



사고위치별 로지스틱 회귀 교통사고 모형

- 청주시 4지 신호교차로를 중심으로 -

Logistic Regression Accident Models by Location in the Case of Cheong-ju 4 - Legged Signalized Intersections

박 병 호*

양 정 모**

김 준 용***

Park, Byung Ho

Yang, Jeong Mo

Kim, Jun Young

Abstract

The goal of this study is to develop Logistic regression model by accident location(entry section, exit section, inside intersection and pedestrian crossing section). Based on the accident data of Chungbuk Provincial Police Agency(2004~2005) and the field survey data, the geometric elements, environmental factor and others related to traffic accidents were analyzed. Developed models are all analyzed to be statistically significant(chi-square $p=0.000$, Nagelkerke $R^2=0.363\sim 0.819$). The models show that the common factors of accidents are the traffic volume(ADT), distant of crossing and exclusive left turn lane, and the specific factors are the minor traffic volume(inside intersection model) and U-turn of main road(pedestrian crossing model). Hosmer & Lomeshow tests are evaluated to be statistically significant($P \geq 0.05$) except the entry section model. The correct classification rates are also analyzed to be very predictable(more than 73.9% to all models).

Keywords : accident model, independent samples t-test, hosmer & lomeshow test, logisitic regression model

요 지

본 연구의 목적은 사고위치별(유입부, 유출부, 교차로내 및 횡단보도) 로지스틱 회귀 교통사고 모형을 개발하는 것이다. 충북지방경찰청의 2004~2005년도 사고 자료와 현장조사 자료를 근거로, 교통사고와 관련된 기하구조 요소, 환경 요소 등이 분석되었다. 개발된 모형은 카이제곱 p 값은 0.000 그리고 Nagelkerke R^2 값 0.363~0.819로 모두 통계적으로 유의한 것으로 분석된다. 개발된 모형의 공통 사고요인은 교통량, 횡단거리 및 좌회전전용차로이며, 특정변수는 교차로내 사고모형의 부도로 교통량, 그리고 횡단보도 사고모형의 주도로 U턴인 것으로 나타나고 있다. Hosmer & Lomeshow 검정은 유입부를 제외한 모형들은 p 값이 0.05보다 크기 때문에 통계적으로 적합한 것으로 평가된다. 또한 정분류율 결과는 모든 모형식이 73.9% 이상으로 높은 예측력을 보이는 것으로 분석된다.

핵심용어: 사고모형, 독립표본 T검정, Hosmer & Lomeshow검정, 로지스틱 회귀모형

* 정회원 · 충북대학교 도시공학과 교수(E-mail : bhpark@chungbuk.ac.kr)

** 비회원 · 충북대학교 도시공학과 석사졸업(E-mail : cibus99@nate.com)

*** 비회원 · 충북대학교 도시공학과 석사과정

1. 서론

1.1 연구배경 및 목적

우리나라는 급격한 산업화 이후 차량의 수가 기하급수적으로 증가하여 왔다. 하지만 급증하는 자동차 이용대수에 맞춰 교통안전 여건이 뒷받침되지 않아 교통사고건수 역시 1980년대 약 12만 건에서 2004년 약 22만 건으로 지속적으로 증가해 왔다.

청주시의 교통사고는 2004년 3,155건으로 이 중 교차로에서 발생한 교통사고는 51.2%인 1,592건이었으며, 신호교차로에서 92.7%에 해당하는 1,474건이 발생하였다. 이 중 4지 신호교차로에서 발생한 사고는 1,183건으로 전체의 80.0%로 매우 높은 비율을 차지하였다.

이에 본 연구는 청주시 4지 신호교차로를 대상으로 사고위치를 4곳으로 세분하고 도로조건, 교통조건 및 교통운영 조건을 조사하여 자료를 구축한 뒤, 통계적 방법을 활용하여 교차로의 위치별 교통사고와 위의 조건들 간의 관계를 파악하고, 개발된 모형식이 통계적으로 어떠한 의미를 가지는지 분석하는데 그 목적이 있다.

1.2 연구범위 및 방법

본 연구는 도시부 신호교차로에서 발생한 교통사고현황 및 특성을 파악하고 모형을 개발하기 위해 청주시의 2004~2005년의 4지 신호교차로 181개소를 대상으로 하였다. 여기에서 발생한 교통사고 2,340건을 사고위치에 따라 유입부, 유출부, 교차로내 및 횡단보도로 구분하였고, 「충북지방경찰청의 2004~2005년도 사고관리시스템(TAMS)과 도로교통공단의 사고 충돌도」를 활용하여 교통사고 유형 및 교통조건을 분석하였다.

