

# 확률적 모수를 고려한 음이항모형에 의한 교통사고와 기하구조와의 관계 - 미국 워싱턴 주(州) 고속도로를 중심으로

박민호\*  
Park, Minho\*

## Relationship between Interstate Highway Accidents and Heterogeneous Geometrics by Random Parameter Negative Binomial Model - A case of Interstate Highway in Washington State, USA

### ABSTRACT

The objective of this study is finding the relationship between interstate highway accident frequencies and geometrics using Random Parameter Negative Binomial model. Even though it is impossible to take account of the same design criteria to the all segments or corridors on the road in reality, previous research estimated the fixed value of coefficients without considering each segment's characteristic. The drawback of the traditional negative binomial is not to explain the integrated variations in terms of time and the distinct characters specific segment has. This results in under-estimation of the standard error which inflates the t-value and finally, affects the modeling estimation. Therefore, this study tries to find the relationship of accident frequencies with the heterogeneous geometrics using 9-years and 7-interstate highway data in Washington State area. 16-types of geometrics are used to derive the model which is compared with the traditional negative binomial Model to understand which Model is more suitable. In addition, by calculating marginal effect and elasticity, heterogeneous variables' effect to the accidents are estimated. Hopefully, this study will help to establish the future policy of geometrics.

**Key words :** Accident model, Random parameter negative binomial regression analysis, Interstate highway geometrics, Heterogeneity

### 초록

이 연구는 확률적 모수를 고려한 음이항회귀분석을 이용하여 고속도로에서의 사고와 기하구조와의 관계를 파악하는데 목적이 있다. 고속도로에서의 기하구조는 모든 구간에 동일한 설계요소가 적용되기에는 현실적으로 불가능함에도 불구하고, 지금까지의 연구에서는 모형을 통해 도출되는 계수값이 구간에 설치된 기하구조의 특성에 관계없이 항상 고정된 값으로 추정되어왔다. 고정된 값을 이용한 일반적인 음이항모형은 시간적 변화 또는 각 대상구간이 가지고 있는 고유한 특성에 따른 변화를 통합하여 설명하지 못한다는 단점이 있으며, 이로 인해 추정된 계수의 표준오차가 과소추정되어 t-값이 부풀려지게 되며, 그 결과 모형의 설명력이 떨어지게 된다. 따라서, 이 연구에서는 워싱턴 주에 위치하고 있는 7개의 고속도로에서 발생한 9년동안의 사고자료 및 기하구조자료를 이용하여 구간별로 상이한 기하구조가 사고에 미치는 영향을 알아보려고 한다. 총 16개의 기하구조 관련 변수가 모형 도출에 이용되었으며, 기존의 음이항모형과의 비교를 통해 이 연구에서 제시하는 모형이 교통사고와 기하구조와의 관계파악에 더욱 적합함을 보이려고 한다. 그리고, 각 변수의 한계효용 및 탄력성 분석을 통해 이질성을 가지는 기하구조가 사고에 미치는 영향을 제시하고자 한다. 이는 향후 기하구조 관련 정책수립에 도움이 될 것으로 판단된다.

**검색어 :** 사고모형, 확률적 모수를 고려한 음이항회귀분석, 고속도로 기하구조, 이질성

\* 정회원 · 펜실베이니아 주립대학교 (Corresponding Author · The Pennsylvania State University · minhopark77@gmail.com)

Received April 11, 2013/ revised May 20, 2013/ accepted June 10, 2013

Copyright © 2013 by the Korean Society of Civil Engineers

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

# 1. 서론

## 1.1 연구의 배경 및 목적

오랜기간동안 교통사고의 특성을 파악하기 위해 많은 연구들이 진행되어왔으며, 대표적인 연구가 교통사고와 도로기하구조와의 관계파악이다. 이러한 연구들을 통해서 교통사고에 영향을 미치는 도로기하구조 요소를 파악하여 이를 개선함으로써 교통사고를 줄이는 노력을 하여왔다. 이러한 노력은 주로 통계모형을 통해서 이루어져 왔으며, 이를 위해 사용된 주요통계모형은 가산모형(Count Model)이며, 선형회귀모형(Linear Regression Model), 포아송(Poisson Model) 그리고 음이항모형(Negative Binomial Model)이 대표적인 가산모형이다.

하지만 이러한 모형들은 모든 관측값(도로구간 혹은 교차로 등)에 대한 추정된 모수(parameter)가 고정되어있다고 가정하고 있다. 즉, 도로의 구간에 설치되어 있는 기하구조의 길이폭등의 설치형태에 관계없이 동일한 값이 적용되어 사용되어 왔다. 도로기하구조의 경우, 현실적으로 동일한 기하구조가 도로전체에 적용될 수가 없음에도 불구하고, 지금까지 모형에 적용된 변수들은 이러한 현실을 고려하지 않은채 사용되고 있는 실정이다. 그 결과, 모형을 통해 추정된 계수의 표준오차값이 과소추정되었으며, 이는 계수의 t-값을 과도하게 부풀리게 되어 모형의 설명력이 떨어지게 되는 결과를 초래하게 된다.

따라서, 이번 연구에서는 도로의 각 구간별로 상이할 수 있는 기하구조의 이질성(heterogeneity)을 고려하여 기하구조와 교통사고와의 관계를 파악하고자 한다.

## 1.2 연구의 내용

이 연구에서는 기존의 음이항모형(Fixed Parameter Model, 이하 FPM)과 이 연구에서 제시하고자 하는 도로기하구조의 이질성을 고려한 음이항모형(Random Parameter Model, 이하 RPM)의 비교를 통해 보다 현실적인 방법을 제시하고자 한다. 이 두가지 형태의 모형은 도출된 계수값 및 계수값들의 탄력성, 한계효과와 함께 RMSE(Root Mean Square Error : 평균제곱근오차), MAPE (Mean Absolute Perchange Error : 평균백분율오차의 평균), 그리고 MAE(Mean Absolute Error : 평균절대오차)방법에 의해 비교되었다. 자세한 연구 수행절차는 Fig. 1에 제시되어 있다.

