

가산자료모형을 이용한 고속도로 입체교차구간의 사고모형 개발

Accident Model Development of Korea Expressway Intersections
by Count Data Models

김 윤 환

(충북대학교 도시공학과 박사과정)

박 병 호

(충북대학교 도시공학과 교수)

김 태 영

(충북대학교 도시공학과 박사과정)

목 차

- | | |
|---|---|
| I. 서론
1. 연구 배경 및 목적
2. 연구 방법
II. 문헌고찰
1. 사고모형의 이론적 고찰
2. 국내외 모형개발 사례
III. 자료 및 기술통계
1. 분석자료
2. 기술통계분석 | IV. 사고모형개발
1. 모형분석 설정
2. 포아송회귀모형(PRM)
3. 음이항회귀모형(NBRM)
4. ZIP 모형
5. ZINB 모형
V. 결론 및 향후 연구과제
참고문헌 |
|---|---|

I. 서론

1. 연구의 배경 및 목적

본 연구는 목적은 고속도로 입체교차구간의 기하구조와 교통사고와의 관계를 설명하기위해 교통사고모형 개발의 최근 경향을 반영한 사고모형을 개발하는데 있다.

사고모형의 연구경향은 선형회귀분석에서 벗어나 비선형회귀분석이나 다양한 통계적 방법들에 의해서 이루어지고 있다. 최근 자료특성에 따라 음이항 분포에 의한 비선형 모형(성낙문, 2002), GLM 이용한 모형(Greibe, 2003), 교차로 사고빈도의 음이항회귀모형(Poch, Mark and Mannering, Fred, 2004) 등이 개발된 바 있다.

또한 포아송회귀모형, 음이항회귀모형(오주택, 성낙문, 하오근, 2005), 0의 값이 과다한 포아송회귀분석(ZIP : Zero-inflated poisson regression) 및 0의 값이 과다한 음이항회귀분석(ZINB : Zero-inflated negative binomial regression)을 적용한 모형(Lee and Mannering, 2002; 이인성, 2004) 등이 사고모형 개발에 이용되고 있다.

류승욱·박병호(2006)는 '고속도로 교차점의 기하구조와 교통사고의 다중회귀분석'에서 일반

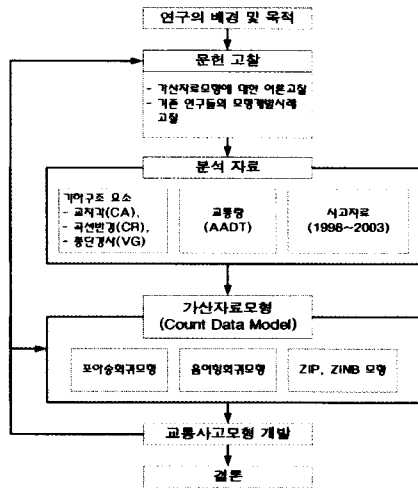
적인 다중선형회귀분석 및 비선형회귀분석에 의한 모형을 검토한 바 있다. 그러나 연구대상인 전국 고속도로 교차점 195개에 대한 사고모형 개발에서 전술한 다중선형 및 비선형회귀분석의 적합도가 적절하지 못한 문제를 가지고 있으며, 개발된 모형 또한 교차점 램프 구간에 대한 사고모형만을 개발한 바 있다.

이에 본 연구에서는 우리나라 고속도로 입체교차구간의 기하구조 및 사고자료를 바탕으로 통계적 모형 중 가산자료모형(count data model)을 적용하여 좀 더 설명력 있는 사고모형을 개발하고자 한다. 이는 전국 고속도로의 일부 구간 및 지점이 아닌 전체 입체교차구간을 대상으로 보다 적절한 통계적 모형을 이용하여 사고모형을 개발한다는 것에 그 의의가 있다.

2. 연구 방법

기존 문헌을 통해 최근 연구경향과 이론을 고찰하여, 고속도로 입체교차구간에서 발생하는 사고자료의 특성을 잘 반영할 수 있는 방법을 검토한다.

사고모형의 개발은 통계패키지 LIMDEP 8.0의 Count Data Model을 이용하여 분석한 전반적 연구과정에 대한 사항은 그림 1과 같다.



<그림 1> 연구수행절차

II. 문헌고찰

1. 사고모형의 이론적 고찰

본 연구에서 적용하고자 하는 가산자료모형¹⁾은 크게 포아송회귀모형, 음이항회귀모형으로 구분할 수 있다. 가산자료모형의 가장 일반적인 모형은 포아송회귀모형이며, 포아송회귀모형에서 발생하는 과분산 문제(dispersion problem)를 해결하기 위해 음이항회귀모형이 개발되었다. 그 밖에 포아송회귀모형에서 발생될 수 있는 분석 오류에 대한 대책으로 절단된 포아송회귀모형(Truncated poisson regression model), 0값이 과다한 포아송회귀모형과 음이항회귀모형(ZIP와 ZINB) 등이 있다.

<표 1> 가산자료모형의 종류 및 적용

구분	내용
포아송회귀모형	가장 일반적 가산자료 모형
절단된 포아송회귀모형	관찰된 자료가 절단된 경우 적용
음이항회귀모형	포아송분포의 가정한 평균과 분산이 같지 않고, 평균보다 분산이 클 경우(과분산) 적용
ZIP, ZINB	0의 관찰값이 과다한 경우 적용

1) 포아송회귀모형(PRM : Poisson regression model)

포아송 회귀모형은 사고건수를 이산적 확률 변수로 해석하는 방법으로 분산과 평균이 같다는 전제하에 적용하되 일반식은 다음과 같다.