교통량과 차로폭, 횡단거리, 현시수, 횡단보도, 제한속도, 유턴 등과 같은 교통사고 관련 변수들은 현장조사를 통해 수집·정리하였다.

정리된 자료의 통계적 분석을 위해 SPSS 12.0K를 이용하여 로지스틱 회귀분석을 통해 사고모형을 개발하였다. 연구의 수행 흐름은 다음과 같다.

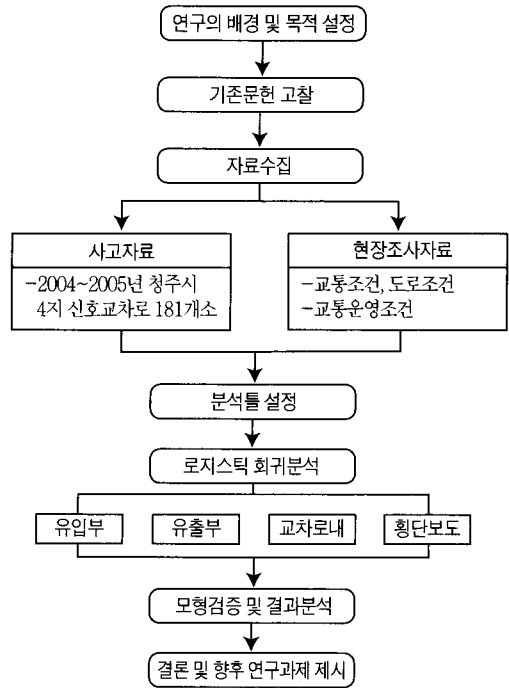


그림 1. 연구수행 흐름도

2. 기존문헌 고찰

2.1 국내연구

하오근(2005)의 연구에서는 포아송 회귀모형을 사용하여 ρ^2 (우도비), 상관계수, MAD, MPB를 알아보고 가장 적합한 결과가 도출된 모델을 주 모델로 선정하여 사고예측모형을 만들고, ρ^2 와 x^2 값으로 모형 설명력을 검증하여 사고 심각도를 분석하였다.

김숙희 등(2005)의 연구에서는 수원시의 사고다발지점을 대상으로 로지스틱 판별분석을 이용하여 사고발생 주요 요인을 추출한 바 있다.

박병호 등(2007)은 차량 방향전환 유형에 따라 사

고모형을 개발하였다. 사고모형은 과산포 검정을 통해 음이항 모형을 개발하였으며, 모형식의 공통변수는 교통량으로 나타났고 특정변수로는 직진의 경우 평균황색신호시간, 우회전의 경우 정류장유무, U턴의 경우 중차량비율이 교통사고에 영향을 미치는 요소로 제시되었다.

2.2 국외연구

Hoong Chor Chin(2003)은 기존의 포아송, 다중선형 회귀분석의 한계를 설명하고, 음이항 모형을 이용하여 싱가포르 신호교차로의 교통과 제어 특성들, 사고발생과 기하구조 사이의 관계를 분석하였다.

Yan 등(2005)은 후미추돌사고와 운전자의 특성, 도로환경 및 차량 유형으로 구분된 일련의 잠재적인 사고요인과의 관계를 조사하여 설명하였으며, 분석 방법에는 Quasi-induced exposure technique와 다중 로지스틱회귀모형을 사용하였다.

2.3 기존연구와의 차별성

본 연구는 기존의 신호교차로 사고예측 모형 연구와는 달리 종속변수를 교차로 위치(4개)별로 세분화한 점이다. 또한 교통사고모형을 도출함에 있어 기존의 포아송이나 음이항 회귀분석이 아닌 로지스틱 회귀분석을 실시하고 있다. 그 이유는 포아송이나 음이항회귀분석의 경우 다음 표와 같이 전체모형식의 설명력이 매우 낮기 때문이다.