## 1.3 연구의 범위

본 연구의 공간적·시간적·내용적 범위는 다음과 같다.

### 1.3.1 공간적 범위

미 워싱턴 주(Washington State)에 위치한 7개의 고속도로

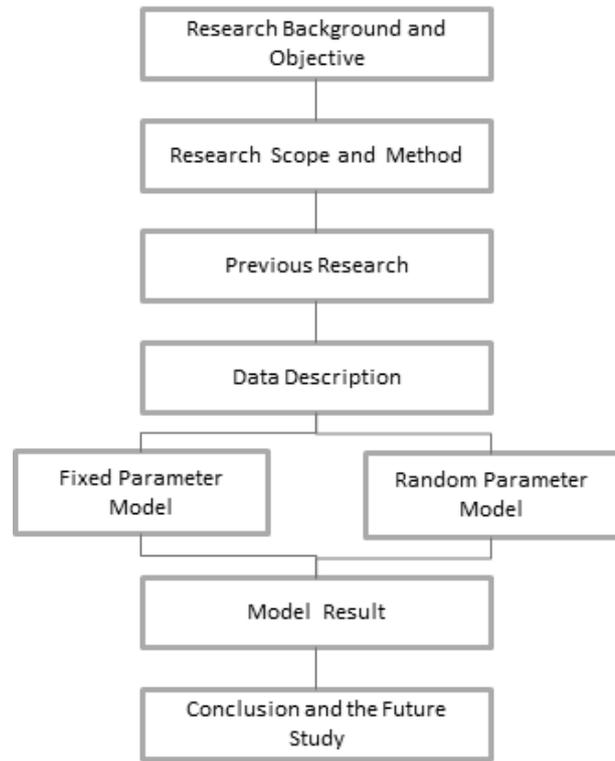


Fig. 1. Research Flow

(interstate highway) : I-5, 82, 90, 182, 205, 405, 그리고 705(1,528 mile≒2,458 km)

### 1.3.2 시간적 범위

9년(1999~2007년) 동안의 고속도로 교통사고, 교통량 및 기하구조 자료

### 1.3.3 내용적 범위

도로구간에 설치된 기하구조의 이질성을 고려한 음이항모형 도출 및 각 변수별 사고에 미치는 영향분석

## 2. 기존 문헌 고찰

### 2.1 교통사고 모형

기존의 사고모형은 유형별로 크게 3가지 형태-선형회귀식, 포아송 회귀식, 음이항 회귀식으로 나누어진다. 우선 선형회귀식은 사고 발생에 영향을 미치는 인자들을 분석하는 가장 단순한 기법이며, 다음과 같이 나타낼 수 있다.

$$y_i = \alpha + \beta x_i + \epsilon_i \quad (1)$$

여기서,  $y_i$  : 구간  $i$ 에서의 사고건수, 사상자수 또는 사고율(종속 변수)

$\alpha$ 와  $\beta$  : 상수와 회귀계수(추정값)

$x_i$  : 구간  $i$ 에서의 사고요인(독립변수)

$\epsilon_i$  : 확률오차항  $N(0, \sigma^2)$

$$\lambda_i = \exp(\beta X_i + \epsilon_i) \tag{3}$$

여기서,  $\exp(\epsilon_i)$  : 평균이 1이고 분산이  $\alpha$ 인 감마분포를 따르는 오차항

사고모형 중 가장 단순한 방법이긴 하지만, 계수 추정에 사용되는 변수값이 증가할수록 분산값 또한 증가하게 되어 선형회귀식의 기본가정인 동분산성(Homoscedasticity)을 위배하게 된다. 그 결과 변수의 유의수준에 변화를 주게 되며, 이는 변수에 대한 통계적 유의성을 낮추게 된다. 뿐만 아니라 교통사고수와 같은 양의 변수에 대해 음(Negative)의 사고수를 예측한다는 단점이 있다. 특히 교통사고의 특성인 특정기간 동안 사고가 발생하지 않았거나 낮은 사고건수에 대해선 음의 값을 예측하게 된다(Jovanis and Chang, 1986). 또한, Miao et al.(1993)은 기존의 선형모형은 산발적(sporadic)이고 무작위(random)로 발생하는 사고의 모형화에는 적합하지 않다고 주장하였다. 이러한 음의 사고예측값에 대한 문제를 해결하기 위한 한 가지 방법으로 0의 값을 나타내는 사고를 제거하고 분석하는 기법(Left-Truncating the Accident Frequency at Zero)을 사용할 수 있으나, 이는 모형 추정에 사용되는 표본수가 줄어들 뿐만 아니라, 모형의 개념이 타 모형에 비해 열악해진다. 이러한 문제점을 해결하고자, Jovanis and Chang(1986), 그리고 Joshua and Garber(1990)들은 사고건수를 이산적 확률변수(Discrete Random Variable)로 해석하는 포아송회귀식(Poisson Regression)을 제안하였으며, 포아송회귀식의 일반식은 다음과 같다.

$$P(n_i) = \frac{e^{-\lambda} \cdot \lambda^{n_i}}{n!} \tag{2}$$

여기서,  $P(n_i)$  : 사고  $n$ 이 지점  $i$ 에서 발생할 확률

$\lambda$  : 평균사고발생건수( $=\exp(\beta X_i)$ )

$\beta$  : 추정된 계수

$X_i$  : 지점  $i$ 의 속성(교통량, 도로 기하구조·환경등)

포아송모형은 “분산과 평균이 같다.”라는 중요한 기본전제가 있다. 하지만, 실제 발생하는 교통사고는 대부분의 경우, 분산값이 평균값보다 큰 과분산(Overdispersion)의 형태를 보이게 되어 포아송 모형이 적합하지 않은 경우가 발생하게 된다. 이를 보완하기 위해 음이항회귀식(Negative Binomial Regression Model)이 주로 사용되고 있다. 음이항분포에는 사고수( $\lambda_i$ )와 오차항( $\epsilon_i$ )이 포함되며 이는 다음식과 같다.