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

1) 가산자료모형에 대한 이론적 고찰은 Washington, Simon P., et al.(2003), 이기영·이용택(2004)의 연구내용을 참고한다.

여기에서 $P(y_i)$ 는 교차점 i 에 대한 사고발생 확률, y_i 는 회/년, 그리고 λ_i 는 교차점 i 에 대한 포아송 계수이다.

포아송 회귀분석은 설명변수들의 함수가 되는 특정한 포아송 계수 λ_i 에 의해 조정된다.

$$\ln(\lambda_i) = \beta X_i \quad (2)$$

여기에서 X_i 는 설명변수들의 벡터, β 는 추정 가능한 계수들의 벡터이다.

포아송 회귀계수는 식(1)과 (2)에 의해서 우도함수로써의 표준최대우도법에 의해 추정된다.

$$L(\beta) = \prod_i \frac{\exp[-\exp(\beta X_i)] [\exp(\beta X_i)]^{y_i}}{y_i!} \quad (3)$$

log-우도 함수,

$$LL(\beta) = \sum_{i=1}^n [-\exp(\beta X_i) + y_i \beta X_i - \ln(y_i!)] \quad (4)$$

포아송회귀모형의 적합도는 우도비검정과 모형편차의 합(G^2), 표준화된 잔차에 근거하는 R_p^2 및 전체 모형의 적합성을 측정하는 ρ^2 통계치로 이루어진다.

$$\text{우도비 검정} : \chi^2 = -2[LL(\beta_R) - LL(\beta_U)] \quad (5)$$

여기에서 $LL(\beta_R)$ 는 제약된 모형 수렴에 있어서의 로그-우도값, $LL(\beta_U)$ 는 제약되지 않는 모형 수렴에 있어서의 로그-우도값이다.

$$\text{모형편차의 합} : G^2 = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\hat{\lambda}_i} \right) \quad (6)$$

$$\text{결정계수} : R_p^2 = 1 - \frac{\sum_{i=1}^n \left[\frac{y_i - \hat{\lambda}_i}{\sqrt{y_i}} \right]^2}{\sum_{i=1}^n \left[\frac{y_i - \bar{y}}{\sqrt{\bar{y}}} \right]^2} \quad (7)$$

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(0)} \quad (8)$$

여기에서 $LL(\beta)$ 는 계수 벡터 β 로 수렴된 로그 우도이다. $LL(0)$ 는 초기의 로그 우도(모든 계수를 0에 설정)이다.

2) 음이항회귀모형(NBRM : Negative binomial regression model)

일반적 분석 오류는 $E[y_i] = VAR[y_i]$ 일때, 평균과 분산이 같도록 제약되는 포아송분포의 특성을 만족

시키지 못한 결과이다. 만약 이 동질성이 유효하지 않다면, 자료는 저분산(under-dispersed); $E[y_i] > VAR[y_i]$ 또는 과분산(over-dispersed); $E[y_i] < VAR[y_i]$ 되었다고 말한다. 그리고 그 계수벡터는 만약 정확하게 측정되지 않으면 편기(biased)된다. 음이항 모형은 식(2)를 재작성함에 의해 유도되며, 각 관찰 i 에 대한 음이항 모형은 다음과 같다.

$$\lambda_i = \text{EXP}(\beta X_i + \epsilon_i) \quad (9)$$

여기에서 $\text{EXP}(\epsilon_i)$ 는 평균 1과 분산 α^2 를 가지는 감마분포의 오차항이다.

$$VAR[y_i] = E[y_i][1 + \alpha E[y_i]] = E[y_i] + \alpha E[y_i]^2 \quad (10)$$

포아송회귀모형은 계수 α 가 0에 접근할 때 음이항회귀모형의 한계모형으로서 간주된다. 이들 두 모형 사이에 선택된 평균은 α 값에 종속된다. 계수 α 는 종종 과분산 계수(overdispersion parameter)로써 언급된다. 음이항분포는 다음의 형태이다.

$$P(y_i) = \frac{\Gamma((1/\alpha) + y_i)}{\Gamma(1/\alpha)y_i!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i} \right)^{1/\alpha} \left(\frac{\lambda_i}{(1/\alpha) + \lambda_i} \right)^{y_i} \quad (11)$$

여기에서 $\Gamma(\cdot)$ 는 감마함수이다. 이것은 우도 함수에서의 결과이다.

$$L(\lambda_i) = \prod_i \frac{\Gamma((1/\alpha) + y_i)}{\Gamma(1/\alpha)y_i!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i} \right)^{1/\alpha} \left(\frac{\lambda_i}{(1/\alpha) + \lambda_i} \right)^{y_i} \quad (12)$$

자료가 과분산 될 경우, 추정된 분산 항은 포아송회귀분석의 분산 보다 크다. 과분산이 크게 되었을 때, 결정적으로 계수 추정치 모든 표준 오차는 과다 추정된다.

과분산에 대한 검정은 $(y_i - E[y_i])^2 - E[y_i]$ 이 평균 0을 가지는 포아송모형 하의 가정을 기반으로 Cameron and Trevedi(1990)에 의해 제시되었다. 여기에서 $E[y_i]$ 는 예측된 가산 $\hat{\lambda}_i$ 이다. 따라서 다음과 같이 귀무 및 대립가설을 생성시킨다.

$$\begin{aligned} H_0: VAR[y_i] &= E[y_i] \\ H_A: VAR[y_i] &= E[y_i] + \alpha g(E[y_i]) \end{aligned} \quad (13)$$

여기에서 $g(E[y_i])$ 는 $g(E[y_i]) = E[y_i]$ 또는 $g(E[y_i]) = E[y_i]^2$ 의 가장 빈번하게 주어지는 값인

예측된 가산(counts)의 함수이다. 이 검정을 수행하기 위해, 단순선형회귀분석은 여기서 Z_i 는 W_i 에 대하여 회귀되는 것으로 추정된다.

$$Z_i = \frac{(y_i - E(y_i))^2}{E(y_i) \sqrt{2}} \quad (14)$$

$$W_i = \frac{g(E(y_i))}{\sqrt{2}} \quad (15)$$

$g(E[y_i]) = E[y_i]$ 와 $g(E[y_i]) = E[y_i]^2$ 를 가지는 회귀분석($Z_i = bW_i$)를 실행 후에 만약 b 가 각 경우에 있어서 통계적으로 유의하다면 그뎨 귀무가설(H_0)은 관련된 함수 g 에 대하여 기각된다.

