

고속도로 연결로의 교통사고예측모형 개발

Traffic Crash Prediction Models for Expressway Ramps

최윤환 Choi, Yoon-Hwan 오영태 Oh, Young Tae 최기주 Choi, Keechoo 이철기 Lee, Choul Ki 윤일수 Yun, Ilsoo	한국도로공사 사업본부장 (E-mail: choiyh@ex.co.kr) 아주대학교 교통시스템공학과 교수 (E-mail: ytoh@ajou.ac.kr) 정회원 · 아주대학교 교통시스템공학과 교수 (E-mail: keechoo@ajou.ac.kr) 아주대학교 교통시스템공학과 부교수 (E-mail: cklee@ajou.ac.kr) 정회원 · 아주대학교 교통시스템공학과 조교수 · 교신저자 (E-mail: ilsooyun@ajou.ac.kr)
--	---

ABSTRACT

PURPOSES : Using the collected data for crash, traffic volume, and design elements on ramps between 2007 and 2009, this research effort was initiated to develop traffic crash prediction models for expressway ramps.

METHODS : Three negative binomial regression models and three zero-inflated negative binomial regression models were developed for individual ramp types, including direct, semi-direct and loop, respectively. For validating the developed models, authors compared the estimated crash frequencies with actual crash frequencies of twelve randomly selected interchanges, the ramps of which have not been used for model developing.

RESULTS : The results show that the negative binomial regression models for direct, semi-direct and loop ramps showed 60.3%, 63.8% and 48.7% error rates on average whereas the zero-inflated negative binomial regression models showed 82.1%, 120.4% and 57.3%, respectively.

CONCLUSIONS : Conclusively, the negative binomial regression models worked better in traffic crash prediction than the zero-inflated negative binomial regression models for estimating the frequency of traffic accidents on expressway ramps.

Keywords

crash prediction, Poisson regression model, negative binomial regression model, zero-inflated negative binomial regression model, expressway ramp

Main Author : Choi, Yoon-Hwan, Executive Director
 Korea Expressway Corporation, 805, Daewang Pangyoro, Sujeong-gu,
 Seongnam-si 461-703, Korea
 Tel : +82.2.2230.4206 Fax : +82.2.2230.4105
 email : choiyh@ex.co.kr

International Journal of Highway Engineering
<http://www.ksre.or.kr/>
 ISSN 1738-7159 (Print)
 ISSN 2287-3678 (Online)

1. 서론

1.1. 연구의 배경 및 목적

2010년 말 우리나라 주요 간선도로망인 고속도로는 31개 노선 3,579km에 이르고, 일평균 377만대의 차량이 고속도로를 이용하고 있다(한국도로공사, 2011). 2010년 한 해 동안 고속도로에서는 총 3,924건의 교통

사고가 발생하였으며, 그로 인해 총 389명이 사망하고, 9,669명이 부상을 당하였다(경찰청, 2011). 또한 지난 3년(2007~2009년)간 고속도로에서 발생한 구간별 사고현황을 보면, 본선, 연결로, 영업소 그리고 기타 구간에서 각각 74.5%, 11.0%, 6.7%, 7.8%로 나타났다.¹⁾ 분석 결과, 고속도로 본선 연장과 연결로 연장 비율을 감

안할 때, 연결로 구간은 고속도로 본선 등에 비해 교통사고가 상대적으로 많이 발생함을 알 수 있다.

고속도로 나들목/분기점 상의 연결로는 본선을 포함한 다른 도로시설에 비해 평면 및 종단선형 등의 도로선형 요소들이 복잡하게 변화하는 구간이며, 이러한 복잡한 기하구조 변화가 운전자의 실수를 유발하여 상대적으로 많은 교통사고가 발생하고 있는 상황이다. 따라서 고속도로 연결로상의 교통사고에 대한 보다 철저한 연구를 통해 연결로 교통안전성을 개선해야 할 필요가 있다.

한국도로공사는 고속도로의 안전성 향상을 위해 시설 개량, 안전시설 확충, 도로안전진단, 사고다발지점 개선 사업 등 다양한 사업을 시행하고 있으나, 교통사고에 대한 사회적 관심이 점점 더 높아지고 있는 현실을 감안한다면, 이러한 사후적 조치 외에도 예방적 차원의 도로안전성을 높이기 위한 안전 사업을 시행해야 할 시점에 와 있다.

이에, 본 연구에서는 고속도로 나들목/분기점 상의 연결로에서 발생하는 교통사고에 대하여 곡선반경, 구배 등 도로선형 요소를 기반으로 하는 교통사고예측모형을 개발하고자 한다. 즉, 교통사고발생 가능성이 높은 도로선형요소들을 사전에 파악하여, 이를 설계기준 개선에 활용되도록 하고, 개통 후 운영 및 관리단계에서는 사고다발지점의 원인분석 및 개선안 도출을 위한 기초자료로 활용할 수 있도록 연결로 설계요소를 중심으로 고속도로 연결로 교통사고예측모형을 개발하는 것이 목적이다.

1.2. 연구의 범위 및 방법

연구의 시간적 범위는 2007년부터 2009년까지 3년간이며, 공간적 범위는 동 기간 동안 공용 중이었던 전국 고속도로를 포함한다. 본 연구에서는 첫 번째 단계로 연구관련 고찰을 통하여 전통적인 교통사고예측방법론에 대한 이론적 고찰을 실시하였다. 두 번째 단계로 연결로를 크게 3가지 유형(루프, 준직결, 직결)으로 분류한 후 각 유형에 대한 교통사고자료 및 도로선형자료를 수집하였다. 세 번째 단계로 수집된 도로선형 자료의 교통사고와의 상관계수를 분석하고 영향력이 존재하는 변수(도로선형요소)를 추출하였다. 선정된 도로선형요소에 대하여 비선형 회귀분석을 적용하여 교통사고예측모형을 개발하였다. 마지막으로 통계적인 기법을 이용하여 개발된 모형을 검증하였다.

2. 연구관련 고찰

2.1. 선행연구 고찰

성낙문(2002)은 미국 미시간주의 고속도로 교통사고 및 교통량 데이터베이스를 이용하여 다이아몬드형 인터체인지와 클로버형 인터체인지를 대상으로 교통사고예측모형을 구축하였다. 모형구축에는 음이항회귀모형(negative binomial regression model)을 이용하였으며 다이아몬드형 인터체인지는 연결로 교통량, 트럭비율, 인터체인지 길이, 직결연결로 길이, 본선 길어깨 폭이 설명변수로 선택되었고 부분적 클로버형 인터체인지는 연결로 교통량, 트럭비율, 인터체인지 길이가 설명변수로 선택되었다.