표 1. 포아송·음이항 회귀분석 결과

모형	모형식	우도비(σ^2)
유입부	Negative binomial model	0.086
유출부	Negative binomial model	0.068
교차로내	Poisson model	0.093
횡단보도	Negative binomial model	0.085

$$\sigma^2 = 1 - \frac{\text{Log Likelihood function}}{\text{Restricted log likelihood}}$$

3. 자료 분석 및 변수 선정

3.1 수집방법

청주시 4지 신호교차로 181개 지점에 대한 사고 자료는 충북지방경찰청의 2004~2005년도 사고관리시스템(TAMS)을 이용하였다. 그리고 사고 관련 변수에 대해서는 다음 그림과 같이 교차로를 유입부(A, B, C, D), 유출부(A'', B'', C'', D''), 교차로내, 횡단보도 및 기타로 분류한 후, 각 지점별로 MS-Excel을 이용하여 자료를 수집·정리하였다.

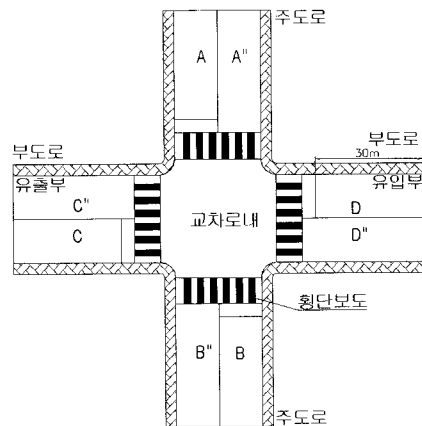


그림 2. 연구 교차로의 위치별 정의

3.2 위치별 교통사고현황

본 연구에서는 ①유입부, ②유출부, ③횡단보도, ④교차로내 및 ⑤기타로 사고위치를 구분한다. 분석 결과, 교차로내가 805건으로 33.8%, 유출부가 432건으로 18.1%, 횡단보도가 217건으로 9.1% 그리고 기타 45건으로 1.9%를 차지하고 있다.

표 2. 위치별 사고건수 현황

구분	유입부	유출부	횡단보도	교차로내	기타	합
2004년	403	205	113	439	23	1,183
2005년	483	227	104	366	22	1,202
전체	886	432	217	805	45	2,385
(비율%)	(37.1)	(18.1)	(9.1)	(33.8)	(1.9)	(100.0)



총 181개 교차로 중 교통사고 발생지점은 144개소, 교통사고 미발생지점은 37개소이다. 위치별 사고발생 현황을 보면, 사고발생 지점은 교차로내가 136개소로 가장 많고, 사고 미발생 지점은 기타가 324개소로 가장 많다.

표 3. 위치별 사고발생 현황

구분	유입부	유출부	횡단보도	교차로내	기타
사고발생	133	116	100	136	38
사고미발생	48	65	81	45	324
전체				181	

3.3 독립변수의 선정

조사된 교통량, 교차로 기하구조, 운영조건 등을 정리하여 기존문헌을 검토한 후, 총 13개의 변수를 선정하였다. 특히 이 연구에서는 기존 문헌들의 변수와는 다르게 교통량, 횡단거리 그리고 제한속도를 주도로(a~b)와 부도로(c~d)로 분류하여 정리하였다.

표 4. 선정된 독립변수

변수명	기호	정의(단위)
1. 교차로내 주도로a~b방향, 부도로c~d 일평균교통량(ADT)	$X_{1-major}$, $X_{1-minor}$	$\ln(\text{첨두시간교통량} \times \text{일평균보정계수})$ (=13.9)(대/일, pcu)
2. 차로폭	X_2	교차로 차로 폭 합
3. 주도로횡단거리	X_3	주도로의 정지선까지의 거리(m)
4. 부도로횡단거리	X_4	부도로의 정지선까지의 거리(m)
5. 횡단보도	X_5	횡단보도의 합(개)
6. 주도로 제한속도	X_6	주도로 제한속도(kph)
7. 부도로 제한속도	X_7	부도로 제한속도(kph)
8. 신호현시	X_8	신호운영에 따른 현시 수(현시)
9. 황색신호시간	X_9	황색신호 시간의 평균(sec)
10. 좌회전전용차로	X_{10}	좌회전전용차로의 유무(없음:0, 있음:1)
11. 우회전전용차로	X_{11}	우회전전용차로의 유무(없음:0, 있음:1)
12. 유턴	X_{12}	주도로 유턴 유무(없음:0, 있음:1)
13. 교통섬	X_{13}	교통섬 유무(없음:0, 있음:1)

우·좌회전 전용차로, 유턴 그리고 교통섬을 제외한 변수들은 자료가 범주형태를 띄어 유·무에 따른 이산변수(없음:0, 있음:1)로 자료를 변형하였다.