3) ZIP 및 ZINB 모형

ZIP 및 ZINB 모형은 앞서 언급한 것처럼 포아송회귀모형 및 음이항회귀모형의 확장된 형태이다. 관찰된 값에서 0값이 과다한 비율을 차지하고 있는 경우에 적용할 수 있는 모형이다. 다시 말해 수집된 자료의 구성 상 0건의 사고 건수가 발생하고 있는 교차점이 많은 경우 적용할 수 있는 모형이다.

이 모형은 Vuong(1989)이 제안한 Vuong's 통계치를 통해 포아송회귀모형과 음이항회귀모형이 적합한지, 혹은 ZIP나 ZINB 모형이 적합한지를 판단하여 적용한다.

(1) ZIP 모형, $Y = (y_1, y_2, \dots, y_n)$

$$\begin{aligned} y_i = 0 & \text{ with probability } p_i + (1-p_i)\text{EXP}(-\lambda_i) \\ y_i = y & \text{ with probability } \frac{(1-p_i)\text{EXP}(-\lambda_i)\lambda_i^y}{y!} \end{aligned} \quad (16)$$

(2) ZINB 모형, $Y = (y_1, y_2, \dots, y_n)$

$$\begin{aligned} y_i = 0 & \text{ with probability } p_i + (1-p_i) \left[\frac{\frac{1}{\alpha}}{(\frac{1}{\alpha}) + \lambda_i} \right]^{\frac{1}{\alpha}} \\ y_i = y & \text{ with probability } (1-p_i) \left[\frac{\Gamma((\frac{1}{\alpha}) + y) u_i^{1/\alpha} (1-u_i)^y}{\Gamma(\frac{1}{\alpha}) y!} \right], \\ & y = 1, 2, 3, \dots \end{aligned} \quad (17)$$

여기에서 최대우도추정치들 이 모형들의 계수 추정을 위해 이용한다.

다음 식은 Vuong(1989)이 제안한 비내포 모형(non-nested model)에 대한 검정통계량이다.

$$m_i = LN \left(\frac{f_1(y_i|X_i)}{f_2(y_i|X_i)} \right) \quad (18)$$

여기에서 $f_1(y_i|X_i)$ 는 모형 1의 확률밀도함수

이며, $f_2(y_i|X_i)$ 는 모형 2의 확률밀도함수이다.

Vuong's 통계치는 표준정규분포와 유사하며, 만약 V 가 95%의 신뢰수준에서 1.96보다 큰 경우는 ZIP나 ZINB 모형을 선택하게 되며, -1.96보다 작은 경우는 포아송이나 음이항회귀모형을 선택한다.

$$V = \frac{\sqrt{n} \left[\left(\frac{1}{n} \sum_{i=1}^n m_i \right) \right]}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n (m_i - \bar{m})^2 \right)}} = \frac{\sqrt{n}(\bar{m})}{S_m} \quad (19)$$

여기에서 $\bar{m} = \left(\frac{1}{n} \sum_{i=1}^n m_i \right)$ 평균이며, S_m 은 표준편차, 그리고 n 은 표본크기이다.

2. 국내외 모형개발 사례

1) 국외사례

조반니스와 창(Jovanis and Chang, 1986)은 미국 인디애나주의 유료도로에 대해 차량간의 충돌유형별로 통행시간과 사고수의 관계를 설명하는 사고모형을 개발하였다. 이 연구는 선형 회귀방식의 문제점을 지적하면서 포아송 회귀모형을 개발하도록 권고하고 있다. 모형분석 결과 트럭과 승용차의 통행거리가 증가할수록 사고건수가 증가하며, 승용차의 사고건수가 트럭의 사고건수에 비해 우천 시 상황에 큰 영향을 받는 것으로 보고하고 있다.

지거(Zeeger, 1988)는 미국의 7개주 5,000마일에 이르는 2차로에 대한 자료를 기초로 사고유형별 사고율(단독차량사고, 전복사고)을 종속변수로 도로종단선형, 일평균교통량, 차로폭, 전체노면폭을 독립변수로 하는 사고모형을 개발하였다. 그 결과 종단선형이 완만할수록 사고율이 줄어들 뿐만 아니라 일평균교통량이 적을수록, 차로폭, 전체노면폭이 넓을수록 사고건수가 적었다. 또한 도로변의 가드레일이나 표지판보다 나무와의 충돌사고가 더 많은 것으로 분석되었고, 중상 및 사망사고의 경우, 방호벽이나 펜스보다 나무, 가로등, 암거 등에 의한 사고율이 높은 것으로 보고하고 있다.

미아우(Miau, 1993)는 미국 북부 캐롤라이나의 고속도로에 대해 기하구조와 트럭사고사이의 관계를 설명하기 위해 차로당 일평균일교통량, 종단곡선, 종단경사와 트럭사고수와의 관계를 포아송 분포를 사용하여 모형을 개발하였다. 모형분석 결과 교통량, 종단경사 증가하고, 종

단곡선이 작을수록 사고건수가 증가한다고 보고하고 있다.