음이항 모형은 포아송모형의 확장된 형태이며, 만약 오차항의  $\alpha$ 값이 통계적으로 0과 다른 경우에는 음이항 모형이 적합하며, 반대의 경우( $\alpha=0$ )에는 포아송모형이 적합하게 된다. Engel (1984), Lawless (1987), Maher (1991) 그리고 Miao (1994)는 교통사고와 기하구조와의 관계 파악을 위한 연구에서 음이항모형이 선형 또는 포아송 모형보다 우수하다는 연구결과를 발표하였다.

### 2.2 교통사고와 기하구조와의 관계

교통사고와 기하구조와의 관계에 관한 연구는 많은 연구자들에 의해 이루어져 왔는데, 연구에 사용된 기하구조는 교통량을 포함하여 길어깨폭(Ogden, 1997), 차선수와 차선넓이, 그리고 제한속도(Noland, 2003), 종횡단 곡선 구성요소의 최소최대값(Zhang et al., 2005), 시거, 구배(Caliendo et al., 2007), 인터체인지 램프(Montella et al., 2008)등을 사고에 미치는 변수로 채택하여 주로 음이항 모형으로 분석을 하였다. 하지만, 추정된 계수값들이 도로의 구간에 관계없이 모든 구간에 대해 고정된 값으로 도로의 각 구간이 가질 수 있는 각각의 이질적인 특성에 관한 가능성은 배제하였다는 단점이 있다.

이러한 이질성과 관련된 연구측면에서는 Shankar et al.(1998)은 2가지 형태의 모형-Random Effects Negative Binomial (RENB)과 Cross-sectional Negative Binomial을 이용하여, RENB 모형이 상수값에서 이질성이 도출되도록 하였으며, Chin et al.(2003)은 싱가포르의 교차로에서 발생한 사고를 이용하여, 관측되지 않은 이질성(Unobserved Heterogeneity)과 시계열 상관(serial correlation)을 다루기 위해 RENB모형을 적용하였다. Washington et al.(2010)은 계수값이 관측값에 따라 다양한 값을 나타내지만, 고정된 값이라고 가정하여 이를 제약할 경우, 일관성이 없고(inconsistent) 편중된(unbiased) 결과가 도출됨을 증명하였다.

### 3. 자료구축

#### 3.1 분석범위

이 연구에서 사용된 교통사고, 교통량 및 기하구조 자료는 총 9년 동안의 미국 워싱턴 주(Washington State)에 속한 고속도로(interstate highway)를 대상으로 불균형 패널데이터(Unbalanced Panel Data)<sup>1)</sup> 형식으로 구축되었다. 데이터 구축에 포함된 고속도

1) 기하구조의 경우, 9년 동안 개선 혹은 변경된 자료의 수집이 불가능하여 시간의 흐름에 관계없이 동일하다고 가정함.

로는 I-5, I-82, I-90, I-182, I-205, I-405, 그리고 I-705의 7개 노선이며, 보다 자세한 노선의 길이 및 위치는 Fig. 2에 나타나있다.

모든 고속도로는 방향별로 본선구간과 인터체인지 구간으로 구분되어있는데, 이는 기본적으로 두 구간의 기하구조가 상이하기 때문임과 동시에 교통사고 분석에서 주요한 요인인 교통량이 인터체인지를 기준으로 바뀌기 때문이다. 구간 분할의 경우, 또 다른 형태로 가능(기하구조가 변경 되는 지점 등)하지만, 이 연구에서는

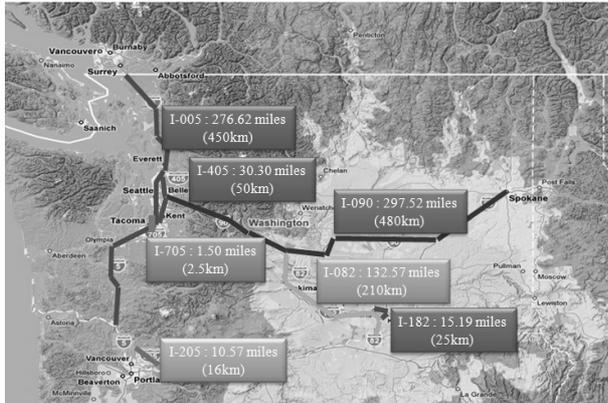


Fig. 2. Interstate Highway in Washington State

Table 1. Number of Segments and Miles by Route and Direction

Route	Increasing Direction			
	Interchange Segment		Non-interchange Segment	
	Miles(Kilometers)	Segments	Miles(Kilometers)	Segments
5	104.84(168.69)	138	171.78(276.39)	139
82	27.51(44.26)	33	105.06(169.04)	34
90	65.73(105.76)	83	231.79(372.95)	83
182	7.08(11.39)	8	8.11(13.05)	9
205	4.14(6.66)	5	6.43(10.35)	6
405	13.53(21.77)	19	16.77(26.98)	20
705	0.5(0.80)	1	1(1.61)	2
Total	223.33(375.43)	287	540.94(870.37)	293
Route	Decreasing Direction			
	Interchange Segment		Non-interchange Segment	
	Miles(Kilometers)	Segments	Miles(Kilometers)	Segments
5	104.48(168.11)	138	172.13(276.96)	140
82	26.5(42.64)	33	106.07(170.67)	34
90	65.28(105.04)	85	232.24(373.67)	85
182	7.55(12.15)	8	7.64(12.29)	9
205	4.06(6.54)	5	6.51(10.47)	6
405	13.84(22.27)	20	16.46(26.48)	21
705	0.55(0.88)	2	0.95(1.53)	2
Total	222.26(357.62)	291	542.00(872.08)	297

이질성을 고려하였기 때문에 모형에서 도출되는 계수값 등을 통해 설명이 되어지며, Milton 등(2008)의 연구에서도 기술되어 있다.