3.4 종속변수 선정

교차로의 사고위치별 분석을 위해 다음과 같이 교차로의 위치별로 4개로 종속변수를 선정하였으며, 기타지역은 교차로의 지역에서 발생한 것으로 본 연구에서 제외하였다.

표 5. 종속변수의 선정

종속변수	정의
1. 유입부사고건수(Y_1)	유입부 사고건수의 합(건)
2. 유출부사고건수(Y_2)	유출부 사고건수의 합(건)
3. 교차로내사고건수(Y_3)	교차로내 사고건수의 합(건)
4. 횡단보도사고건수(Y_4)	횡단보도 사고건수의 합(건)

4. 분석틀 설정

4.1 기술통계(descriptive statistics)

선정된 변수들의 통계적 특징을 파악하기 위해 독립변수의 기술통계를 분석하였다. 기술통계란 변수들의 평균, 분산 등의 기본적인 통계 값을 살펴봄으로써, 대표 값, 변동의 크기, 분포의 형태 등을 파악하여, 방대한 자료집합의 특징을 쉽게 알아 볼 수 있는 통계적 방법이다.

독립변수의 기술통계 분석결과, 교통량의 경우 주도로(a~b)방향의 평균이 부도로(c~d)보다 높아 주도로 교통량이 많은 것으로 나타났고, 모든 독립변수의 왜도와 첨도 값이 크지 않아 심하게 왜곡된 형태의 분포를 보이지 않음을 알 수 있다.



표 6. 독립변수의 기술통계

구분	최소값	최대값	평균	분산	왜도	첨도
$X_{1-major}$	6.30	10.55	8.08	0.75	0.72	0.28
$X_{1-minor}$	3.71	9.38	7.01	1.12	0.34	0.07
X_2	11.70	17.00	13.67	1.13	0.83	0.37
X_3	26.10	46.90	32.00	21.19	1.21	1.35
X_4	26.70	50.20	39.41	34.29	-0.20	-1.10
X_5	0.00	4.00	3.44	0.83	-1.80	2.78
X_6	30.00	80.00	54.97	91.81	0.16	2.18
X_7	30.00	70.00	46.57	71.54	-1.12	0.45
X_8	2.00	6.00	4.22	0.65	-1.00	1.42
X_9	3.00	20.00	15.31	14.63	-1.58	2.86
X_{10}	0.00	1.00	0.71	0.21	-0.95	-1.11
X_{11}	0.00	1.00	0.38	0.24	0.52	-1.75
X_{12}	0.00	1.00	0.43	0.25	0.30	-1.93
X_{13}	0.00	1.00	0.31	0.83	0.83	-1.32

4.2 정규성 검정(normality test)

정규성 검정은 관측 값들이 정규분포의 가정을 만족하는지 검정하는 방법으로 p-value가 0.05보다 크면 정규성이 있는 것으로 판단한다. 이를 위해 이 연구에서는 Kolmogorov-Smirnov 검정을 실시하였다.

표 7. Kolmogorov-Smirnov 검정결과

구분	통계량	자유도	유의확률
$X_{1-major}$	0.169	181	0.000
$X_{1-minor}$	0.089	181	0.001
X_2	0.130	181	0.000
X_3	0.119	181	0.000
X_4	0.147	181	0.000
X_5	0.371	181	0.000
X_6	0.252	181	0.000
X_7	0.453	181	0.000
X_8	0.287	181	0.000
X_9	0.267	181	0.000
X_{10}	0.449	181	0.000
X_{11}	0.405	181	0.000
X_{12}	0.379	181	0.000
X_{13}	0.438	181	0.000

검정결과 독립변수 모두 유의확률이 0.05보다 작아 정규분포를 따른다는 귀무가설을 기각하므로 정규분포를 따르지 않는다고 할 수 있다.

4.3 독립표본 t-검정

로지스틱 회귀분석은 종속변수가 이산적인 값을 나타낼 때 유용한 분석 방법으로 두 집단(사고발생, 사고미발생)의 평균 차이를 통계적으로 검정해보는데 의미가 있다.