강정규와 이성관(2002)은 호남고속도로를 직선부, 곡선부, 완화곡선부로 나누고 각각에 대해 본선구간과 유·출입구간의 교통사고예측모형을 개발하였다. 모형개발에는 다중선형회귀모형을 이용하였으며 도로 길이와 AADT가 6개 모형의 공통된 변수로 선정되었다.

박효신 등(2007)은 2003년~2005년까지의 고속도로 교통사고 속보자료를 이용해 트럼펫 인터체인지 31개를 대상으로 교통사고예측모형을 개발하였다. 음이항회귀모형을 통해 전체 연결로, 직결 연결로, 준직결 연결로 및 루프 연결로 등 총 4개의 모형을 개발하였으며 유의한 변수로는 유·출입여부, 교통량, 중차량 비율, 곡률 등이 선택되었다.

김봉곤(2007)은 2003년에서 2005년까지 고속도로 인터체인지 유·출입부에서 발생한 교통사고를 바탕으로 교통사고 판별모형을 정립하였다. 유출 및 유입 연결로에 대한 판별모형은 곡선 길이, 종단경사, 횡단경사, 가감속차로 길이, 길어깨 폭이 변수로 선정되었고, 유출 연결로에 대한 판별모형은 곡선 길이, 종단경사, 횡단경사가 유의한 변수로 선정되었다.

박병호 등(2008)은 2004년에 청주시에서 발생한 교통사고자료를 사용하여 비신호교차로 사고모형을 개발하였다. 3지교차로는 포아송회귀모형이, 4지교차로는 영과잉포아송회귀모형이 적합한 것으로 분석되었다. 두 모형에서 공통적으로 사고에 영향을 미치는 변수로는 교통량, 횡단보도수, 교차각, 시거, 차로폭으로 나타났다. 특히 이 논문에서는 Vuong 통계값을 이용하여 포아송회귀모형 및 음이항회귀모형과 각 모형에 대한 영과잉회귀모형을 선택하는 과정을 잘 설명하고 있다.

윤병조(2008)는 1999년~2003년 간 트럼펫형 나들목(interchange; IC) 99개소에서 발생한 교통사고자료를 바탕으로 연결로형태 별 교통사고예측모형을 개발하

1) 한국도로공사 2007~2009년(3년간)「교통사고속보자료」를 기준으로 한 고속도로 사고위치별 분석결과임(최윤환, 2011).

였다. 직결형, 준직결형, 루프형으로 연결로형태를 구분하였으며 다중선형회귀모형을 이용하였다. 세 연결로형태에서 공통으로 선택된 변수는 곡선반경과 일평균교통량으로 선정되었다.

노창균 등(2008)은 2002년부터 2005년까지 고속도로의 트럼펫인터체인지와 클로버인터체인지에서 발생한 사고에 대해 음이항회귀모형을 이용하여 교통사고 예측모형을 수립하였다. 트럼펫인터체인지의 교통사고 예측모형은 곡률, 속도차, 유·출입여부, 교통량이 변수로 포함되었으며 클로버인터체인지 또한 동일한 변수가 선택되었다.

Shankar et al.(1995)는 미국 시애틀의 도시고속도로 48km 구간에 대해 도로 기하구조 및 환경요소가 사고에 미치는 영향을 음이항회귀모형을 통해 도출하였다. 또한, 이 연구에서는 곡선구간의 수, 최소곡선반경, 도로구간의 최대경사 등의 도로 기하구조요소를 모형식에 적용하였다.

Bauer와 Harwood(1997)는 미국 워싱턴 주의 200개의 연결로를 대상으로 교통사고와 기하구조와의 관계를 모형화하였다. 1993년부터 1995년에 발생한 교통사고를 기반으로 포아송회귀모형 및 음이항회귀모형을 이용하여 사고등급 별, 지역 별 교통사고예측모형을 개발하였다. 인터체인지 연결로 교통사고에 영향을 주는 주요 인자가 교통량, 연결로의 형식, 가감속차로 길이, 연결로 길이, 유·출입여부라고 결론을 제시하였다.

기존 선행연구를 고찰한 결과, 사고예측모형 개발을 위해 다양한 도로선형요소를 통해 활용하였으며, 초기 모형개발에는 다중선형모형을 통한 연구를 중심으로 이루어졌으며 점차 교통사고는 임의적이고 산발적으로 발생하는 과분산 현상으로 인해 음이항분포를 적용한 연구로 발전하였음을 알 수 있다.

본 연구에서는 기존 연구에서 주로 적용한 포아송회귀모형 및 음이항회귀모형을 고려함과 동시에 각 모형에 대한 영과잉회귀모형을 함께 고려하고 또한 그 결과를 Vuong 통계값 및 검정(validation)을 이용하여 비교하였다. 또한 교통사고 예측모형식에 교통사고노출계수(exposure)를 포함시켜 예측의 정확도를 높이고자 시도하였다. 이러한 시도는 Interactive Highway Safety Design Model(IHSDM)과 Highway Safety Manual(HSM)에서 사용된 기본모형에서 사용된 방식이다. IHSDM와 HSM에서는 다른 기하구조 요인보다 지대하게 교통사고에 영향을 줄 수 있는 구간 길이 및 AADT 등을 옵션변수로 보고 구간 거리 및 AADT를 포

함하는 방정식을 예측모형식에 곱하는 형태로 삽입하여 해당구간의 특성을 반영하는 방법을 사용하고 있다(AASHTO, 2010; FHWA, 2011).