칼라프티스(Karlaftis, et al., 2002)는 농촌도로 교통사고율에 대한 도로의 기하구조와 교통량의 효과를 위계적 수목기반 회귀(hierarchical tree-based regression)기법을 이용하여 분석하였다. 그는 독립변수의 중요성이 차로 수에 따라 다르지만, 가장 중요한 두 가지 요소는 기하구조 설계요소와 포장상태임을 밝히고 있다. 여기에서 독립변수로는 구간길이, 차로 수, 차로폭, 노면폭, 중앙분리대 폭, 노면유형, 포장유형, 콘크리트 포장, 중앙분리대의 유형, 회전차로, 연석 수, 주차 차로 수, 마찰계수, 포장서비스지수, 접근제어, 연평균일교통량(AADT) 등을 이용하여 분석하였다.

아울러 그레이베(Greibe, 2003)는 도시구간과 교차로의 간단한 사고모형을 정립하였다. 그는 GLM을 이용한 모형을 추정하였으며, 여기에서 사용된 변수는 AADT, 속도제한, 도로폭, km당 유출지점 수, km당 부도로의 수, 주차상황, 토지 이용 등이 이용되었다.

2) 국내사례

김태완(1996)은 경부고속도로 1993~1994년 사고자료를 근거로 차로수, 차로당 AADT, 선형, 경사, 유출입부, 기본구간에 따른 사고의 영향을 일반선형회귀식과 포아송회귀식으로 개발하였다. 그 결과 경부고속도로의 경우 차로수가 많을수록 사고가 감소하며, 곡선반경 1,000m이하와 종단경사 3%이상인 지점부터 사고가 급격히 증가한다고 보고하고 있다.

김경석(1997)은 도로 설계요소와 교통사고의 상호관계에 관한 연구에서 교통사고의 특성을 분류하고 교통사고와 관련된 도로설계요소의 특성을 분석하여 이들의 상관관계를 규명함으로써 도로설계요소의 개선점을 제시하였다.

박병호·류승욱(2002)은 트럼펫 인터체인지의 기하구조와 교통사고의 상관관계를 규명하고 있다. 그들은 인터체인지의 형식이 교통사고율에 미치는 영향을 검토하였으며, 다중회귀분석을 통해 교통사고에 미치는 독립변수들을 검증하여 적정 회귀식을 제시하고 있다.

성낙문(2002)은 교통사고의 과분산현상(overdispersed)을 반영하기 위하여 음이항분포에 근거한 비선형 교통사고 예측모형을 제시하고 있다. 그는 교통특성과 도로특성에 대한 독립변수를 GLIM(Generalized Linear Interactive Model)모형

을 통해 다이아몬드형과 부분적 클로버형 인터체인지에 대한 교통사고 예측모형을 개발하였다.

이기영, 이용택(2004)은 고속도로에서의 버스 와 화물차 각각에 대한 사고모형의 개발과 적용방안에 대하여 고찰하고 있다. 자료는 1998년 한국도로공사 사고조서를 기초로 고속도로 전역을 대상으로 있다. 도로구간에 대하여 종속변수로는 사망자, 중상자, 경상자 및 이를 종합한 사상자 수를 이용하였고, 독립변수로 운전자(인적)요인, 도로요인(기하구조 및 환경요인) 및 기타요인을 가지고, 포아송과 음이항 모형을 이용하여 분석하였다. 분석결과 기존의 관련 연구들의 결과와 유사한 결과를 도출하였으며, 국내 처음으로 인적요인을 포함시킨 연구로 그 인자가 미치는 영향을 파악하는 이론적 틀을 제공하였다고 언급하고 있다.

이인성(2004)은 미국의 각 주(州)에서 구축된 3년간의 도로구간 사고자료의 사고건수를 종속변수로 하고 독립변수를 교통량/용량(V/C), ADT, 속도제한(speed limit), 차로수, 도로유형(고도도로, 간선도로, 집산도로)을 가지고 가산자료모형의 4가지 모형(포아송, 음이항, ZIP 및 ZINB)을 분석하였다. 분석결과 4가지 모형비교에서는 ZIP 모형이 우수한 모형으로 나타났고, 관측치와 예측치의 사고빈도 확률분포 비교에서는 음이항 회귀모형이 우수한 것으로 분석하였다.

III. 자료 및 기술통계

자료 및 기술통계에서는 기하구조와 사고의 관계 및 모형분석에 앞서 분석에 이용되는 자료의 속성을 파악하고자 한다.

1. 분석자료

1) 분석대상 교차점

분석대상 교차점은 고속도로 교차점 233개 중 트럼펫 I.C 142개소, 기타 I.C 33개소 및 J.C 20개소 총 195개이다.

<표 2> 분석대상 교차점 유형

요금정산소 유무	교차점 형식	해당 교차점 수	분석 교차점 수	비 고
인터체인지 (I.C)	트럼펫형	156	142	분석대상 교차점 수는 195개로 전체의 약 83%
	기 타	33	33	
교차점(J.C)	-	37	20	
계		233	195	

주 : 분석 교차점 수는 해당 교차점 중 교통량, 사고통계기록 및 기하구조 자료획득이 가능한 지점의 수를 표현.

2) 조사된 기하구조 설계요소

조사된 기하구조 설계요소로는 교차점의 교차각, 차로수, 종단경사 및 곡선반경이며, 이중 분석에 사용된 요소는 램프구간 분석에 적용될 수 있는 차로수를 제외한 교차각, 종단경사 및 곡선반경이다. 추후에 보다 다양한 요소를 영향을 검토하기 위해서는 추가 조사해야 할 필요가 있다.

