

로지스틱모형을 이용한 가로구간 사고모형

박병호[†] · 임진강 · 한수산

충북대학교 도시공학과

(2010. 2. 9. 접수 / 2010. 7. 23. 채택)

Accidents Model of Arterial Link Sections by Logistic Model

Byung Ho Park[†] · Jin Kang Lim · Su San Han

Department of Urban Engineering Chungbuk National University

(Received February 9, 2010 / Accepted July 23, 2010)

Abstract : This study deals with the accident model of arterial link section in Cheongju. The objective is to develop the accident model of arterial link section using the logistic regression. In pursuing the above, the study uses the 258 accident data occurred at the 322 arterial link section. The main results are as follows. First, Nagellerke R^2 of developed accident model is analyzed to be 0.309 and t-values of variable that explains goodness of fit are evaluated to be significant. Second, the variables adopted in the model are AADT, the number of exit and entry. These variables are all analyzed to be statistically significant. Finally, the analysis of correct classification rate shows that the total accident of correct classification rate is analyzed to be 72.7% at the arterial link section.

Key Words : arterial link section, accident model, logistic regression model, correct classification rate

1. 서론

1.1. 연구의 배경 및 목적

우리나라는 급격한 산업화 이후 차량의 수가 기하급수적으로 증가하여 왔다. 하지만 급증하는 자동차 이용대수에 맞춰 교통안전 여건이 뒷받침되지 않아 교통사고 건수 역시 지속적으로 증가해 왔다. 이에 맞춰 도시 내 교차로 사고에 관한 연구는 다각적으로 진행되어 왔지만, 가로구간 사고는 발생 건수가 교차로 사고보다 적고, 그 대상범위가 넓어 교차로 사고연구에 비해 연구가 미비한 실정이다.

이에 본 연구는 청주시 가로구간의 사고를 대상으로 사고자료와 기하구조자료를 조사하여 로지스틱모형을 이용한 사고모형을 개발하는데 목적이 있다.

1.2. 연구의 내용 및 수행과정

이 연구는 청주시의 2007년 13개 주간선 도로의 교통사고를 대상으로 하며, 사고자료는 청주시 가로구간 주간선도로 13개 구간을 세분화하여 사용

하였고, 기하구조 자료는 현장조사를 통해 수집하였다.

본 연구의 수행절차는 다음과 같다.

제1단계는 관련문헌 고찰로서 가로구간의 교통사고와 관련된 연구논문 및 분석방법론을 검토하여 연구의 방향을 설정한다.

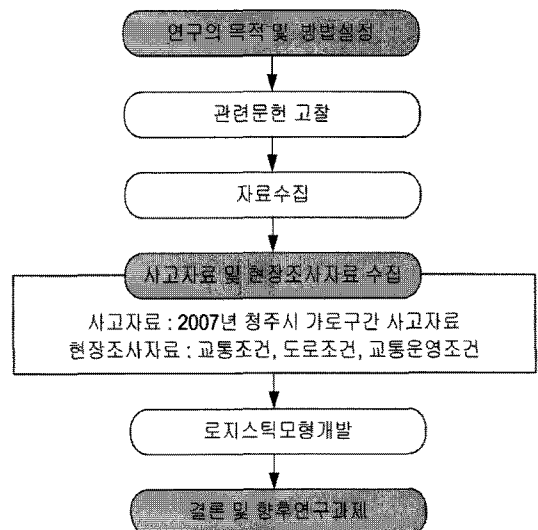


Fig. 1. Flow chart of this study.

[†] To whom correspondence should be addressed.
bhpark@chungbuk.ac.kr

제2단계는 청주시 가로구간 중 사고자료와 기하구조자료가 수집가능한 구간선도로 13개 구간을 대상으로 사고자료 및 기하구조자료를 정리한다.

제3단계는 분석틀의 설정단계로 조사수집된 자료를 바탕으로 SPSS 통계패키지를 이용하여 사고자료를 분석한다.

마지막 단계는 결론을 정리하고 본 연구의 미진한 부분들과 향후 연구과제를 제시한다.