표 8. 독립표본 t-검정(Independent Samples t-test)

독립변수	Levene의 등분산 검정		평균의 동일성에 대한 t-검정			
	F	유의확률	t	자유도	유의확률(양측)	평균차
$X_{1-major}$	57.81	0.000	6.26	53.48	0.000	1.07
$X_{1-minor}$	17.38	0.000	7.77	62.23	0.000	1.37
X_2	1.48	0.225	1.44	179.00	0.151	0.26
X_3	13.13	0.000	-5.38	155.75	0.000	-3.04
X_4	5.14	0.025	-7.63	88.97	0.000	-6.38
X_5	4.53	0.035	-0.67	65.07	0.506	-0.13
X_6	2.78	0.097	-4.61	179.00	0.000	-7.05
X_7	6.99	0.009	-2.23	76.74	0.029	-3.28
X_8	20.76	0.000	-1.49	142.99	0.139	-0.16
X_9	2.05	0.154	-0.84	179.00	0.402	-0.54
X_{10}	20.26	0.000	-7.88	66.90	0.000	-0.57
X_{11}	1.38	0.242	-0.62	179.00	0.534	-0.06
X_{12}	348.02	0.000	-8.85	163.09	0.000	-0.49
X_{13}	114.53	0.000	-4.55	131.84	0.000	-0.28

주: $X_{1-major}$, $X_{1-minor}$, X_3 , X_4 , X_5 , X_6 , X_7 , X_8 , X_9 , X_{10} , X_{11} , X_{12} , X_{13} 은 Levene의 등분산 검정결과 $P \leq 0.05$ 로 나타나 등분산이 가정되지 않음.

즉, 각 범주의 변수 값이 비슷하다면 그 변수 값에 따라 범주를 판별하기가 힘들지만, 만일 각 범주의 변수 값이 다르다면 변수 값의 대소에 따라 범주를 판별할 수 있게 된다. 이렇게 각 범주 간 평균에 대한 검정을 위해서 독립표본 t-검정을 실시하였다.



검정 결과, 13개 변수 중 9개 변수인 주도로 교통량($X_{1-major}$), 부도로 교통량($X_{1-minor}$), 주도로횡단거리(X_3), 부도로횡단거리(X_4), 주도로제한속도(X_6), 부도로제한속도(X_7), 좌회전전용차로(X_{10}), 유턴(X_{12}) 그리고 교통섬(X_{13})의 유의확률이 0.000(≤ 0.05)으로 나타나, 두 집단(사고발생과 사고미발생) 간 차이가 있는 것으로 분석되었다.

4.4 독립변수들 간 상관분석

독립변수들 간의 독립성을 보기위해 상관분석을 실시하였다. 독립변수 간의 유의한 관계, 특히 주도

표 9. 변수 간 상관관계

구 분	$X_{1-major}$	$X_{1-minor}$	X_2	X_3	X_4	X_5							
$X_{1-major}$	1	-	-	-	-	-							
$X_{1-minor}$	0.63	1	-	-	-	-							
X_2	0.09	-0.22	1	-	-	-							
X_3	-0.08	0.20	-0.32	1	-	-							
X_4	0.05	-0.30	-0.03	0.24	1	-							
X_5	-0.12	0.26	-0.25	0.27	-0.15	1							
X_6	0.04	-0.26	0.04	0.16	0.65	-0.20							
X_7	-0.02	0.01	-0.13	0.32	0.20	0.10							
X_8	0.02	0.25	0.26	0.29	0.16	0.31							
X_9	-0.04	0.27	0.21	0.35	0.03	0.36							
X_{10}	-0.20	-0.43	0.06	0.27	0.74	-0.08							
X_{11}	-0.05	0.18	-0.20	0.31	-0.02	0.19							
X_{12}	-0.14	-0.29	-0.08	0.31	0.69	-0.11							
X_{13}	-0.10	0.04	-0.16	0.43	0.16	0.12							
구 분	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}						
X_5	-	-	-	-	-	-	-						
X_6	-	-	-	-	-	-	-						
X_7	1	-	-	-	-	-	-						
X_8	0.11	1	-	-	-	-	-						
X_9	0.16	0.68	1	-	-	-	-						
X_{10}	0.18	0.13	0.04	1	-	-	-						
X_{11}	0.02	0.21	0.26	0.09	1	-	-						
X_{12}	0.22	0.07	-0.05	0.55	0.03	1	-						
X_{13}	0.10	0.20	0.25	0.24	0.42	0.20	1						

주: pearson 상관계수

로횡단거리와 다른 독립변수들 간의 유의확률이 높은 것으로 분석되었다.