3) 교통량 자료

교통량은 사고의 위험성을 대별할 수 있는 중요한 독립변수이다. 교통량은 '전국고속도로 교통량 통계(1998~2003년)' 자료를 이용하였으며, 입체교차구간의 회전교통량을 산출하기 위해 교통계획 패키지인 EMME/2가 이용되었다.

4) 사고자료

사고자료는 종속변수로 이용되는 중요한 자료이며, 이 연구에서는 1998~2003년까지의 6개년의 사고자료를 수집·분석하였다.

<표 3> 교차점 사고자료

년 도	통계연보자료	수집자료	사용자료
1998	215	52	44
1999	209	159	133
2000	211	211	103
2001	211	211	134
2002	252	252	140
2003	181	186	166
총계	1,279	1,071	720

2. 기술통계분석

교차점 자료의 기술통계는 분포의 중심을 나타내는 중심경향 값을 중심으로 분석하였다.

표 4를 참조하면, 교차점의 평균사고건수는 3.69건이며, AADT의 평균은 43,479대이다. 또한 중심에서 좌우로 치우친 경향을 나타내는데 왜도는 교차각과 최대종단경사는 왼쪽으로 긴 꼬리를 형성하며, 나머지 최소곡선반경, 평균사고건수, AADT는 오른쪽으로 긴 꼬리를 형성하는 것으로 나타났다.

분석결과, 교차각은 직각교차인 90°를 기준으로 약간 오른쪽으로 치우쳐 있으며, 좌우로 균일하게 분포되어 있는 것으로 판단된다. 최소곡선반경의 경우는 왜도가 제일 크며 그 값들이 왼쪽으로 많이 치우쳐 분포되어 있는 것으로 나타났다. 최대종단경사는 분포 치우침은 다소 적으나, 곡선반경의 반대 경향을 보이고 있다.

<표 4> 자료의 기술통계

구분	교차각 (°)	최소곡선 반경(m)	최대종단 경사(%)	AADT (대수/일/년)	사고건수 (건/년)
평균	89.64	77.84	5.64	43,479	3.69
최소값	30	40	1.04	1,373	0.00
최대값	145	600	6.00	290,208	29.00
범위	115	560	4.96	288,835	29.00
표준편차	20.75	92.33	0.76	42,618	5.13
분산	431	8,524.33	0.58	1,816,264.566	26.32
왜도	-0.07	4.16	-2.97	2.33	2.60
첨도	-0.34	16.97	10.45	7.54	8.05
관측수	195	195	195	195	195

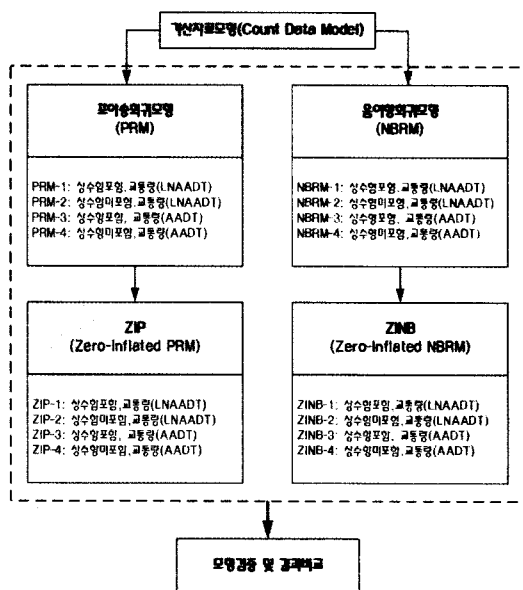
IV. 사고모형 개발

1. 모형분석 설정

모형 분석을 위해 사용된 변수들은 종속변수로 사고건수(ACC : accident), 독립변수로는 교차각(CA : cross angle), 최소곡선반경(CR : curve radius), 최대종단경사(VG : vertical grade) 및 연평일일교통량(AADT)을 이용하였다.

<표 5> 모형적용 변수

구분	기호	설명		
		평균	범위	
독립 변수	교차각(°)	CA	87	30~145
	최소곡선반경(m)	CR	78	40~600
	최대종단경사(%)	VG	5.6	1.04~6.00
	교통량(대수/일/년)	AADT	43,479	1,373~290,208
종속 변수	사고건수(건)	ACC	3.7	0~29



<그림 2> 사고모형 분석과정

이중 독립변수인 교통량(AADT)은 대부분 사고건수와 선형적 관계보다는 비선형 관계를 갖는 것으로 나타나고 있다(Pendleton, 1996). 따

라서 본 연구에서는 AADT와 LN(AADT)값을 병행하여 독립변수로 이용한다.

이 연구에서는 가산자료모형의 4가지 형태인 포아송회귀(PRM: Poission Regression Model), 음이항회귀(NBRM: Negative Binomial Regression Model), 0 과다 포아송회귀(ZIP: Zero-Inflated PRM), 및 0 과다 음이항회귀모형(ZINB: Zero-Inflated NBRM)을 이용하였다. 또한 각 모형에서 상수항 포함 여부와 교통량(AADT)의 변수의 적용형태에 따라 4가지 하부 모형으로 세분하였다. 모형의 개발과정은 그림 2와 같다.

2. 포아송회귀모형(PRM)

1) 모형분석

PRM-1 모형 계수의 p값은 95% 신뢰수준에서 유의하지만, LNAADT의 방향성(-0.120)은 타당하지 못하다. 즉 기존 연구들에서 제시하고 있는 '교통량이 증가할수록 사고건수는 증가하는 예상'과 반대의 결과이기 때문이다. 또한 모형의 설명력을 나타내는 $\rho^2 = 0.04$, $R_p^2 = 0.105$ 로 매우 낮다.