2. 기존연구 고찰

2.1. 로지스틱 모형

종속변수가 두 개의 값을 가질 경우 일반 회귀 분석에 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 사고유무 값으로 변형하여 사고 여부를 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($F_i = 0$) 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 같이 변환을 로지스틱 변환이라 하고, P' 을

로짓(logit)이라 부른다. 로지스틱 회귀분석에서는 우도(likelihood), 즉 사건의 발생가능성을 크게 하는 최대우도추정법을 이용하여 계수를 추정한다.

2.2. 국·내외 관련연구

Hoong Chor Chin¹⁾은 기존의 포아송, 다중선형 회귀분석의 한계를 설명하고, 음이항 모형을 이용하여 싱가포르 신호교차로의 교통과 제어 특성등, 사고발생과 기하구조 사이의 관계를 분석하였다.

Xuedong Yan 등²⁾은 후미추돌사고와 운전자의 특성, 도로환경 및 차량유형으로 구분된 일련의 잠재적인 사고요인과의 관계를 조사하여 설명하였으며, 분석 방법에는 Quasi-induced exposure technique와 다중로지스틱회귀모형을 사용하였다.

박병호 등³⁾은 사고 위치별 로지스틱 회귀 교통사고모형에서 청주시 4지 신호교차로의 교통사고 자료를 바탕으로 사고모형을 개발하였으며, 모형에 사용된 변수로는 교통량, 좌회전 전용차로수 등이 있다.

김숙희 등⁴⁾은 사고다발지점의 안전성능 진단 및 위치별 사고요인분석에서 수원시의 사고다발지점을 대상으로 로지스틱 판별분석을 이용하여 사고 발생 주요 요인을 추출하였다.

오주택 등⁹⁾은 국도변 신호교차로 안전성 향상을 위한 사고예측모형개발과 심각도 분석에서 포아송 회귀모형을 사용하여 ρ^2 , 상관계수, MAD, MPB를 알아보고 가장 적합한 결과가 도출된 모델을 주 모델로 선정하여 사고예측모형을 만들고, 모형을 검증하여 사고 심각도를 분석하였다.

홍정렬⁵⁾은 원주시의 2001년 64개 신호교차로의 기하구조 도면과 현장조사를 통하여 얻어진 자료를 이용해 각 조사대상 변수들이 사고에 얼마나 영향을 주는지를 분석하였으며, 교차로 사고에 영향을 주는 요소들을 선택하여 비선형회귀모형을 사용해 사고모형을 개발하고 그에 따른 교차로 위험수준을 결정하였다.

이수범 등¹⁰⁾은 지방지역의 도로특성 및 사고특성 조사를 통해 도로유형별로 물리적인 특성을 반영하여 보다 체계적인 도로등급과 그 특성에 따른 사고모형을 개발하기 위해 분석구간을 세 가지 유형으로 선정하였다. 자료는 전북지역 2001년과 2002년, 2년간의 지방지역별 사고자료를 수집하여 다중회귀분석을 통해 사고모형식을 도출하였으며, 모형을 실제자료로써 검증하였다.

오주택 등⁹⁾은 충청남·북도의 교차로 중 사고가

많은 3지, 4지 신호교차로에 대해 계층적 샘플링방법을 통해 77개의 조사지점을 선정하였다. 또한 교차로 교통사고에 영향을 미치는 24개 항목을 정하고 이를 다시 50개의 변수로 재분류하였으며, 포아송회귀모형을 이용하여 사고모형을 개발하였다.

이주연 등⁶⁾은 구조방정식을 이용한 고속도로 교통사고 심각도 분석에서 요인들을 도로요인, 운전자요인, 환경요인으로 구분하고, 구조방정식 모형을 이용하여 사고심각도를 분석하고 있다.

박정순 등⁷⁾은 신호교차로 측면직각 충돌사고 특성과 심각도에서 교차로의 교통사고를 대상으로 순서형 프로빗 모형을 이용하여 사고심각도 모형을 개발하였다.

김태영 등⁸⁾은 트럼펫 IC 램프의 운전조건과 교통사고 분석에서 고속도로 트럼펫 IC의 교통사고와 기하구조와의 관계를 분석하고 이를 바탕으로 사고모형을 개발하였다.

