Thomas F. Golob(1996), “A model of a activity participation and travel interactions between household heads”, Transportation Research Part B 31, pp. 177~194
11. Jin-Hyuk Chung(2002), “Structural model of automobile demand in Korea”, Transportation Research Record 1807, pp. 87~91
12. Jin-Hyuk Chung(2002), “Structural equation models of day-to-day activity participation and travel behavior in a developing country”, Transportation Research Record 1807, pp. 109~118