2.2. 이론적 고찰

2.2.1. 교통사고예측모형

교통사고예측모형 개발을 위하여 초기에는 선형 또는 비선형 다중회귀모형(multiple regression model)이 빈번하게 사용되었다. 하지만 교통사고는 불연속적이며 의도하지 않고 산발적인 형태로 발생하기 때문에 다중회귀모형을 이용할 경우 분석에 한계가 존재한다(박효신 등, 2007). 따라서 교통사고의 이산적 특성(discrete event)반영을 위한 포아송회귀모형(Poisson regression model)을 도입하는 것이 적절하다(Washington et al., 2003). 그러나 포아송회귀모형은 기댓값과 분산값이 동일하다는 조건으로 교통사고의 과분산 현상(over-dispersion)을 고려할 수 없다. 즉, 포아송회귀모형의 주요 특징은 조건부 평균이 커질수록 분포의 밀도가 오른쪽으로 이동하며, 평균과 분산이 같은 등분산(equi-dispersion)의 가정이 있다는 것이다. 그러나 현실에서의 교통사고 자료는 분산이 평균보다 큰 과분산 문제가 발생하게 된다. 이러한 포아송회귀모형의 과분산 문제를 해결하기 위하여 음이항회귀모형(negative binomial regression model)이 적용될 수 있다(황경성, 2010; Washington et al., 2003). 즉, 대부분의 교통사고 분포는 평균보다 분산이 큰 과분산 현상을 보이는 경향이 있기 때문에 이를 해결할 수 있는 음이항분포를 적용한다. 예를 들어, Eq. (4)에서 음이항 분포의 과분산계수 α 의 값이 0이라면 평균과 분산이 같아져 포아송분포가 적합하다는 결론을 내릴 수 있지만 교통사고 건수의 경우 α 의 값이 0이 되는 경우가 많지 않기 때문에 음이항회귀모형을 적용하는 것이 더 적절하다(노창균 등, 2008).

또한 수집된 교통사고자료 중 교통사고가 0건인 지점의 자료 비율이 높을 경우 이를 해결하기 위해 영과잉포아송회귀모형(zero-inflated Poisson regression models) 또는 영과잉음이항회귀모형(zero-inflated negative binomial regression models)을 사용한다. 이때 영과잉회귀모형의 사용여부를 판단하기 위해 Vuong 통계값이 이용된다. 95%의 신뢰수준일 때 Vuong 통계값이 1.96보다 크면 영과잉회귀모형을 사용하는 것이 적합하며, -1.96보다 작으면 포아송회귀모형 또는 음이항회귀모형이 더 적합하다. -1.96과 1.96

사이에 있게 되면 두 모형간의 차이를 판단할 수가 없다 (박병호, 2008). 이러한 경우 두 모형 중에 단순한 모형을 선택하는 것이 일반적이다. Table 1은 상기 네가지 모형의 특징을 정리하고 있다.

Table 1. Characteristics of Regression Models

Regression Models	Characteristics
Poisson	<ul style="list-style-type: none"> - Most basic model - Applicable when traffic accident data meet the equi-dispersion condition(i.e., if α is equal to 0, or α is greater than 0 while the t-ratio α is smaller than 1.96)
Negative Binomial	<ul style="list-style-type: none"> - Applicable when traffic accident data violate the equi-dispersion condition(i.e., if α is equal to 0, or α is greater than 0 while the t-ratio α is greater than 1.96)
Zero-inflated Poisson	<ul style="list-style-type: none"> - Applicable when the Poisson regression model is not apt to explain the frequency of traffic accident data because the frequency of zero-traffic-accident location is high - Applicable when the value of Vuong statistic is greater than 1.96
Zero-inflated Negative Binomial	<ul style="list-style-type: none"> - Applicable when α is greater than 0 and the value of α is statistically significant (i.e., the t-ratio value of α is greater than 1.96) - Applicable when the value of Vuong statistic is greater than 1.96

2.2.2. 포아송회귀모형

포아송회귀모형에서 1년 동안 y_i 건의 사고가 발생한 연결로 i 에서 1년 동안 발생할 사고의 확률은 다음과 같다(Washington et al., 2003).

$$P(y_i) = \frac{\text{EXP}(-\lambda_i)\lambda_i^{y_i}}{y_i!} \quad (1)$$

$P(y_i)$ 는 1년 동안 y 건의 사고가 발생한 연결로 i 에서 1년 동안 발생할 사고의 확률이고 λ_i 는 연결로 i 의 포아송 파라미터로서 연결로 i 에서 1년 동안 기대되는 사고발생 건수(즉, $E[y_i]$)이다. 포아송회귀모형은 설명 변수의 함수인 포아송 파라미터 λ_i 에 의해 추정이 된다. 설명변수와 포아송 파라미터는 대부분 로그-선형(log-linear) 모형의 관계를 갖으며 다음과 같다.

$$\lambda_i = \text{EXP}(\beta X_i) \text{ 또는 } LN(\lambda_i) = \beta X_i \quad (2)$$

여기서 X_i 는 설명변수의 벡터이고 β 는 추정 파라미터이다. 이 식에서 일정 기간 동안 기대되는 사건 수는 $E[y_i] = \lambda_i = \text{EXP}(\beta X_i)$ 로 표현할 수 있다.

2.2.3. 음이항회귀모형

포아송분포를 이용한 분석 시 발생하는 대부분의 오차는 평균과 분산이 같다는 조건을 만족시키지 못하기 때문에 발생한다. 이러한 조건을 만족시키지 못하는 경우 저분산(under-dispersion) 또는 과분산(over-dispersion) 현상이 나타나다(Hilbe, 2008). 이러한 경우 파라미터 추정 시 한쪽으로 치우친(biased) 결과도 출되어 정확한 분석이 어렵다. 이러한 오차를 극복하기 위해 음이항회귀모형이 적용되며, 음이항회귀모형을 기반으로 하는 연결로 i 의 1년 동안 기대되는 사고발생 건수는 λ_i 는 Eq. (3)과 같다(Washington et al., 2003).

$$\lambda_i = \text{EXP}(\beta X_i + \epsilon_i) \quad (3)$$

여기서, $\text{EXP}(\beta X_i + \epsilon_i)$ 는 평균이 1이고 분산이 α^2 인 감마분포를 갖는 오차(error term)이다. 이 오차로 인해 분산이 평균보다 커지게 되어 과분산 현상을 해결하게 된다.

$$\text{VAR}[y_i] = E[y_i][1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2 \quad (4)$$

위의 식을 살펴보면 과분산계수 α 가 0인 경우에 분산과 평균이 같아지는 것을 확인할 수 있다. 즉, 포아송회귀모형은 음이항회귀모형에서 α 가 0인 모형이며, α 에 의해 어떠한 모형을 선택할 것인지 달라진다. 음이항회귀모형에서 1년 동안 y_i 건의 사고가 발생한 연결로 i 에서 1년 동안 발생할 사고의 확률은 다음과 같다.

$$P(y_i) = \frac{\Gamma((1/\alpha) + y_i)}{\Gamma(1/\alpha)y_i!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i} \right)^{1/\alpha} \left(\frac{\lambda_i}{(1/\alpha) + \lambda_i} \right)^{y_i} \quad (5)$$