2.3. 연구의 차별성

본 연구의 차별성은 다음과 같다. 첫째, 교통사고 특성 분석을 다루고 있는 기존 연구들은 주로 교차로 사고를 다루고 있고, 가로구간 논문 역시 세부구간으로 나누어 분석하고 있지 않는 반면에, 본 연구는 도시 내 가로구간을 세부구간으로 나누어 사고 발생 유·무를 분석하고 있다. 둘째, 종속변수를 사고건수로 선정하여 로지스틱 회귀모형을 개발하는 점에서 기존연구와의 차별성이 있다.

3. 자료분석 및 변수 선정

3.1. 자료정리 및 분석

청주시 13개 주간선도로를 연구의 대상지역으로 하고 있으며, 2007년 청주시 주간선도로의 사고자료는 도로교통공단 사고통계자료를 이용하였다.

주간선도로를 교차로와 교차로 사이의 세부구간으로 나누어 사고자료를 조사정리하였다. 2007년 청주시 주간선도로의 총 사고건수는 258건이고, 322개 세부구간 중 사고가 발생한 지점은 125개 지점으로 나타났다.

3.2. 변수선정

조사된 교통량, 가로구간 관련변수 등을 정리하여 기존문헌을 검토 한 후, 총 8개의 변수를 선정하였다. 특히 이 연구에서는 기존문헌들의 변수와는 다르게 가로구간 전체의 변수가 아닌 세부구간

Table 1. Present condition of arterial link section and number of accident

Classification	Number of link section	Number of accident	Number of accident/ Number of link section
Uiam	14	11	0.79
Sajik	20	40	2.00
Mochung	14	5	0.36
Cheongnam	22	13	0.59
Sangdang	22	25	1.14
Heungdeok	34	14	0.41
Konghang	12	13	1.08
Danjae	16	11	0.69
Garosugil	10	26	2.60
West-bypass	20	2	0.10
East-bypass	26	13	0.50
Beltway 1	54	62	1.15
Beltway 2	58	23	0.40
Total	322	258	0.80

Table 2. Dependent variables

Variables	Abbreviation	Variable description
1. AADT	X_1	Annual Average Daily Traffic
2. Crosswalk	X_2	Number of crosswalk
3. Lane	X_3	Number of average lane
4. Entry and Exit	X_4	Length of entry and exit
5. Median	X_5	Length of median(m)
6. Link section	X_6	Length of link section(m)
7. Busstop	X_7	Number of busstop
8. Busbay	X_8	Length of busbay(m)

으로 나누어 변수들을 분류하여 정리하였다.

4. 분석틀의 설정

4.1. 정규성 검정

정규성 검정은 관측값들이 정규분포의 가정을 만

Table 3. Results of Kolmogorov-Smirnov test

Variables	Mean	Kolmogorov-Smirnov의 Z	P(> Z)
X_1	1522.6739	1.51	0.021
X_2	1.8447	4.128	0.000
X_3	2.3447	5.663	0.000
X_4	2.3447	3.32	0.000
X_5	116.5031	7.012	0.000
X_6	435.6142	3.394	0.000
X_7	0.8106	4.448	0.000
X_8	4.8665	9.027	0.000

축하는지 검정하는 방법으로 유의확률이 0.05보다 크면 정규성이 있는 것으로 판단한다. 이를 위해 이 연구에서는 Kolmogorov-Smirnov 검정을 실시하였다. 검정결과 독립변수 모두 유의확률이 0.05보다 작아 정규분포를 따른다는 귀무가설을 기각하므로 정규분포를 따르지 않는다고 할 수 있다.

4.2. 독립표본 t-검정

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

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. 검정결과, 8개 변수 중 교통량(X_1), 횡단보도(X_2), 차로수(X_3), 진출입구수(X_4), 정류장수(X_7)의 90% 유의수준에서 두 집단 간의 차이가 있는 것으로 분석되었다. 이렇게 분석된 사고발생과 미발생, 두 집단 간의 독립표본 t-검정은 로지스틱 사고모형을 개발할(제5장)에 있어, 독립변수를 선정하는 기준으로 활용하게 된다.

4.4. 독립변수간 상관분석

어떠한 변수가 모형에 선정되는가에 따라 모형의 설명력에 큰 차이가 난다. 따라서 변수들 간의 상관분석을 통해 유의한 변수를 찾아야 하며, 상관관계분석결과 변수들 간의 상관성이 크게 떨어질 경우 자료의 신빙성이 저하되는 문제가 제기 될 수 있다.