PRM-2 모형 경우는 계수 벡터의 방향성(CA: +, CR: -, LNAADT: +)은 모두 유의하지만 LNAADT의 p값이 유의수준($p \leq 0.05$)에 들지 못하고, PRM-1 모형과 마찬가지로 모형의 설명력이 매우 낮다.

<표 6> 포아송회귀모형 1과 2의 결과

구분	PRM-1			PRM-2		
	계수	t-값	p값	계수	t-값	p값
상수	1.591	3.307	0.001	-	-	-
CA	0.014	6.742	0.000	0.014	6.841	0.000
CR	-0.002	-2.628	0.009	-0.002	-3.242	0.001
LNAADT	-0.120	-2.739	0.006	0.013	0.790	0.430
R_p^2	0.105			0.069		
χ^2	59.901(0.000)			49.251(0.000)		
G^2	995.297			1019.160		
ρ^2	0.040			0.033		

주 : p값은 95%신뢰수준에서 유의해야 함.

PRM-3 모형은 상수항이 유의수준($p \leq 0.05$)에 적합하지 않아 적절치 못한 모형으로 판단된다.

PRM-4 모형은 분석된 포아송회귀모형 중 가장 적절한 모형으로 판단된다. 계수의 방향성(CA: +, CR: -, AADT: +)이 타당하며, 모두 유의수준($p \leq 0.05$)의 범위에 들며, 모형편차의 합인 R_p^2 값이 4개의 모형 중 가장 우수하다. 하지만

모형전체의 적합도를 나타내는 ρ^2 는 0.152로 다소 낮은 것으로 판단된다.

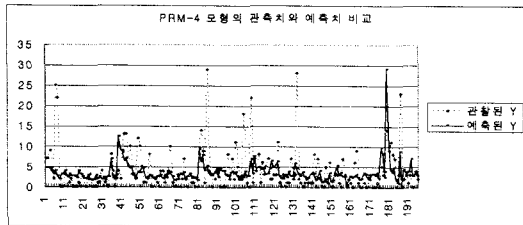
<표 7> 포아송회귀모형 3과 4의 결과

구분	PRM-3			PRM-4		
	계수	t-값	p값	계수	t-값	p값
상수	-0.056	-0.300	0.764	-	-	-
CA	0.011	6.742	0.000	0.010	6.841	0.000
CR	-0.002	-3.740	0.000	-0.002	-3.242	0.000
AADT	0.142D-05	15.405	0.000	0.142D-05	15.758	0.000
R_p^2	0.240			0.242		
χ^2	227.493(0.000)			227.402(0.000)		
G^2	827.706			824.573		
ρ^2	0.152			0.152		

2) 결과 및 검증

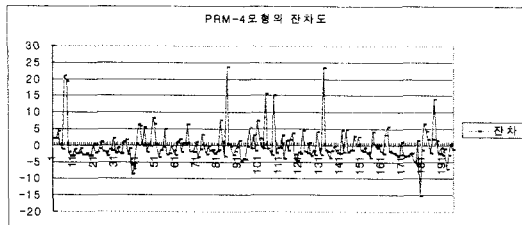
포아송회귀모형의 분석에서는 PRM-4 모형이 가장 적절한 모형으로 판단된다.

아래의 그림은 PRM-4 모형의 사고건수 예측치와 실제 사고건수를 비교한 그림이다. 전반적인 흐름은 유사하나, 몇몇 지점에서 크게 오차가 발생함을 볼 수 있다.



<그림 3> PRM-4 모형의 예측치와 관측치 비교

잔차도에서도 마찬가지로 눈에 띄는 4~5개 지점에서 큰 오차가 발생되고 있는 것으로 나타났다.



<그림 4> PRM-4 모형의 잔차도

3.. 음이항회귀모형(NBRM)

1) 모형분석

음이항회귀모형은 포아송회귀모형에 비해 모형의 적합성 즉 설명력을 나타내는 ρ^2 값이 2배

가량 높다. 이는 포아송회귀모형에서 평균과 분산 같다는 가정이 부적합하여 과분산이 발생되고 있다는 것으로 판단된다.

NBRM-1 모형에서는 LNAADT의 계수 벡터 방향성(-)이 부적절하며, 모형의 상수, CR 및 LNAADT의 유의수준이 적절치 못하다.

NBRM-2 모형에서는 LNAADT가 벡터의 적절한 방향성(+)을 가지나 신뢰수준 95%에 들지 못하고 있다.

<표 8> 음이항회귀모형 1과 2의 결과

구분	NBRM-1			NBRM-2		
	계수	t-값	p값	계수	t-값	p값
상수	1.186	1.070	0.285	-	-	-
CA	0.013	2.826	0.005	0.014	2.987	0.003
CR	0.002	-1.502	0.133	-0.002	-1.765	0.078
LNAADT	-0.078	-0.822	0.411	0.016	0.448	0.654
χ^2	507.366(0.000)			516.859(0.000)		
ρ^2	0.354			0.357		
α	1.349(7.430, 0.000)			1.361(7.459, 0.000)		

주1 : α 는 과분산 계수, ()안은 t-statistics와 p-value임.

주2 : p값은 95%신뢰수준에서 유의해야 함.

NBRM-3 모형에서는 계수벡터의 적절한 방향성(+)을 가지나 신뢰수준 95%에 있어서 상수 계수벡터가 적절치 못하다.

NBRM-4 모형은 계수 벡터의 방향성(+)과 유의수준(0.05) 모두 만족시키며, 분석된 음이항회귀모형 중 가장 우수한 모형으로 판단된다. 모형의 설명력도 $\rho^2 = 0.295$ 로 포아송회귀모형에 비해 매우 높다.