2.2.4. 영과잉음이항회귀모형

교통사고의 특성 상 교통사고의 확률이 존재함에도 불구하고 일정기간 동안 특정 구간에서 교통사고가 일어나지 않는 경우가 있다. 수집된 교통사고건수 분포에서 사고가 0건이 발생한 경우에 대한 비율이 높을 때는 일반적인 포아송회귀모형 또는 음이항회귀모형이 적절

하지 않다(강윤희, 2008). 이와 같이 교통사고자료에서 교통사고 건수가 ‘영(zero)’ 인 비율이 높은 경우에는 일반적인 포아송회귀모형 또는 음이항회귀모형보다는 영과잉회귀모형(zero-inflated regression model)을 적용하는 것이 보다 적합하다(Washington et al., 2003). 예를 들어, 영과잉음이항회귀모형은 음이항회귀모형의 한계점을 극복하기 위해 영인 부분은 로지스틱 분포를 따른다고 가정하며 영이 아닌 부분은 음이항 분포를 따른다고 가정한다. 따라서 과분산 현상 문제뿐만 아니라 영과잉 문제까지 동시에 해결할 수 있는 모형이다(최보미, 2011). $\mu_i = (1/\alpha) / [(1/\alpha) + \lambda_i]$ 일 때 영과잉음이항회귀모형에서 1년 동안 y_i 건의 사고가 발생한 연결로에서 1년 동안 발생할 사고의 확률은 다음의 식과 같다(Washington et al., 2003).

$$\begin{cases} y_i = 0 \text{ 일 때 } p_i + (1-p_i) \left[\frac{1/\alpha}{(1/\alpha) + \lambda_i} \right]^{1/\alpha} \\ y_i = y \text{ 일 때 } (1-p_i) \left[\frac{\Gamma((1/\alpha) + y) \mu_i^{1/\alpha} (1-\mu_i)^y}{\Gamma(1/\alpha) y!} \right] \end{cases} \quad (6)$$

3. 자료수집 및 분석방법론

3.1. 자료수집

본 연구에서는 고속도로 전 노선을 대상으로 2007~2009년까지 3년간 C급 이상²⁾의 교통사고가 발생한 총 491개 나들목/분기점 중에서 AADT, 도로선형 자료 등 자료수집이 가능한 56개 나들목/분기점을 대상으로 분석을 수행하였다. 나들목/분기점 상에는 크게 3가지 유형(직결형, 준직결형, 루프형)의 연결로가 혼재되어 있다. 따라서 최종적으로 직결형 연결로는 99개, 준직결형 연결로는 52개, 루프형 연결로는 50개로 총 201개 연결로의 AADT, 도로선형자료와 교통사고자료를 수집하였다. 참고로, 56개 나들목/분기점 중 무작위(random)로 선정된 48개 나들목/분기점은 교통사고예측모형 개발에 사용되었고, 나머지 12개 나들목/분기점은 모형 검증(validation)에 사용되었다.

또한, 나들목/분기점 내의 연결로의 형태 차이(Fig. 1 참조)를 고려하기 위하여 최종적으로 Table 2와 같은 도로선형 자료를 구축하였다. Table 2에 나와 있는 각종 도로선형요소를 추출하기 위해서는 한국도로공사가

2) 한국도로공사에서는 사고의 피해 정도에 따라 크게 A, B, C, D급으로 분류하고, 그 중 C급 이상의 사고는 인적·물적 피해가 발생한 사고를 나타내고 A급이 가장 피해 규모가 크다.



--- Direct — Semi-Direct - - Loop

Fig. 1 Ramp Types for Trumpet Interchanges

보유하고 있는 도면을 활용하였다. 하지만 한국도로공사의 HGIS 상에서는 대부분 평면도만 존재하여 자료수집에 한계가 발생하였으며 준공도면이 확보 가능한 56개 나들목/분기점을 대상 도로선형 자료를 정리하였음을 밝힌다. 직결형 연결로의 경우, 하나의 곡선반경만이 존재하므로 해당 연결로를 대표할 곡선반경 또는 곡률을 선정하는 데 문제가 없다. 하지만, 준직결형과 루프형 연결로의 경우, 두 개의 곡선반경을 존재하기 때문에 두 곡선반경 중 작은 값(이하, 최소곡선반경이라 칭함) 또는 곡률차를 모형개발에 사용하였다.

Table 2. Road Design Elements by Ramp Types

Ramp Types	Road Design Elements				
Direct	Length of curve	Radius	Grade	Length of acceleration/deceleration lane	Curvature
Semi-Direct	Length of curve	Minimum radius*	Grade	Length of acceleration/deceleration lane	Deviation of curvature
Loop					

* Minimum radius means the minimum radius between two radius involved in the ramp

3.2. 분석방법론

연결로 유형별로 도로선형자료(독립변수)와 교통사고 건수(종속변수)의 상관계수를 분석하고 다중공선성 분석, 모형식의 회귀계수 추정, 적합도 검증, 환류(feedback)의 과정을 통해 적정 변수를 선정하였다. 참고로, 각 독립변수의 규모효과(scale effect)를 줄이기 위하여 Table 3에서 보인 바와 같이 정규화(normalization)를 하였다.

마지막으로 통계적인 모형의 추정과 검정을 수행하여 도로선형에 기반을 둔 교통사고 예측모형 개발을 위한 흐름은 Fig. 2와 같다.

Table 3. Normalization of Road Design Elements

Type	Contents			
Design elements	Radius	Grade	Length of curve	Curvature (deviation)
Independent variable	Radius /100	Grade	Length of curve/100	Curvature (deviation)×100