본 연구는 모형개발에 앞서 독립변수들 간의 다중공선성 문제를 해결하기 위해 상관분석을 실시

Table 4. Independent sample t-test

Variables	t-value	Degree of freedom	P(> t)	Difference of mean
X_1	-2.783	320	0.006	-0.13198
X_2	-3.941	320	0.000	-0.46303
X_3	1.729	320	0.085	0.11886
X_4	-9.086	320	0.000	-1.86116
X_5	0.661	320	0.509	23.11716
X_6	-0.021	320	0.983	-0.78595
X_7	-3.472	320	0.001	-0.32272
X_8	0.123	320	0.902	0.18709

Table 5. Results of correlation analysis

Variables	X_1	X_2	X_3	X_4
X_1	1.000	0.031	0.060	0.113
X_2	0.031	1.000	-0.092	0.480
X_3	0.060	-0.092	1.000	-0.084
X_4	0.113	0.480	-0.084	1.000
X_5	0.086	-0.197	0.034	-0.025
X_6	0.089	0.109	0.042	0.199
X_7	0.139	0.300	-0.012	0.413
X_8	0.031	0.053	-0.002	-0.008
Variables	X_5	X_6	X_7	X_8
X_1	0.086	0.089	0.139	0.031
X_2	-0.197	0.109	0.300	0.053
X_3	0.034	0.042	-0.012	-0.002
X_4	-0.025	0.199	0.413	-0.008
X_5	1.000	0.747	-0.135	-0.099
X_6	0.747	1.000	0.128	-0.035
X_7	-0.135	0.128	1.000	0.313
X_8	-0.099	-0.035	0.313	1.000

하였으며, 신뢰수준을 95%($\alpha=0.05$)로 하여 Pearson 상관계수를 통해 변수들 간의 상관성을 분석하였다.

분석결과, X_2 와 X_4 , X_2 와 X_7 , X_4 와 X_7 및 X_5 와 X_6 의 상관도가 높기 때문에, 두 변수간의 다중공선성 문제를 검토하였다. 횡단보도가 있는 지점에 버스 정류장이 있는 경우와 횡단보도와 진출입구가 있는 가로구간의 경우에 다중공선성의 문제가 제기될 수 있다. 따라서 모형 개발에서 즉, 횡단보도 수(X_2), 진출입구 수(X_4) 및 버스정류장(X_7)변수가 동시 채택되는 경우를 배제하는 것이 바람직한 것으로 판단된다.

5. 가로구간 사고모형

5.1. 로지스틱 회귀분석

독립표본 t-검정결과에서 두 집단 간의 차이나는 변수를 우선 가로구간 사고모형에 채택 가능한 변수로 고려하였고, 그 변수들 중 상관분석 결과에서 나타난 다중공선성의 문제를 가진 변수를 제거한 후에 모형을 구축하였다.

가로구간 사고모형의 Nagellerke R^2 의 값이 0.309 인 모형이 개발되었으며, 로지스틱 모형의 변수들의 유의확률이 0.000~0.074로 유의수준 90%에서 유의한 것으로 나타났다. 또한 교통량(X_1) 및 진출입구수(X_4)가 증가할수록 사고가 발생하는 모형이 개발되어 모형의 계수 부호도 적절한 것으로 판단

Table 6. Results of logistic regression model

Variables	B	S.E.	Wald	Degree of freedom	P(> t)	Exp(B)
X ₁	0.001	0.000	24.403	1	.000	1.001
X ₄	0.474	0.075	40.023	1	.000	1.606
Constant	-3.106	0.397	61.304	1	.000	0.045

-2log likelihood : 346.867
Nagellerke R-square : 0.309

Logit = -3.106 + 0.001X₁ + 0.474X₄

Table 7. Results of corrected classification rate

Classification		Prediction		Correct classification rate(%)
		No accident	Accident	
Observation	No accident	169	28	85.8
	Accident	60	65	52.0
Total		-	-	72.7

된다. 또한 독립변수들 간의 상관관계가 낮아 다중공선성 문제도 해결하였다.