인터체인지 구간은 본선에서 인터체인지로의 분류부 시점에서 인터체인지에서 본선으로의 합류부 시점까지의 구간이며, 본선구간은 그 외의 구간으로 정의하였다. 이러한 전제조건 하에 총 7개노선의 1,528mile(2,458km)구간이 1,168개의 구간으로 구분되었으며, 구간의 노선별 방향별 개수 및 길이는 Table 1에 나타나있다. 본 연구에서 사용된 기하구조에 대한 자료는 정의된 각각의 구간에 대한 값으로 변환되어 적용되었다. 그리고 교통량과 사고는 9년동안 각 년도별로 적용되었다.

### 3.2 개별자료분석

Table 2는 이 연구에서 사용된 확률적 모수를 고려한 음이항 모형에 사용된 변수에 대한 통계값을 나타내고 있다. 주요한 변수로는 연간 사고건수, 차선수, 좌우 길어깨폭, 횡단 곡선에 관한 변수, 구간의 길이 및 교통량이 사용되었으며, 연간 사고건수의 평균값은 10.70 표준편차는 18.92 그리고 최대값은 388건을 보여 주고 있다.

차선수 및 좌우 길어깨폭의 경우, 구간내에서 동일한 설계가 유지되는 곳도 있는 반면, 달라지는 경우도 고려하기 위해 이 연구에서는 구간별 길이비율로 정의하였다. 따라서 기하구조가 동일하게 유지되는 구간에서의 최대값은 1을 넘지 않으며, 상이한 기하구조가 구간내에 다소 존재하더라도 구간내에서의 합은 최대값인 1을 넘지 않게된다. 고속도로를 대상으로 분석이 진행되었기 때문에 2차선 도로는 많이 존재하지 않으며, 길이는 0.52의 평균값과 0.494의 표준편차값을 가진다. 3차선 도로의 경우에는 0.336, 0.461, 4차선은 0.135, 0.331 그리고 5차선 도로의 경우에는 0.002, 0.043의 평균값과 표준편차값을 가진다.

길어깨의 경우에는 5가지의 경우(2-foot 이하, 3~4, 5~9, 10, 10-foot 이상)로 구분하였으며, 차선수와 마찬가지로 구간내에서 차지하는 비율로 정의하였다.

구간내에서 횡단곡선의 개수는 평균적으로 1.805개가 존재하는 것으로 파악되었으며, 최대 37개의 횡단곡선이 존재하는 것으로 나타났다. 또한, 가장 짧은 횡단곡선은 평균적으로 0.182mile(0.29km)의 길이를 가지고 있으며, 반면 가장 긴 횡단곡선은 0.275mile(0.44km)의 평균값을 가지는 것으로 파악되었다. 횡단곡선의 최소교각의 평균값은 12.186°이며, 최대교각은 21.080°의 평균값을 가지는 것으로 나타났다.

종단곡선의 경우에는 3.125개의 곡선이 구간내에서 평균적으로 존재하며, 최대개수는 30개인 것으로 파악되었다. 종단곡선과 관련된 변수로는 최소최대 종단구배가 있으며, 최소 구배는 1.232%, 1.385%의 평균값과 표준편차값을 가지는 것으로 나타난 반면,

Table 2. Statistics of Accident, Geometrics and ADT Variables

Variable	Variable Description	Mean	Std.Dev.	Minimum	Maximum
ACC	Number of Accidents per year	10.703	18.917	0	388
NLN2	Portion of segment with 2 lanes	0.525	0.494	0	1
NLN3	Portion of segment with 3 lanes	0.336	0.461	0	1
NLN4	Portion of segment with 4 lanes	0.135	0.331	0	1
NLN5	Portion of segment with 5 lanes	0.002	0.043	0	1
LSHW2	2-ft(60cm) left-shoulder width segment proportion	0.175	0.337	0	1
LSHW34	3- to 4-ft(90~120cm) left-shoulder width segment proportion	0.199	0.387	0	1
LSHW59	5- to 9-ft(150~270cm) left-shoulder width segment proportion	0.133	0.320	0	1
LSHW10	10-ft(300cm) left-shoulder width segment proportion	0.472	0.478	0	1
LSHW1126	Over 10-ft(over 300cm) left-shoulder width segment proportion	0.020	0.132	0	1
RSHW2	2-ft(60cm) right-shoulder width segment proportion	0.178	0.341	0	1
RSHW34	3- to 4-ft(90~120cm)t right-shoulder width segment proportion	0.207	0.392	0	1
RSHW59	5- to 9-ft(150~270cm)t right-shoulder width segment proportion	0.121	0.308	0	1
RSHW10	10-ft(300cm) right-shoulder width segment proportion	0.464	0.477	0	1
RSHW1124	Over 10-ft(over 300cm) right-shoulder width segment proportion	0.030	0.155	0	1
NHORZ	Number of hirizontal curves in segment	1.807	2.459	0	37
MINHLMI	Shortest horizontal curve-in-segment length (miles)	0.182(0.29km)	0.193	0	1.219
MAXHLMI	Longest horizontal curve-in-segment length (miles)	0.275(0.44km)	0.277	0	2.402
MINHDEL	Smallest horizontal curve-in-segment central angle (degrees)	12.186	14.692	0	98.426
MAXHDEL	Largest horizontal curve-in-segment central angle (degrees)	21.080	21.001	0	111.294
NVERT	Number of vertical curves in segment	3.125	3.210	0	30
MINVCRVG	Smallest absolute vertical curve gradient (%)	1.232	1.385	0	7.430
MAXVCRVG	Largest absolute vertical curve gradient (%)	2.780	2.075	0	10
LENGTH	Segment length (miles)	1.309(2.11km)	1.753	0.01	20.380
AADT	Annual Average Daily Traffic Volume	13043.300	8378.490	916.45	44223.800

최대 구배의 경우에는 2.78%의 평균값과 2.075%의 표준편차값을 가지는 것으로 나타났다.