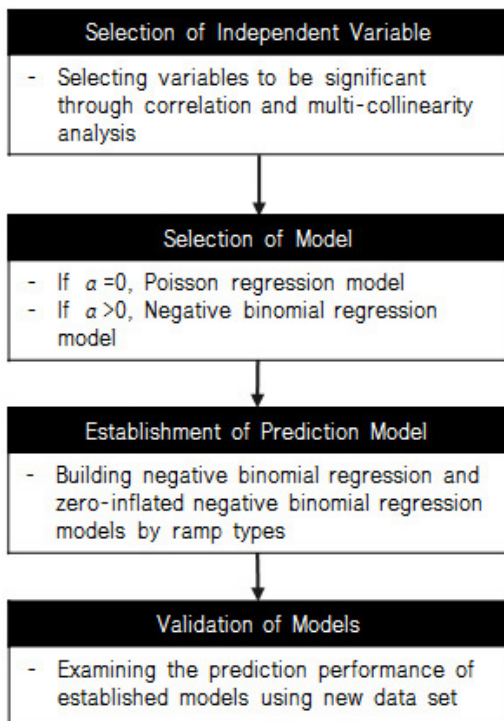


Fig. 2 Flow Chart for Modeling

4. 도로선형 기반 교통사고예측모형 개발

4.1. 변수선정

연결로 유형별(직결, 준직결, 루프) 도로선형요소(독립변수)들과 종속변수인 교통사고 건수와와의 상관관계³⁾는 Table 4와 같고, 상관관계의 설명력이 높은 선형변수를 변수선정에 우선적으로 고려하였다. 또한, 다중공선성 분석을 통해 독립변수들 간의 선형독립여부를 확인하고 모형식의 계수 추정을 통해 p-value의 유의미하게 설명 가능한 변수를 최종적으로 Table 5와 같이

3) 선형변수들과 종속변수인 교통사고와의 상관관계는 유의확률 90%에서 분석함.

선정하였다. 또한, 회귀분석 수행을 위해 다중공선성 존재여부의 판단이 필요하며, 이를 위해 분산팽창인자 (variance inflation factor; VIF)와 허용치 (tolerance limit)를 이용하였다(서혜선, 2009).

Table 4. Correlation Coefficient Analysis for Variable Selection

Type	Curvature	Radius	Grade	Length of acc./dec. lane
Direct	-0.040	0.202	-0.305	-0.047
Types	Curvature Deviation ⁴⁾	Minimum Radius	Grade	Length of acc./dec. lane
Semi-direct	-0.057	-0.138	-0.213	0.070
Loop	-0.131	-0.142	-0.174	-0.047

Table 5. Multicollinearity Analysis for Variable Section

Type	Radius	Grade	Length of acc./dec. lane
Direct	Tolerance Limit	0.990	0.990
	VIF	1.010	1.010
Types	Minimum Radius	Grade	Length of acc./dec. lane
Semi-direct	Tolerance Limit	0.857	0.753
	VIF	1.166	1.328
Loop	Tolerance Limit	0.977	0.993
	VIF	1.024	1.007

$$VIF_i = \frac{1}{1 - R_i^2} \quad (7)$$

여기서,

VIF_i : i 번째 독립변수에 대한 분산팽창인자 (variance inflation factor)

$1 - R_i^2$: 공차한계 또는 허용치(tolerance)

R_i : i 번째 독립변수와 종속변수에 대한 상관관계수

Eq. (7)에서는 i 번째 독립변수가 나머지 독립변수들과 상관관계가 없다($R_i^2=0$)면 VIF_i 와 허용치는 1이 되고, i 번째 독립변수가 나머지 독립변수들과 상관관계가 있다($R_i^2=1$)면 $VIF_i=\infty$ 이고 허용치는 0이 될 것이다. 또한, 일반적으로 가장 큰 VIF_i 값이 10을 넘으면 다중공

4) 곡률은 1/곡선반경(r)을 의미하지만 매우 작은 값임. 곡률차는 차량의 진행방향에 대해 1번째 주행하는 곡선의 곡선반경(r_1)의 역수에서 2번째 주행하는 곡선의 곡선반경(r_2)의 역수를 뺀 값을 말함.

선성의 존재를 의심하게 된다. 따라서 n개의 허용치 중에서 가장 작은 값이 0.1 이하일 때도 다중공선성이 있다고 판단한다. 최종적으로 선정된 선형변수들의 다중공선성 분석 결과, Table 5에서 보인 바와 같이 변수들 간 다중공선성은 존재하지 않는 것으로 분석되었다.

변수선정 결과, 최종적으로 선정된 변수는 직결형, 준직결형, 루프형 모두 곡선반경, 종단경사, 가감속차로 길이로 선정되었다. 곡률과 곡선반경은 서로 반비례의 관계를 갖는 변수이므로 전체적으로 상관관계가 높은 곡선반경을 선정하였다. 종단경사는 전체적으로 상관관계가 가장 높았기 때문에 선정하였다. 가감속차로 길이는 세 가지 램프형태에서 모두 상관관계가 낮고 부호 또한 일정치 않았다. 하지만 이는 각 램프의 교통사고 노출 정도(exposure)에 대한 영향을 보정하지 않은 것이다. 따라서 교통사고 노출 정도를 고려한 최종 모형식에서 변수의 적용여부를 판단하기로 하였다.

4.2. 모형 개발

교통안전성을 평가하는데 있어 분석대상구간을 어떻게 설정하는지 여부가 결과에 많은 영향을 미칠 수 있다. 따라서 공평한 교통안전성 평가를 위하여 분석대상구간은 기하구조측면과 교통측면에서 유사한 구간으로 선정하는 것이 필요하다(최윤환, 2011). 따라서 일반적으로 교통사고예측모형식 내에서 이러한 유사성을 보장하기 위하여 대표적인 도로 설계요소인 구간의 길이와 교통요소인 교통량을 교통사고노출계수(exposure)란 형태로 모형식에 포함시키고 있다(Harwood, et al., 2005; AASHTO, 2010; FHWA, 2011). 본 연구에서 사용한 교통사고노출계수(exposure)는 다음 식과 같다.

$$\begin{aligned} EXPO(\text{exposure}) \\ = AADT_i \times L_i \times 365 \text{일} \times \text{year} \times 10^{-6} \end{aligned} \quad (8)$$

여기서,

$AADT_i$ = 교통사고지수예측 대상 i 의 AADT (대/일)

L_i = 교통사고지수예측 대상 i 의 연장 (m)

$year$ = 분석기간, 본 연구에서는 3년

연결로의 유형별로 상관분석 결과를 이용하여 교통사고에 영향을 미치는 도로선형변수들에 변수선정에 우선적으로 고려하고, 다중공선성 분석을 통해 도로선형변수들 간의 선형독립 여부를 파악하였다. 그리고 마지막

으로 고속도로 나들목/분기점 연결로 별 사고예측모형에 대해 적합도 검증(Goodness-of-Fit)을 통하여 우도비(ρ^2)와 추정된 계수의 p-value 통하여 적합한 결과가 도출되었는지 검증을 실시하였다.