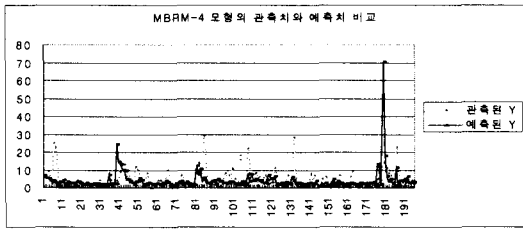
<표 9> 음이항회귀모형 3과 4의 결과

구분	NBRM-3			NBRM-4		
	계수	t-값	p값	계수	t-값	p값
상수	-0.172	-0.430	0.667	-	-	-
CA	0.010	2.444	0.015	0.009	5.324	0.000
CR	-0.002	-1.896	0.058	-0.002	-2.010	0.044
AADT	0.207D-05	5.478	0.000	0.205D-05	5.476	0.000
χ^2	374.471(0.000)			374.376(0.000)		
ρ^2	0.295			0.295		
α	1.027(6.785, 0.000)			1.030(6.788, 0.000)		

2) 결과 및 검증

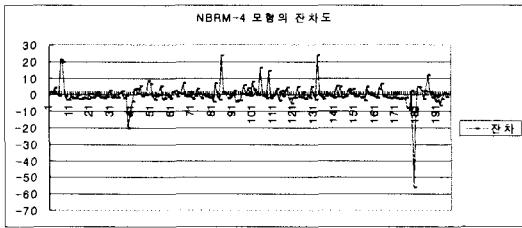
음이항회귀모형으로 구축된 모형 중 가장 우수한 모형은 NBRM-4 모형이며, 과분산 계수 $\alpha = 1.030$ 으로 95% 신뢰수준을 만족하고 있어 포아송회귀모형의 α 값에 의해 간접적으로 드러나는 과분산 문제를 적절히 해결하는 것으로 판단된다. 음이항회귀모형 중 NBRM-4 모형으

로 예측/관측 값의 비교를 통해서 보면, 포아송 회귀모형의 예측치와 유사한 패턴을 보이나 다소 낮은 값으로 예측된 것을 알 수 있다.



<그림 5> NBRM-4 모형의 예측치와 관측치 비교

PRM-4 모형의 잔차도와는 달리 실제 관측치에 예측치가 많이 못 미치는 경향을 나타내고 있다.



<그림 6> NBRM-4 모형의 잔차도

4. ZIP 모형

ZIP 및 ZINB 모형은 0값이 과다하게 분포된 포아송과 음이항회귀모형의 확장된 형태이다. 본 연구의 사고자료의 빈도를 보면 0건인 사고 지점 수가 195개 교차점 중 53개소(27.2%)를 차지하고 있어, ZIP 모형이나 ZINB 모형은 검토해 볼 필요가 있다.

<표 10> 사고자료의 빈도분포

사고건수	빈도	%	사고건수	빈도	%
0	53	27.2	11	3	1.5
1	31	15.9	12	1	0.5
2	25	12.8	13	2	1.0
3	24	12.3	14	2	1.0
4	13	6.7	18	1	0.5
5	6	3.1	22	2	1.0
6	3	1.5	23	1	0.5
7	10	5.1	25	1	0.5
8	7	3.6	28	1	0.5
9	5	2.6	29	1	0.5
10	3	1.5	계	195	100

1) 모형분석

분석결과, ZIP-1 모형은 모든 변수들이 유의수준에 들었지만, LNAADT의 계수 방향성이

음의(-)값으로 적절치 못하였다. ZIP-2 모형은 ZIP-2 모형이 계수의 방향성 및 유의수준이 적절한 것으로 나타났다. 모형의 채택여부에서 Vuong 통계치가 $V = 4.336$ 으로 +1.96(신뢰수준 95%)보다 그 값이 크므로, 포아송 모형의 선택 보다는 ZIP-2 모형의 선택이 적절한 것으로 판단된다.

ZIP-3 모형은 각 변수들에 대한 계수 값이 도출되지 않아 모형의 효력이 없으며, ZIP-4 모형은 모형 계수의 방향성(CA: +, CR: -, AADT: +)은 타당하나, Vuong 통계치가 0.626으로 -1.96(신뢰수준 95%) 보다 작아 모형채택은 어려운 것으로 판단된다.

<표 11> ZIP 모형 1과 2의 결과

구분	ZIP-1			ZIP-2		
	계수	t-값	p값	계수	t-값	p값
상수	2.301	12.238	0.000	-	-	-
CA	0.010	8.839	0.000	0.010	9.182	0.000
CR	-0.001	-2.675	0.008	-0.002	-4.826	0.000
LNAADT	-0.125	-6.250	0.000	0.063	6.821	0.000
ρ^2			0.175			0.162
τ		-0.672(-6.250, 0.000)			-0.708(-6.328, 0.000)	
Vuong's st.		4.4134			4.336	

주1 : τ 계수는 ZIP 또는 ZINB 모형을 형상화하는 계수(shape parameter)임.

주2 : Vuong's statistics는 ZIP 또는 ZINB가 포아송이나 음이항모형을 대체하는 판단 기준임.

주3 : p값은 95%신뢰수준에서 유의해야 함.

분석된 ZIP 모형 중에서는, ZIP-2 모형이 가장 적절하게 구축된 것으로 판단되나 모형의 설명력을 나타내는 ρ^2 값은 0.162로 높지는 않은 것으로 판단된다.