대상구간의 평균 길이는 1.309mile(2.11km)이며, 구간내 연평균 일 교통량(방향별 차선별)은 평균 13,043대로 나타났으며, 최대 교통량은 44,223대로 조사되었다. 구간의 길이와 교통량은 로그값으로 변환된 값이 모형에 적용되었다.

#### 4. 모형의 구축 및 결과분석

##### 4.1 확률적 모수를 고려한 음이항모형

###### (Random Parameter Negative Binomial)

이 연구에서 제시하고자 하는 확률적 모수를 고려한 음이항 모형이 기존의 음이항 모형과 다른 가장 큰 특징은 계수값( $\beta$ )이 고정되어 있지 않다는 것이다. 일반적으로 독립변수( $x$ )는 관측되거나 수집된 자료에 제약되어 있기 때문에 독립변수로부터 이용가능

한 최대한의 정보를 얻는것에 모델링의 목적이 있으며, 이러한 관점에서 각 계수를 고정된 값이 아니라, 공간별( $i$ ) 혹은 시간별( $t$ )로 다양하다고 가정하여 접근한다면 이질성을 고려한 계수값을 도출할 수 있게 된다. 예를 들면, 어느 특정분포가 확률모수의 지속적인 변화를 가장 잘나타낸다고 가정하는 것이며, 이러한 가정에 기반한 우도함수(likelihood function)는 고정된 계수가 아닌 변화하는 계수로부터 계산된 확률을 기반으로 해서 다음과 같이 수정될수 있게 된다.

$$\beta_{it} = \beta + \Delta h_i + \Gamma w_{it} \quad (4)$$

여기서,  $h$ 가 속한 두 번째 항은  $h_i$ 에 속한 변수들의 평균값에 대한 이질성을 나타내며, 세 번째 항은 평균으로 부터의 확률편차(random deviation)를 나타내게 된다. 따라서, Eq. (4)는 확률적 이질성(parameter heterogeniety)을 나타내는 기본적인 식이며,

특히 델타( $\Delta$ )는 계수( $\beta$ )에 영향을 미치는 외생변수(exogenous variable)를 포함하며, 만약 변수 평균값내에 이질성이 외생변수로 존재하지 않는다면, 이 항은 사라지게 될 것이다. 즉, 모든 이질성은 세 번째 항에 의해서 모형화 되어지게 된다. 이는 사용된 독립변수의

특성에 따른 분포의 형태를 바탕으로 적용될 것이며, 만약 독립변수의 형태가 이항변수(dummy variable)라면, 균일(uniform)분포가, 연속변수(continuous variable)라면 정규(normal) 또는 로그정규(lognormal)분포가 될 것이다. 따라서, 세 번째 항이 통계적으로

Table 3. Modeling Estimation Results for Fixed-and Random-Parameters Negative Binomial Models

Variable	Fixed Parameter Model (FPM)		Random Parameter Model (RPM)	
	Coefficient	t-value	Coefficient	t-value
<b>Exposure and Context</b>				
Constant	-8.409	-73.556	-7.624	-73.294
Logarithm of Length of segment in miles	0.743	68.992	0.787	83.210
Standard deviation of parameter distribution		-	0.087	11.882
Logarithm of ADT	1.143	99.726	1.036	98.715
Standard deviation of parameter distribution		-	0.008	4.283
Interchange indicator(1 if segment is an interchange segment; 0 otherwise)	-0.027	-1.991	-0.051	-4.418
<b>Number of Lanes by Length Proportion</b>				
Three-lane cross-section segment proportion	0.375	15.947	0.473	23.879
Standard deviation of parameter distribution		-	0.094	7.784
Four-lane cross-section segment proportion	0.902	33.995	0.856	38.592
Standard deviation of parameter distribution		-	0.167	10.530
Five-lane cross-section segment proportion	0.785	3.456	1.190	7.868
<b>Left Shoulder Width by Length Proportion</b>				
3- to 4-ft(90~120cm) left-shoulder-width proportion	-0.273	-8.528	-0.231	-8.430
Standard deviation of parameter distribution		-	0.287	16.698
5- to 9-ft(150~270cm) left-shoulder-width proportion	-0.496	-18.081	-0.393	-18.274
10-ft(300cm) left-shoulder-width proportion	-0.416	-20.219	-0.300	-16.131
Standard deviation of parameter distribution		-	0.154	14.000
<b>Right Shoulder Width by Length Proportion</b>				
3- to 4-ft(90~120cm) right-shoulder-width proportion	-0.286	-8.347	-0.279	-10.380
5- to 9-ft(150~270cm) right-shoulder-width proportion	-0.319	-13.737	-0.344	-14.733
Standard deviation of parameter distribution		-	0.127	6.930
10-ft(300cm) right-shoulder-width proportion	-0.395	-17.875	-0.320	-18.813
<b>Horizontal-Vertical Curvature</b>				
Number of horizontal curves in segment	0.029	7.080	0.019	6.150
Shortest horizontal curve-in-segment length in miles	-0.625	-12.484	-0.514	-12.109
Largest degree of curvature in segment	0.006	13.850	0.005	13.023
Largest vertical curve gradient in segment	0.020	5.327	0.029	8.955
Standard deviation of parameter distribution		-	0.021	13.638
<b>Dispersion parameter for negative binomial distribution</b>				
Dispersion parameter	0.348	51.092	5.960	41.550
Number of observations	10512			
Log-likelihood with constant only	-105,377.5			
Log-likelihood at convergence	-29,210.79		-28,315.69	
$\rho^2$	0.722		0.731	

유의하면, 계수값이 표준편차값과 함께 각 구간별로 상이한 값이(이질성을 가진다) 추정되며, 통계적으로 유의하지 않으면, 일반적인 음이항모형이 적용되어 계수값이 구간에 상관없이 고정된 값으로 추정된다.