먼저, 연결로 형태 별로 음이항회귀모형을 수립하여 α (과분산계수)를 확인해 본 결과 모두 0을 초과하는 것으로 확인되었다. 따라서 포아송회귀모형은 배제하고 연결로 형태 별로 각각 음이항회귀모형과 영과잉음이항회귀모형으로 교통사고예측모형을 수립하였으며 신뢰수준은 90%에서 이루어졌다. 먼저 음이항회귀모형으로 수립된 고속도로 나들목/분기점 연결로 별 교통사고예측 모형은 다음 식과 같다.

• 직결형 연결로

$$\begin{aligned} Y &= EXPO \times \exp(\alpha + \beta_1 X_1 + \beta_2 X_2) \\ &= EXPO \times \exp(-0.575 - 0.221 X_1 - 0.103 X_2) \end{aligned} \quad (9)$$

여기서,

X_1 : 최소곡선반경

X_2 : 종단경사

• 준직결형 연결로

$$\begin{aligned} Y &= EXPO \times \exp(\alpha + \beta_1 X_1 + \beta_2 X_2) \\ &= EXPO \times \exp(-3.565 + 0.884 X_1 + 0.412 X_2) \end{aligned} \quad (10)$$

여기서,

X_1 : 최소곡선반경

X_2 : 가감속차로 길이

• 루프형 연결로

$$\begin{aligned} Y &= EXPO \times \exp(\alpha + \beta_1 X_1 + \beta_2 X_2) \\ &= EXPO \times \exp(1.449 - 1.563 X_1 + 0.167 X_2) \end{aligned} \quad (11)$$

여기서,

X_1 : 최소곡선반경

X_2 : 종단경사

다음으로 음이항회귀모형에 대한 영과잉모형으로서 영과잉음이항회귀모형을 수립하였다. 수립된 영과잉음이항회귀모형으로 수립된 고속도로 나들목/분기점 연결로 별 교통사고예측 모형은 다음 식과 같다.

• 직결형 연결로

$$Y = EXPO \times \exp(\alpha + \beta_1 X_1 + \beta_2 X_2) \\ = EXPO \times \exp(-1.195 - 0.407 \times X_1 - 0.650 \times X_2) \quad (12)$$

여기서,

- X_1 : 최소곡선반경
- X_2 : 가감속차로 길이

• 준직결형 연결로

$$Y = EXPO \times \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) \\ = EXPO \times \exp(-1.114 + 1.116 \times X_1 - 0.384 \times X_2 - 0.439 \times X_3) \quad (13)$$

여기서,

- X_1 : 최소곡선반경
- X_2 : 종단경사
- X_3 : 가감속차로 길이

• 루프형 연결

$$Y = EXPO \times \exp(\alpha + \beta_1 \times X_1 + \beta_2 \times X_2) \\ = EXPO \times \exp(2.287 - 0.218 \times X_1 - 0.657 \times X_2) \quad (14)$$

여기서,

- X_1 : 종단경사
- X_2 : 가감속차로 길이

사고예측모형의 우도비(ρ^2)와 p-value는 Table 6과 같고, McFadden(1976)에 의하면 우도비(ρ^2)의 값이 0.2~0.4사이의 값을 가지면 모형은 아주 좋은 설명력을 가진다고 평가되고 있다. 본 연구의 분석 결과, 우도비(ρ^2)의 값은 약 0.07~0.34 사이의 값으로 각 교통사고 예측모형 별 설명력의 차이가 있음을 확인할 수 있었다. 또한 직결형, 준직결형 그리고 루프형 연결로에 대한 영과잉음이향회귀모형의 Vuong 통계값은 각각 -0.957, 1.996, 그리고 -1.741로 분석되었다. 이는 직결형과 루프형의 경우는 두 모형 중 어떠한 모형이 적합한지 판단할 수 없었다. 단, 준직결형은 영과잉음이향회귀모형을 사용하는 것이 더 좋음을 의미한다.

4.3. 모형검증

앞 절에서 보인 바와 같이 연결로 형태별로 음이향회귀모형과 영과잉음이향회귀모형의 각종 통계량이 서로 상반된 결과를 보이고 있다. 따라서 본 연구에서 개발된 모형에 실제 자료를 적용한 후 실제 교통사고 건수와 추

Table 6. Comparison of Traffic Crash Prediction Models by Ramp Types

Type		Constant	Radius	Grade	Length of acc./dec. lane	α	
Negative binomial regression model	Direct	coeff.	-0.575	-0.221	-0.103	-	2.24
		p-value	0.164	0.085	0.088	-	0.021
		ρ^2	0.072				
	Semi-direct	coeff.	-3.565	0.884	-	0.412	3.17
		p-value	0.006	0.069	-	0.072	0.018
		ρ^2	0.236				
	Loop	coeff.	1.449	-1.563	0.167	-	1.33
		p-value	0.018	0.028	0.089	-	0.023
		ρ^2	0.085				
Zero-inflated negative binomial regression model	Direct	coeff.	-1.195	-0.407	-	0.650	3.378
		p-value	0.248	0.018	-	0.068	0.011
		ρ^2	0.173				
	Semi-direct	coeff.	-1.114	1.116	-0.394	-0.439	4.210
		p-value	0.165	0.001	0.000	0.014	0.023
		ρ^2	0.344				
	Loop	coeff.	2.287	-	-0.218	-0.657	0.961
		p-value	0.011	-	0.099	0.075	0.075
		ρ^2	0.082				

Note) If α is greater than 0 and the p-value for α is statistically significant, then the negative binomial regression model may be applicable.

정된 교통사고 건수를 비교함으로써 최종적으로 고속도로 연결로에 적합한 교통사고예측모형을 제시할 수 있을 것으로 판단하였다. 또한 통사고예측모형이 도출되면, 개발된 교통사고예측모형이 실제 교통사고 건수를 얼마나 적절하게 예측할 수 있는 검증할 필요가 있다. 이를 위하여 모형 개발에 사용되지 않은 고속도로 연결로에서 일어난 실제 교통사고 건수와 교통사고예측모형을 통해 예측된 교통사고 건수를 Table 7에서 보인 바와 같이 비교해 보았다. 모형에 포함되지 않은 인터체인지는 총 12개이고 이 중 교통사고가 발생한 연결로만을 대상으로 검증을 실시하였다. 이를 통해 음이항회귀모형과 영과잉음이항회귀모형에 대한 교통사고예측력을 살펴보았으며 결과는 다음과 같다. 여기서, 오차율은 실제 교통사고 건수와 예측된 교통사고 건수 차이의 절대값을 실제 교통사고건수로 나눈 값이다.