개발된 사고모형의 결과와 다중선형회귀 사고모형을 비교하였다. 다중선형회귀모형에서는 신뢰수준 95%에서 교통량(X₁)과 중앙분리대 길이(X₅)가 모형의 변수로 채택되었고, 개발된 다중선형회귀 사고모형의 R²는 0.236로 분석되었다. 그렇지만 중앙분리대 길이변수의 상수값이 0.090으로 중앙분리대가 길수록 사고가 많이 나는 일반적인 상식에 어긋나는 모형이어서 이 논문에서 개발된 사고모형이 더 적합한 것으로 평가된다.

5.2. 분류 정확도

분류표의 분류행렬을 살펴보면, 사고발생 유무에 대한 관측값과 예측값이 나타나있다. 가로구간 사고의 정확도는 분류 기준값을 0.5로 사용하였을 때, 전체 분류 정확도는 72.7%로 분석되었다. 사고 미발생 건수에 대한 예측은 높은 것으로 나타났다.

6. 결론

본 연구에서는 로지스틱 회귀모형을 이용하여 청주시 주간선도로 가로구간을 대상으로 사고예측모형을 개발하였다.

첫째, 개발된 가로구간 사고모형은 Nagellerke R²의 값이 0.309이고, 변수들의 t값이 유의수준 90%에서 유의한 모형이 개발되어, 로지스틱 모형이 변수들을 잘 설명하는 것으로 분석된다. 둘째, 모형

에 선정된 변수들은 교통량(X₁) 및 진출입구수(X₄)이며, 이 변수들은 모두 통계적으로 유의한 것으로 평가되고, 변수들 간의 다중공선성 문제도 없는 것으로 판단된다. 마지막으로, 가로구간 전체사고 분류정확도는 72.7%로 나타나며, 사고 미발생구간의 분류정확도가 발생구간의 정확도보다 높아 미발생구간의 예측력이 높은 것으로 분석된다.

아직까지 가로구간의 사고예측모형 개발이 미비한 실정에서 이 논문의 로지스틱 사고예측모형은 어떤 가로구간이 위험할지 예측하여 가로구간의 위험도를 측정할 수 있는 하나의 기준으로 활용가능하다.

연구의 한계점은 사고 발생구간의 정확도가 미발생구간의 정확도보다 낮은 점이며, 향후 연구로는 이 연구의 한계점을 보완하는 차원에서 사고발생구간의 예측력을 높일 수 있는 변수추가 및 모형의 개발이 필요할 것으로 판단된다.

참고문헌

- 1) Hoong Chor Chin, Mohammed Abdul Quddus, "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections", *accident analysis & prevention* 35 pp. 253~259, 2005.
- 2) Xuedong Yan, Essam Radwan, Mohamed Abdel -Aty, "Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model", *accident analysis & prevention* 37, pp. 35~46, 2005.
- 3) 박병호, 양정모, 김준용, "사고위치별 로지스틱 회귀 교통사고 모형", 한국도로학회지, 제11권, 제2호, pp. 17~25, 2009.
- 4) 김숙희, 장정아, 최기주, "사고다발지점의 안전성 진단 및 위치별 사고요인분석", 대한교통학회지, 제23권, 제1호, pp. 9~20, 2005.
- 5) 홍정열, "신호교차로에서의 사고 예측모델개발 및 위험수준결정 연구", 한양대학교 석사학위논문, 2002.
- 6) 이주연, 정진혁, 손봉수, "구조방정식모형을 이용한 고속도로 교통사고 심각도 분석", 대한교통학회지, 제26권, 제2호, pp. 17~24, 2008.
- 7) 박정순, 김태영, 유두선, "신호교차로 측면직각 충돌사고 특성과 심각도(청주시 4지 신호교차로를 중심으로)", 대한교통학회지, 제25권, 제2호, pp. 63~72, 2007.
- 8) 김태영, 박병호, "트럼펫 IC 램프의 운전조건과

- 교통사고 분석”, 대한교통학회지, 제25권, 제1호, pp. 73~79, 2007.
- 9) 오주택, 성낙문, 하오근, “국도변 신호교차로 안전성 향상을 위한 사고예측모형개발”, 대한토목학회논문집, 대한토목학회, 제25권, 제1호 pp. 9~15, 2005.
- 10) 이수범, 김정현, 홍다희, 유창남, “도로등급 및 특성에 따른 교통사고예측모형개발”, 대한토목학회논문집, 대한토목학회, 제23권, 제4호, pp. 133~144, 2003.