#### 4.2 결과분석

이 연구에서는 Table 2에서 제시된 변수들을 확률적 모수를 이용한 음이항모형에 이용하였다. 종속변수는 워싱턴 주 7개의 고속도로에서 발생한 교통사고건수를 독립변수로는 기하구조, 교통량 및 구간의 길이로 설정하였으며, 통계프로그램인 LIMDEP을 이용하여 기존연구에서 주로 사용되는 음이항 모형(FPM)과 본 연구에서 제시하고자 하는 확률적 모수를 이용한 음이항 모형(RPM)을 도출하였다. 포이송 모형과 음이항 모형의 선택에 있어서 앞서 설명한대로 과분산값( $\alpha$ )이 0.348, t-값이 51.092임을 보여 통계적으로 0과 다르다는 유의성을 보이기 때문에 도출된 모형은 포이송분포보다 음이항분포를 따르는 것으로 나타나 결과제시는 음이항모형에 의한 결과만 제시하고자 한다.

모형의 도출은 Halton draws를 이용한 최대우도법을 사용하였으며, Bhat (2003) 그리고 Milton et al. (2008)의 논문에서 정확한 계수추정을 할 수 있다고 언급된 200 Halton Draws를 적용하였다. 계수추정의 또 다른 방법으로 이용될 수 있는 Random Draws도 있지만 계수값을 수렴하기 위해 Halton Draws 방법에 비해 많은 Draws가 필요하여 효율적이지 못하다고 알려져 있다(Train, 2003).

그리고 이 연구에서 제시하는 확률적 모수를 고려한 음이항 모형의 계수추정 시, 여러 형태의 확률분포(정규, 로그정규, 균일분포)를 고려할 수 있는데, 이중 정규분포가 통계적으로 가장 유효한 값을 보이는 것으로 나타났다.

도출된 계수의 한계효과(Marginal Effect)가 Table 4에 제시되어 있다. 이를 통해 기하구조의 단위(unit)변화가 사고발생에 미치는 상대적인 영향을 알 수 있다. 그리고 이용된 변수의 특성에 따라서 탄력성(Elasticity)을 제시하고자 한다.

도출된 모형의 전체적인 설명력은 로그우도함수 값이 -105,377.5에서 -28,315.69로 우도비( $\rho^2$ )가 0.731로 일반적인 음이항모형(-29,210.79,  $\rho^2=0.722$ )에 비해 상대적으로 조금 나은 설명력을 가지는 것으로 나타났다.

앞서 설명한대로 확률적 모수를 이용한 음이항 모형의 경우, 추정된 계수의 표준편차가 통계적으로 유의한 경우( $s.d \neq 0$ ,  $t\text{-value} \geq 1.96$ )에 사용될 수 있으며(RPM), 만약 추정된 계수의 표준편차가 통계적으로 유의하지 않으면( $s.d=0$ ,  $t\text{-value} < 1.96$ ) 추정된 계수가 일반적인 음이항 모형의 경우처럼, 구간에 관계없이 모든 구간에 대해서 동일한 값을 가지게 된다(FPM).

Table 3에 따르면, 구간의 길이가 길어지게 되면 사고발생확률이

높아짐을 알 수 있다. RPM에서는 평균값이 0.787 표준편차는 0.087이며 t-값에 의하면 상당히 높은 유의함을 보임을 알 수 있다. RPM의 탄력성 측면에서 1%의 구간길이 증가는 0.787%의 사고증가에 영향을 주는 것으로 나타났다. 그리고 FPM에서는 0.743%의 영향이 있는 것으로 나타났다. 이는 노출량이 많아질수록 사고발생 확률이 높아진다는 기존의 연구와 일치함을 알 수 있다. 또한, 구간 길이의 경우 통계적으로 유의한 표준오차 값을 보이기에 사고에 대한 영향은 구간에 따라 다르게 나타날 수 있으며, 최대 1168개의 다른 값을 가지게 된다.

교통량 또한 구간의 길이와 마찬가지로 교통량이 증가할수록 사고발생에 영향을 많이 주는 것으로 나타났다. FPM에서는 1%의 교통량 증가가 1.143%의 사고 증가로 이어짐을 알 수 있다. 반면, RPM에서도 상당히 유의한 1.036의 평균값과 0.008의 표준편차값을 가지는 것으로 나타났다. 탄력성 측면에서 두 가지 경우에서 모두 탄력적임을 보이는데, 1%의 교통량 증가는 1.143%(FPM)와 1.036%(RPM)의 사고증가에 영향을 미치는 것으로 나타났다. 구간의 길이의 경우와 마찬가지로 노출량이 증가할수록 사고증가에도 영향을 미치는 것임을 알 수 있다. 표준편차에서 유의한 값을 나타내기 때문에 최대 1168구간에서 교통사고에 각기 다른 영향(%)을 미치는 것으로 볼 수 있다.

인터체인지 구간의 경우, 본선구간에 비해 사고발생이 작게 발생함을 알 수 있다.

차선수의 경우에는 2차선을 가진 구간을 제외한 나머지 구간에서 통계적으로 유의한 결과를 보였다. 우선 구간내에서 차선수가 많아질수록 사고발생도 증가함을 알 수 있다. 이는 구간의 길이 및 교통량과 함께 노출량이 증가하는 측면에서 사고발생에도 영향을 미치는 것으로 판단된다. 그리고, 한계효과의 관점에서는 4차선에서 5차선 비율에 따라 그 증가효과가 상대적으로 3차선에서 4차선으로의 증가 비율에 비해 작거나 미비함을 보여 구간내 4차선의 비율이 교통안전 및 투자비용의 측면에서 효율적인 것으로 판단된다.

왼쪽 길어깨 폭의 경우, 2-ft(60cm)미만의 길어깨 폭을 제외한 나머지에서 유의한 결과가 도출되었으며, 길어깨 폭이 넓어질수록 교통사고가 감소함을 알 수 있다.