5. 사고모형 개발

5.1 로지스틱 회귀분석

종속변수가 두 개의 값만을 가질 경우 일반 회귀분석에 필요한 가정들을 필연적으로 위반하게 되어 로지스틱 회귀분석이나 판별분석을 사용하게 된다.

로지스틱 회귀분석에서는 정규성과 등분산성에 대한 가정을 만족해야 하는 판별분석과는 달리 이러한 가정이 엄격히 적용되지 않는다. 오히려 판별분석에 필요한 가정이 만족될 경우에도 로지스틱 회귀모형이 보다 잘 설명하고 있음을 보여주고 있다.

본 연구에서는 이항 로지스틱 회귀분석 기법을 교통사고 확률모형에 적용함에 있어, 사고건수 자료를 "0"과 "1"의 사고 유무 값으로 변형하여 사고여부를 예측한다. 여기에서 종속변수 "1"은 사고확률이 100%임을 뜻하게 된다.

$$P(F_i = 1|X_i) = \frac{\exp[f(X_i, \beta)]}{1 + \exp[f(X_i, \beta)]} \quad (1)$$

여기서, F_i : 대상교차로 i 의 교통사고 발생($F_i = 1$) 또는 교통사고 없음($F_i = 0$)을 나타내는 종속변수

X_i : 교차로 교통사고에 영향을 미치는 독립변수

$f(X_i, \beta)$: X_i 와 파라미터 β 로 구성된 함수

위의 로지스틱 함수는 X_i 와 β 에 대하여 비선형이지만 다음과 같이 선형식으로 변환시킬 수 있다.

$$P' = \ln\left(\frac{P}{1-P}\right) \quad (2)$$

위의 식에 식(1)을 대입하면 다음과 같다.



$$P' = f(X_i, \beta) \quad (3)$$

이와 같은 변환을 로지스틱 변환이라 하고, P' 을 로짓(Logit)이라 부른다. 로지스틱 회귀분석에서는 우도(Likelihood), 즉 사건의 발생가능성을 크게 하는 최대우도추정법을 이용하여 계수를 추정한다(이용준, 2006).

5.2 유입부 사고모형

유입부 모형의 카이제곱 p 값이 0.000으로 통계적으로 유의하며, Nagelkerke R^2 값도 0.799로 추정되어, 로지스틱 모형이 변수를 잘 설명하고 있음을 보여준다.* 유입부 모형에서는 주도로교통량, 부도로교통량, 주도로횡단거리와 부도로횡단거리가 $\alpha = 0.05$ 하에서 모두 통계적으로 유의한것으로 분석된다.

표 10. 유입부 로지스틱 회귀분석 결과

독립변수	B	S.E.	Wald	자유도	유의 확률	Exp(B)
주도로교통량	1.274	0.471	7.330	1	0.007	0.280
부도로교통량	0.97	0.486	3.980	1	0.046	0.379
주도로횡단거리	0.295	0.085	11.894	1	0.001	1.343
부도로횡단거리	0.242	0.063	14.887	1	0.000	1.274
-2log likelihood : 85.322 Model Chi-square : 165.597(p-value 0.000) Nagellerke R-square : 0.799						

5.3 유출부 사고모형

유출부 모형의 카이제곱 p 값은 0.000으로 통계적으로 유의하며, Nagelkerke R^2 값도 0.531이어서, 로지스틱 모형이 변수들을 어느 정도 잘 설명하고 있음을 보여준다. 사고발생 요인은 주도로교통량, 주도로횡단거리, 부도로횡단거리와 좌회전 전용차로

이며, 모두 $\alpha = 0.05$ 하에서 통계적으로 유의한 요인으로 분석된다.