Table 7에서 보인 바와 같이 전체적으로 두가지 교통사고예측모형이 모든 연결로 유형에 대하여 높은 오차율을 보이고 있다. 이는 교통사고와 관련된 여러 요인(운전자특성, 환경요인, 교통요인, 도로운영요인, 기하구조요인 등)들을 모두 교통사고예측모형에 담지 못한 것이 주요 원인이라고 판단된다. 또한 각종 설계도면의 누락으로 인하여 교통사고예측모형 개발에 사용된 자료수의 부족도 영향을 주었을 것으로 판단된다.

직결형 연결로에서의 검증 결과를 살펴보면 평균오차율이 음이항회귀모형이 영과잉음이항회귀모형보다 더 작은 것으로 나타났다. 이와 같은 결과는 준직결형과 루프형 연결로에서도 동일하게 나타났다. 특히, 영과잉음이항회귀모형의 경우 남구미 나들목의 직결형 및 준직결형 연결로와 점촌·함창 나들목의 루프형 연결로에서 음이항회귀모형보다 상대적으로 큰 오차율을 보였다. 전체적으로 살펴보면 고속도로 연결로에서의 교통사고예측모형은 영과잉음이항회귀모형보다는 음이항회귀모형의 사고예측력이 좀 더 높은 것으로 판단된다.

교통사고예측모형의 예측력 검증에서 루프형 연결로 중 남구미 나들목과 같이 오차율이 상대적으로 큰 것을 볼 수 있다. 이와 같이 오차율이 다른 연결로에 비해 유난히 커지는 것은 사고예측모형에 포함된 교통사고노출계수(exposure)의 영향인 것으로 판단된다. 교통사고노출계수의 특성 상 교통량이 늘어날수록 교통사고도 함께 증가하게 되는데 실제로는 교통량이 매우 높음에도 불구하고 사고건수가 적은 경우가 있어서 이와 같이 오차율이 높아지는 경우가 발생하는 것으로 판단된다. 참고로, 모형개발에 사용된 준직결형 연결로들의 평균 교통사고 건수, AADT 그리고 곡선길이는 각각 0.83건/3년, 4,064대/일 그리고 578m인데 비해 남구미 나들목의 직결형 연결로의 교통사고 건수, AADT 그리고 곡

Table 7. Validation of Two Types of Traffic Crash Prediction Models

Type	Location	Actual Frequency of Crashes	NBR Model		Zero-inflated NBR Model	
			Predicted Frequency	Error Rate	Predicted Frequency	Error Rate
Direct	Nam-Gumi(off-ramp)	3	1.49	50%	9.43	214%
	Nam-Gumi(on-ramp)	2	0.15	93%	0.33	84%
	Seo-Cheongju(off-ramp)	1	0.48	52%	1.66	66%
	Seo-Cheongju(on-ramp)	1	1.11	11%	0.91	9%
	Yeongcheon	2	0.09	95%	0.39	81%
	Jeomchon-Hamchang	1	0.09	91%	0.08	92%
	Chilgok	2	1.39	30%	1.42	29%
	Average Error Rate			60.3%		82.1%
Semi-direct	Nam-Gumi	1	2.03	103%	4.06	306%
	Dong-Chanwon	2	0.67	66%	4.28	114%
	Bibong	1	1.23	23%	1.76	76%
	Yeongcheon	2	1.37	32%	1.62	19%
	Okgok	2	0.10	95%	0.26	87%
	Average Error Rate			63.8%		120.4%
Loop	Nam-Gumi	3	1.59	47%	2.46	18%
	Okgok	2	1.31	34%	1.46	27%
	Jeomchon-Hamchang	2	3.30	65%	4.54	127%
	Average Error Rate			48.7%		57.3%

선길이는 각각 1건/3년, 8,076대/일 그리고 600m이다. 따라서 교통사고 예측모형에 교통사고 노출계수(exposure)를 적용할 때 해당 연결로의 AADT 값이 과도하게 클 경우 주의할 필요가 있다고 판단된다.

5. 결론 및 향후 연구과제

본 연구에서는 2007~2009년(3년간) 우리나라 고속도로 전 노선의 나들목/분기점 연결로의 교통사고자료와 도로선형자료를 수집하고, 모형도출 시 교통사고 특성인 과분산현상을 반영하기 위해 음이항분포를 이용하여 통계적 유의성을 확보한 연결로 별(직결형, 준직결형, 루프형)로 교통사고예측모형을 음이항회귀모형과 영과잉음이항회귀모형으로 도출하였다.

분석결과 음이항회귀모형에서는 곡선반경이 모든 연결로 형태에서 공통된 변수로 선정되었고 직결형과 루프형 연결로에서는 중단경사, 준직결형에서는 가감속차로장이 유의한 변수로 선정되었다. 영과잉음이항회귀모형에서는 가감속차로장이 공통된 변수로 선정되었으며 직결형에서는 곡선반경, 준직결형에서는 곡선반경과 중단경사, 루프형에서는 중단경사가 유의한 변수로 선정되었다.

모형의 검증을 위해 실제 발생한 교통사고와 모형에 의해 예측된 교통사고를 비교해 보았다. 검증은 모형 개발에 포함되지 않은 나들목 12개소에서 실제 관측된 교통사고와 예측된 교통사고를 비교해 본 결과 평균오차율이 음이항회귀모형은 각각 60.3%, 63.8%, 48.7%였고 영과잉음이항회귀모형에서는 각각 82.1%, 120.4%, 57.3%였다. 전체적으로 살펴보면 음이항회귀모형이 영과잉음이항회귀모형보다 평균오차율이 낮아 음이항회귀모형의 예측력이 상대적으로 우수한 것으로 나타났다.

본 연구의 개발된 모형은 도로의 설계단계에서 사고발생 가능성이 높은 도로선형요소들의 조합을 사전에 파악하는데 도움을 줄 수 있을 것으로 판단한다. 또한, 이러한 도로선형요소의 특징을 이해하고 설계에 임할 경우 향후 발생할 교통사고 감소에 기여할 것으로 사료된다.