3-4ft(90~120cm)와 10ft(300cm)이상의 경우에서는 RPM의 표준편차가 유의함을 알 수 있다. 우선 3-4ft는 78.9%의 구간에서 사고건수의 감소가, 나머지 21.1%의 구간에선 사고건수가 증가한다. 10ft 이상의 길어깨폭을 가진 경우에는 97%의 구간에서 사고건수의 감소가, 3%의 구간에서 사고건수가 증가하는 것으로 나타나, 운전자의 부주의 혹은 차량의 문제등으로 인한 차선이탈등의 예상치 못한 상황의 경우에도 이를 방지할수 있는 충분한 폭의 완충지역이 사고발생건수 감소에 많은 기여함을 알 수 있다. 하지만, 한계효

과측면에선 10ft 이상의 경우가 5-9ft(150~270cm)경우보다 상대 효과가 줄어드는 것으로 보이는데, 사회적 비용등의 경제적 측면에서 5-9ft의 길어깨 폭을 가진 길어깨가 더욱 효율적임을 알 수 있다.

오른쪽 길어깨 폭의 경우도, 왼쪽 길어깨 폭과 마찬가지로 2-ft 구간은 유의하지 않은 결과를 보이는 반면, 나머지 경우에는 유의한 결과가 도출되어 오른쪽 길어깨의 폭이 넓어질수록 사고발생은 줄어드는 사실을 보인다. 5-9ft폭의 경우에서 평균값과 표준편차가 통계적으로 유의함을 보여, 각 구간마다 다른 값을 가짐을 알 수 있으며, 거의 대부분의 구간(99.6%)에서 사고건수가 감소하는 것을 알 수 있다. 한계효과 측면에서는 왼쪽 길어깨 폭의 경우와 달리 10-ft이상의 길어깨 폭을 가진 구간에서는 사고감소에 대한 효과가 늘어나거나(FPM의 경우), 상대적 효과크기가 왼쪽 길어깨의 경우보다 조금 줄어듦(RPM의 경우)을 알 수 있다. 도로의 왼쪽 길어깨 폭과 달리 오른쪽 길어깨 폭에는 속도제한등의 표지판, 시설물 혹은 단속을 위한 순찰차등의 충돌사고를 유발시킬 수 있는 인자가 존재하기 때문에 넓은 어깨폭이 사고감소 효과에 큰 영향을 미칠것으로 사료된다.

구간내에 존재하는 횡단곡선의 개수의 경우, 곡선의 개수가 증가할수록 사고건수도 증가함을 보여준다. 반면, 종단 곡선의 개수는 통계적으로 유의하지 않음을 나타내어 교통사고발생에는 횡단곡선의 경우가 더 많은 영향을 주는 것을 알 수 있다.

짧은 횡단곡선의 길이가 길어짐에 따라 사고발생을 감소시키는

Table 4. Average Marginal Effects for Fixed-and Random-Parameters Negative Binomial Models

	Fixed Parameter Model	Random Parameter Model
LNLEN	7.788	4.251
LNADT	11.971	5.597
INTERCHANGE	-0.287	-0.275
NLN3	3.932	2.556
NLN4	9.454	4.625
NLN5	8.219	5.429
LSHW34	-2.858	-1.245
LSHW59	-5.193	-2.125
LSHW10	-4.359	-1.623
RSHW34	-2.998	-1.504
RSHW59	-3.343	-1.860
RSHW10	-4.134	-1.726
NHORZ	0.302	0.105
MINHLMI	-6.544	-2.777
MAXHDEL	0.066	0.026
MAXVCRVG	0.208	0.156

것으로 나타났다. 물론 구간내에서 1-mile(1.609km) 길이의 횡단 곡선의 단위증가는 현실적으로 불가능하지만, 그에 따른 한계효과 측면에서는 교통사고에 대한 감소효과를 보여준다.

구간내에 존재하는 횡단곡선의 최대교각의 경우, 교각의 크기가 증가할수록 교통사고 발생에 영향을 주는 것으로 나타났다. 이는 횡단곡선구간의 교각이 증가할수록 곡선구간의 도로 선형의 형태가 직선화됨에 따른 차량 속도의 증가에 따른 영향으로 판단된다. 따라서, 도로구간내에서 횡단곡선의 최대교각발생을 최소화하는 방향의 설계가 필요할 것이다.

최대중단구배의 경우, 절대값으로 산출되었으며, 구배가 증가할수록 사고건수도 증가함을 보여준다. 중단구배가 1°증가함에 따라서 사고발생에는 FPM에서는 0.208, RPM에서는 0.156의 증가에 영향을 미치는 것으로 나타났다.

### 4.3 모형비교

FPM 모형과 RPM 모형의 추정값 비교를 위해 3가지 방법으로 추정값을 비교하였다. RMSE를 제외한 추정에서는 미비하지만 보다 나은 추정값을 보임을 알 수 있다.

Table 5. Prediction Accuracy

	FPM	RPM
RMSE (Root Mean Square Mean)	18.01	13.13
MAPE (Mean Absolute Perchange Error)	0.00332	0.00322
MAE (Mean Absolute Error)	5.473	5.429

## 5. 결론 및 향후 연구과제

이번 연구에서는 기존에 사용되던 일반적인 음이항 모형에서 보다 발전된 형태의 확률적 모수를 고려한 음이항 모형을 제시하여 교통사고와 기하구조와의 관계를 파악하였다. 기존의 음이항 모형은 각 도로구간에 설치된 기하구조의 종류 및 설치형태가 동일하지 않음에도 불구하고, 구간의 이질성을 고려하지 못한 단점으로 인해 사고예측시 사고의 예측값에 대한 불확실성(uncertainty)와 임의성(randomness)을 내포하고 있었다. 이러한 이유로, 개선이 필요한 구간이 투자우선순위에서 밀려나, 적절한 예산이 분배되지 못하여 개선 사업이 제때 이루어지지 못하는 원인이 되기도 하였다.