표 11. 유출부 로지스틱 회귀분석 결과

독립변수	B	S.E.	Wald	자유도	유의 확률	Exp(B)
주도로교통량	1.124	0.228	24.358	1	0.000	0.325
주도로횡단거리	0.146	0.044	10.794	1	0.001	1.157
부도로횡단거리	0.112	0.048	5.449	1	0.020	1.118
좌회전 전용차로	1.054	0.603	3.055	1	0.040	2.869
-2log likelihood : 158.981 Model Chi-square : 91.938(p-value 0.000) Nagellerke R-square : 0.531						

5.4 교차로내 사고모형

교차로내 모형의 카이제곱 p 값이 0.000으로 통계적으로 유의하며, Nagelkerke R^2 값도 0.363이므로 로지스틱 모형이 변수들을 어느 정도 설명하고 있음을 보여준다. 사고발생 요인들을 보면 부도로교통량, 주도로횡단거리와 좌회전 전용차로가 모두 통계적으로 유의한 요인으로 분석된다.

표 12. 교차로내 방정식에 포함된 변수

독립변수	B	S.E.	Wald	자유도	유의 확률	Exp(B)
부도로교통량	0.873	0.185	22.365	1	0.000	0.418
주도로횡단거리	0.175	0.043	16.354	1	0.000	1.191
좌회전 전용차로	0.987	0.425	5.407	1	0.020	2.684
-2log likelihood : 193.404 Model Chi-square : 57.516(p-value 0.000) Nagellerke R-square : 0.363						

5.5 횡단보도 사고모형

횡단보도 모형의 카이제곱 p 값이 0.000으로 통계적으로 유의하며, Nagelkerke R^2 값도 0.819이

* Nagelkerke(1991)은 Cox & Snell R-square가 "1"의 값을 가질 수 있도록 수정된 통계량을 제안하였다. 일반적으로 잘 적합한 회귀모형의 결정계수(R-square)에 비하여, Nagelkerke R-square는 비교적 작은 값을 가진다. 이런 이유로 이 통계량은 통계적 유의성 판단에 참고정보만을 줄 뿐이다.



어서 로지스틱 모형이 변수들을 매우 잘 설명하고 있음을 보여준다. 사고발생 요인은 주도로교통량, 주도로횡단거리, 좌회전 전용차, 주도로 유턴으로, 모두 $\alpha = 0.05$ 하에서 통계적으로 유의한 요인으로 분석된다.

표 13. 횡단보도 방정식에 포함된 변수

독립변수	B	S.E.	Wald	자유도	유의 확률	Exp(B)
주도로교통량	2.327	0.472	24.311	1	0.000	0.098
주도로횡단거리	0.620	0.128	23.497	1	0.000	1.860
좌회전전용차로	1.821	0.675	7.281	1	0.007	6.175
주도로 유턴	1.556	0.789	3.891	1	0.049	4.738
-2log likelihood : 78.418 Model Chi-square : 172.501(p-value 0.000) Nagellerke R-square : 0.819						

6. 사고모형 검증

로지스틱 회귀분석을 실시하여 교차로 위치별 추정된 모형식은 다음 표와 같다. 이들 모형식의 검증을 위해 로지스틱 회귀모형의 일반적인 검증방법인 Hosmer & Lomeshow 검정과 정분류율(correct classification rate) 분석을 실시하였다.

표 14. 사고위치별 로지스틱 모형

모형	모형식
유입부	$Logit = -1.274X_{1-major} - 0.97X_{1-minor} + 0.295X_3 + 0.242X_4$
유출부	$Logit = -1.124X_{1-major} + 0.146X_3 + 0.112X_4 + 1.054X_{10}$
교차로내	$Logit = -0.873X_{1-minor} + 0.175X_3 + 0.987X_{10}$
횡단보도	$Logit = -2.327X_{1-major} + 0.620X_3 + 1.821X_{10} + 1.556X_{12}$
각 모형별 사고발생확률 = $\frac{EXP(Logit)}{1 + (EXP(Logit))}$	

6.1 Hosmer & Lomeshow 검정

Hosmer & Lomeshow 검정은 적합도 검정법의 하나로 유의확률 p 값이 0.05보다 클 경우 추정된 모형

이 자료를 적절하게 적합시키지 못한다는 가설을 기각하므로 모형이 통계적으로 적합하다고 판단한다. 분석 결과, 유입부를 제외한 모형들의 p 값이 0.05보다 크기 때문에 통계적으로 적합한 모형이라 판단된다.

표 15. Hosmer & Lemeshow 검정결과

모형	카이제곱	자유도	유의확률
유입부	22.228	8	0.005
유출부	10.957	8	0.204
교차로내	7.534	8	0.480
횡단보도	14.470	8	0.070

6.2 분류 정확도

분류표의 분류행렬을 살펴보면, 사고발생 유무에 대한 관측값과 예측값이 나타나 있다. 유입부의 분류 정확도는 분류기준값을 0.5로 사용하였을 때, 92.8%로 높은 정확도를 보이고 있다.