하지만, 교통사고는 기하구조, 운전자 및 차량특성, 안전시설, 기후조건, 노면포장상태 등 여러 가지 요인이 상호 복합적으로 작용하여 발생하고 있다. 이에 선형적인 요인 외의 교통사고에 영향을 줄 수 있는 자료를 정량화시켜 보다 다양한 요인을 사고예측모형 개발에 활용할 수 있는 연구가 필요하다. 또한 연결로의 기하구조 요소 외에도 연결로 상에 설치된 각종 교통안전시설도

교통사고 발생 건수에 영향을 미칠 수 있기 때문에 이에 대한 연구가 진행될 필요가 있다. 그리고 무엇보다도 이제는 교통사고분석에 활용할 수 있는 교통사고자료, 교통량자료 그리고 기하구조자료를 보다 체계적으로 구축 및 관리함으로써 보다 신뢰성 높은 분석을 도모할 필요가 있다고 판단된다.

감사의 글

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2010-0028693).

References

- Kang, Y., 2008. *Bayesian Model Selection for the Zero-inflated Models*, Master's Thesis, Hanyang Univ. Seoul, Korea.
- (강윤희, 2008. *영과잉 모형의 베이지안 모형 선택론*, 석사학위논문, 한양대학교, 서울, 대한민국)
- Kang, J. and Lee, S., 2002. (Traffic Accident Prediction Model by Freeway Geometric Types, *Journal of Korean Society of Transportation*, Vol. 20, No. 4, 163-175.
- (강정규, 이성관, 2002. 고속도로 선형조건별 교통사고 위험도 평가모형 개발(호남고속도로를 중심으로), *대한교통학회지*, 제20권, 제4호, 163~175)
- National Police Agency, 2011. *Traffic Accident Statistics (Year 2010)*, Seoul, Korea.
- (경찰청, 2011. *교통사고통계(2010년 통계)*, 경찰청, 서울, 대한민국)
- Kim, B. 2008. *The Study on Discriminant Model of Traffic Accidents at the Interchange of Freeway*, Master's Thesis, Ajou University, Suwon, Korea.
- (김봉곤, 2008. *고속도로 유출·입부의 교통사고 판별모형에 관한 연구*, 석사학위논문, 아주대학교, 수원, 대한민국)
- Rho, C., Park, J. and Son, B., 2008. A study of Accident Models for Highway interchange Ramps, *Journal of Korean Society of Transportation*, Vol. 26, No. 4, 21-40.
- (노창균, 박중서, 손봉수, 2008. 고속도로 연결로의 교통사고 추정모형, *대한교통학회지*, 제26권, 제4호, 21~40)
- Park, H., Son, B., Kim, H., 2007. Development of Accident Prediction Models for Freeway Interchange Ramps, *Journal of Korean Society of Transportation*, Vol. 25, No. 3, 123-135.
- (박효신, 손봉수, 김형진, 2007. 고속도로 인터체인지 연결로에서의 교통사고 예측모형 개발, *대한교통학회지*, 제25권, 제3호, 123~135)
- Park, B., Park, S., Lee, Y., In, B., 2008. Accident Analysis of Unsignalized Intersections Using ZAM, *Journal of Korea Planners Association*, Vol. 43, No. 6, 69-78
- (박병호, 박상혁, 이영민, 인병철. 2008. ZAM 모형을 이용한

- 비신호교차로 사고특성 분석, *대한국토·도시계획학회지*, 제43권, 제6호, 69~78)
- Seo, H., 2009, *Regression Analysis, 3rd Ed.*, Hannarea Academy, Seoul, Korea.
- (서혜선, 2009. *회귀분석 제3판*, 한나래아카데미, 서울, 대한민국)
- Seong, N., 2002. The Development of Traffic Crash Prediction Models at Freeway Interchanges, *Journal of Civil Engineering*, Vol. 22, No. 4-D, 617-625.
- (성낙문, 2002. 고속도로 인터체인지에서 교통사고 예측모델의 개발, *대한토목학회지*, 제22권, 제4-D호, 617~625)
- Yoon, B., 2008. Development of Traffic Accident Rate Forecasting Models for Trumpet IC Exit Ramp of Freeway using Variables Transformation Method, *International Journal of Highway Engineering*, Vol. 10, No. 4, 139-150.
- (윤병조, 2008. 변수변환 기법을 이용한 고속도로 트럼펫IC 형식 유출연결로 교통사고율 예측모형 개발, *한국도로학회 논문집*, 제10권, 제4호, 139~150)
- Choi, B., 2011. *Study on Decision Tree for Zero-Inflated Count Data*, Master's Thesis, Seongshin Woman University, Seoul, Korea.
- (최보미, 2011. *영과잉 이산형 자료를 위한 의사결정나무 연구*, 석사학위논문, 성신여자대학교, 서울, 대한민국)
- Choi, Y., 2011. *Study on Development and Application of Accident Modification Factors(AMF) for Expressway Ramps*, Ph.D. Dissertation, Ajou University, Suwon, Korea.
- (최윤환, 2011. *고속도로 연결로 구간의 사고예측계수 개발 및 활용방안 연구*, 박사학위논문, 아주대학교, 수원, 대한민국)
- Korea Expressway Corporation, 2010. *Statistics for Administration Use*, Korea Expressway Corporation, Seongnam, Korea.
- (한국도로공사, 2010. *2009년 업무통계*, 한국도로공사, 성남, 대한민국)
- Hwng, K., 2010. *Development of Vehicle Crash Model Based on Road Characteristics*, Master's Thesis, University of Seoul, Seoul, Korea.
- (황경성, 2010. *구간분할에 따른 고속도로 사고빈도 모형에 관한 연구*, 석사학위논문, 서울시립대학교, 서울, 대한민국)
- AASHTO, 2010. *Highway Safety Manual*, American Association of State Highway and Transportation Officials, Washington, D.C.
- Bauer, K. M. and Harwood, D. W., 1997. Statistical Models of Accidents on Interchange Ramp and Speed-Change Lanes, *FHWA-RD-97-106*.
- FHWA, 2011. *IHS DM User's Manual*.
- Harwood, D. W., Council, F. M., Hauer, E., Hughes, W. E. and Vogt, A., 2005. Prediction of the Expected Safety Performance of Rural Two-lane Highways, *FHWA*
- Hilbe, J. M., 2008. Negative Binomial Regression, *Cambridge University Press*, 39-135.
- Shankar, V., Mannering, F. and Barfield, W., 1995. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies, *Accident Analysis and Prevention*, Vol.27. No.3.
- Washington, S. P., Karlaftis, M. G., Mannering, F., 2003 *Statistical and Econometric methods for Transportation Data Analysis*, Chapman & Hall/CRC.
- (접수일 : 2012. 7. 12 / 심사일 : 2012. 7. 15 / 심사완료일 : 2012. 9. 27)