이를 해결하기 위한 방법으로 이 연구에서는 구간의 이질성을 고려하였으며, 도출된 모형의 결과는 비교를 위해 기존에 사용되었던 일반적인 음이항과 함께 제시되었다. 모형의 설명력은 기존의 음이항모형에 비해 높은 것으로 나타났다. 총 16개의 교통량, 구간 길이 및 다양한 기하구조(차선수, 좌우 길어깨 폭, 총횡단 곡선)와 관련된 변수들이 사용되었다. 이 중, 8개의 변수가 각 구간마다

다른 특성을 나타내는 이질성을 가지는 것으로 나타났으며, 나머지 8개의 변수는 각 구간에 관계없이 동일한 특성을 나타내는 변수로 나타났다. 구간의 길이, 교통량, 차선수(3, 4차선), 왼쪽 길어깨폭(3-4, 10ft이상 - 90~120, 300cm이상), 오른쪽 길어깨폭(5-9ft - 150~270cm), 구간내 중단곡선의 최대구배가 이질성을 가지는 변수로 도출되었다. 차선수의 경우, 4차선에서 5차선이상으로 변하는 경우, 사고감소에 미치는 영향이 상대적으로 줄어드는 것을 알 수 있었는데, 이는 차선수의 증가에 따른 비용과 함께 그에 따른 사고비용측면에서 보다 자세한 연구가 필요할 것으로 사료된다. 이러한 특성은 길어깨 폭에서도 나타났는데, 차선수와 마찬가지로 비용-이익측면에서의 연구가 필요할 것이다. 또한, 도로건설 혹은 기하구조의 변경 시, 자연환경의 제약 등으로 인해 기준에 미치지 못하는 기하구조가 건설되어야 할 경우, 최대한의 안전성을 고려할 수 있는 기하구조의 지침이 될 수 있을 것이다.

이 연구는 안전분야에서 유의한 결과가 도출된다고 판단되는 5년 동안의 사고자료보다 훨씬 오랜 기간 동안의 사고자료가 이용되어 모형의 설명력이 높아졌음에도 불구하고 독립변수로 사용된 기하구조의 경우, 그 변화를 매년 추적하기에는 자료의 수집 등에 어려움이 있었다. 또한, 날씨시간 등에 관련된 다른 환경적인 요소 및 운전자특성(성별, 나이, 음주 등)에 관련된 자료가 적용되지 못하였다. 이 모든 자료가 수집되기에는 보다 많은 비용과 시간이 소요되겠지만, 지속적인 노력이 요구되며, 이러한 자료가 이용가능하다면, 교통사고뿐만 아니라, 비용-효과적 측면에서의 기하구조 지침에 대한 보다 자세하고 세밀한 연구에 많은 도움이 될 것이다.

## References

- Alfonso, M., Lucio, C. and Renato, L. (2008). "Crash prediction models for rural motorways." *Transportation Research Record: Journal of the Transportation Research Board*, No. 2083, pp. 180-189, Transportation Research Board, Washington D.C.
- Bhat, C. (2003). "Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences." *Transportation Research Part B*, Vol. 37, No. 1, pp. 837-855.
- Chin, H. C. and Quddus, M. A. (2003). "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections." *Accident Analysis and Prevention*, Vol. 35, pp. 253-259.
- Ciro, C., Maurizio, G. and Parisi, A. (2007). "A crash-prediction model for multilane roads." *Accident Analysis and Prevention*, Vol. 39, pp. 657-670.
- Engel, J. (1984). "Models for response data showing extra-poisson variation." *Statistical Neerlandica*, Vol. 38, No. 3, pp. 159-167.
- Jovanis, P. P. and Chang, H. L. (1986). "Modeling the Relationship of Accidents to Miles Traveled." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1068, pp. 42-51, Transportation Research Board, Washington D.C.
- Lawless, J. (1987). "Negative binomial and mixed poisson regression." *Canadian Journal of Statistics*, Vol. 15, No. 3, pp. 209-225.
- Maher, M. (1991). "A new bivariate negative binomial model for accident frequencies." *Traffic Engineering and Control*, Vol. 32, No. 9, pp. 422-433.
- Miaou, S. P. and Lum, H. (1993). "Modeling vehicle accidents and highway at geometric design relationships." *Accident Analysis and Prevention* Vol. 25, No. 6, pp. 689-709.
- Milton, J., Shankar, V. and Mannering, F. (2008). "Highway accident severities and the mixed logit model: an exploratory empirical analysis." *Accident Analysis and Prevention*, Vol. 40, No. 1, pp. 260-266.
- Noland, R. B. (2003). "Traffic fatalities and injuries: The effect of changes in infrastructure and other trends." *Accident Analysis and Prevention*, Vol. 35, pp. 599-611.
- Ogden, K. W. (1997). "The effects of paved shoulders on accidents on rural highways." *Accident Analysis and Prevention*, Vol. 29, No. 3, pp. 353-362.
- Sarath, C. J. and Nicholas, J. G. (1990). "Estimating truck accident rate and involvements using linear and poisson regression models." *Transportation Planning and Technology*, Vol. 15, No. 1, pp. 41-58.
- Shankar, V. N., Albin, R. B., Milton, J. C. and Mannering, F. L. (1998). "Evaluating median cross-over likelihoods with clustered accident counts: An empirical inquiry using random effects negative binomial model." *Transportation Research Record: Journal of the Transportation Research Board*, No. 1635, pp. 44-48, Transportation Research Board, Washington D.C.
- Train, K. (2003). *Discrete choice models with simulation*, Cambridge University Press, Cambridge, UK.
- Washington, S., Karlaftis, M. and Mannering, F. (2010). *Statistical and econometric models for transportation data analysis*, Chapman Hall/CRC, Boca Raton, Fla.
- Zhang, C. and Ivan, J. N. (2005). "Effects of geometric characteristics on head-on crash incidence on two-lane roads in connecticut." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1908, pp. 159-164, Transportation Research Board, Washington D.C.