유출부의 분류정확도는 분류기준값을 0.5로 사용하였을 때, 81.8%로 높은 정확도를 나타내고 있다. 교차로내 분류정확도는 73.9%로 비교적 높은 정확도를 보이고 있으나, 다른 모형들에 비교해 가장 낮다. 또한 횡단보도 분류정확도는 93.4%로 다른 모형에 비해 가장 높은 정확도를 보이고 있다.

표 16. 분류 정확도 결과(유입부)

분류	예측값		분류 정확(%)
	사고미발생	사고발생	
관측	사고미발생	37	77.1
	사고발생	2	98.5
전체	-	-	92.8

표 17. 분류 정확도 결과(유출부)

분류	예측값		분류 정확(%)
	사고미발생	사고발생	
관측	사고미발생	41	63.1
	사고발생	9	92.2
전체	-	-	81.8



표 18. 분류 정확도 결과(교차로내)

분류	예측값		분류 정확(%)	
	사고미발생	사고발생		
관측	사고미발생	47	34	58.0
	사고발생	15	85	85.0
전체	-	-	-	72.9

표 19. 분류 정확도 결과(횡단보도)

분류	예측값		분류 정확(%)	
	사고미발생	사고발생		
관측	사고미발생	36	9	80.0
	사고발생	3	133	97.8
전체	-	-	-	93.4

7. 결론

본 연구에서는 청주시 4지 신호교차로를 대상으로 유입부, 유출부, 교차로내 그리고 횡단보도의 4개 사고위치별 사고예측 모형을 개발하였다.

개발된 모형식의 카이제곱 p 값이 0.000으로 통계적으로 모두 유의하며, Nagelkerke R^2 값이 0.363~0.819로 추정되어, 로지스틱 모형이 변수들을 잘 설명하는 것으로 분석되었다.

위치별 사고모형의 변수들을 분석한 결과, 교통량, 횡단거리 및 좌회전 전용차로가 $\alpha = 0.05$ 하에서 통계적으로 유의하여, 사고위치별 사고에 공통적으로 영향을 미치는 변수로 분석된다. 또한 교차로 내 사고모형은 부도로 교통량이, 횡단보도 사고모형은 주도로 유턴이 특정변수로 분석된다.

추정된 모형식의 검증을 위해, Hosmer & Lomeshow 검정과 정분류율을 분석하였다. Hosmer & Lomeshow 검정결과, 유입부를 제외한 모형들의 p 값이 0.05보다 크기 때문에 통계적으로 적합한 모형인 것으로 분석되었다. 또한 정분류율 결과 모든 모형식이 73.9% 이상으로 높은 예측력을 보이는 것으로 평가되었다.

감사의 글

이 논문은 2008년도 충북대학교 학술연구지원사업의 연구비지원에 의하여 연구되었음.

참고문헌

- 김순귀 · 정동빈 · 박영술(2003), 로지스틱 회귀모형의 이해와 응용, 자유아카데미.
- 김숙희 · 장정아 · 최기주(2005), “사고다발지점의 안전 성능 진단 및 위치별 사고요인분석(수원시를 중심으로)”, 대한교통학회지, 대한교통학회, 제23권 제1호, pp. 9~20.
- 하오근(2005), “국도변 신호교차로 안전성 향상을 위한 사고예측모형개발과 심각도 분석”, 한양대학교 대학원 석사학위 논문.
- 박병호 · 한상욱 · 양정모 · 박정순(2007), “차량 방향전환 유형에 따른 사고모형”, 국토계획, 대한국토 도시계획학회, 제42권 제7호, pp. 183~194.
- Hoong Chor Chin·ohammed Abdul Quddus (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections, *accident analysis & prevention* 35, pp. 253~259.
- James A, Bonneson and Patrick T, McCoy, Estimation of safety at two-way stop-controlled intersections on rural highways, *TRR* 1401.
- Xuedong Yan·Essam Radwan·Mohamed Abdel - Aty(2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model, *accident analysis & prevention* 37, pp. 35~4.

접 수 일: 2008. 8. 25
 심사 일: 2008. 8. 28
 심사완료일: 2009. 4. 22