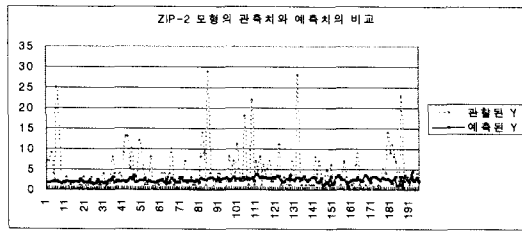
<표 12> ZIP 모형 3과 4의 결과

구분	ZIP-3			ZIP-4		
	계수	t-값	p값	계수	t-값	p값
상수	-	-	-	-	-	-
CA	-	-	-	0.011	32.706	0.000
CR	-	-	-	-0.002	-6.383	0.000
AADT	-	-	-	0.142D-05	34.923	0.000
ρ^2			-			0.162
τ			-		0.000(0.000, 1.000)	
Vuong's st.			-			0.626

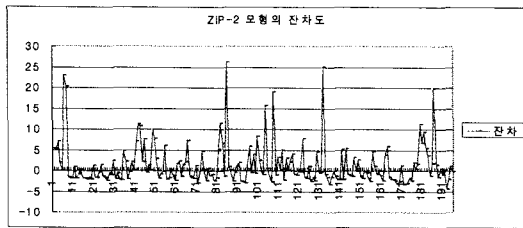
2) 결과 및 검증

ZIP 모형 중 가장 적절한 모형으로 판단되었던 ZIP-2 모형의 예측치와 실제 관측치 비교는 아래의 그림7과 같으며, 두 값의 차이를 의미하는 잔차(residuals)는 그림 8과 같다.

ZIP-2 모형의 예측치는 큰 변동 없이 일정한 패턴을 보이고 있으며, 잔차도에서 볼 수 있듯이 관측치보다 오차범위가 큰 것으로 나타났다.



<그림 7> ZIP-2 모형의 예측치와 관측치 비교



<그림 8> ZIP-2 모형의 잔차도

5. ZINB 모형

ZINB 모형 1, 2 및 3 모형에서는 모두 계수의 방향성 및 유의성이 적절치 못한 것으로 분석된다. ZINB-4 모형은 방향성과 유의성 모두 좋지만, $V = -6.273$ 으로 95%의 신뢰수준에서 -1.96보다 작아 ZINB 모형은 선택될 수 없게 된다. 따라서 현재의 자료로는 ZINB 모형에서의 적절한 모형 도출은 어려울 것으로 판단된다.

<표 13> ZINB 모형 1과 2의 결과

구분	ZINB-1			ZINB-2		
	계수	t-값	p값	계수	t-값	p값
상수	1.079	1.221	0.222	-	-	-
CA	0.011	2.807	0.005	0.011	2.880	0.004
CR	-0.001	-1.538	0.124	-0.001	-1.857	0.063
LNAADT	-0.052	-0.683	0.495	0.033	0.967	0.333
ρ^2	0.378			0.378		
α	1.241(6.723, 0.000)			1.267(6.940, 0.000)		
τ	-2.562(-2.469, 0.014)			-2.709(-2.334, 0.020)		
Vuong's st.	-0.712			-0.704		

주1 : τ 계수는 ZIP 또는 ZINB 모형을 형상화하는 계수(shape parameter)임.

주2 : Vuong's statistics는 ZIP 또는 ZINB가 포아송이나 음이항모형을 대체하는 판단 기준임.

주3 : p값은 95%신뢰수준에서 유의해야 함.

<표 14> ZINB 모형 3과 4의 결과

구분	ZINB-3			ZINB-4		
	계수	t-값	p값	계수	t-값	p값
상수	-0.172	-0.364	0.716	-	-	-
CA	0.010	2.009	0.045	0.009	5.324	0.000
CR	-0.002	-1.760	0.079	-0.002	-2.010	0.061
AAADT	0.207D-05	5.171	0.000	0.205D-05	5.476	0.000
ρ^2	0.324			0.324		
α	1.028(2.604, 0.009)			1.030(2.607, 0.009)		
τ	0.000(0.000, 1.000)			0.000(0.000, 1.000)		
Vuong's st.	-6.361			-6.273		

V. 결론 및 향후 연구과제

본 연구는 고속도로 입체교차구간의 기하구조와 교통사고자료를 바탕으로 사고모형 개발을 하고자 하였다. 이에 사고모형에 적용되는 모형들의 이론 검토와 개발 사례를 검토하였으며, 이를 통하여 사고 특성을 잘 설명할 수 있는 가산자료모형을 적용한 사고모형을 개발하였다. 적용한 가산자료모형은 포아송회귀모형과 음이항회귀모형 그리고 이 두 모형을 확장시킨 ZIP와 ZINB 모형이다. 유의미한 모형은 추정된 계수의 유의성과 모형의 적합도로서 판단된다.

연구의 주요결과로는 첫째, 포아송회귀모형(PRM)과 음이항회귀모형(NBRM)의 분석결과에서는 PRM-4 모형($\rho^2 = 0.152$)과 NBRM-4 모형($\rho^2 = 0.295$)이 적절한 모형으로 나타났다.

둘째, ZIP 모형에서는 각 변수의 방향성과 유의성이 타당한 ZIP-2 모형($\rho^2 = 0.162$)이 적절한 모형으로 분석되었으나, ZINB 모형에서는 모형 계수의 유의성과 Vuong's statistics에 의한 모형 채택 여부에 문제가 있었다.

마지막으로, 음이항회귀모형의 NBRM-4 모형이 가장 적합한 모형으로 판단되며, 이와 같은 판단은 각 모형의 검증에서 검토된 예측치와 관측치 비교, 잔차도를 통해서 확인할 수 있다.

향후 연구과제로는 좀 더 다양한 접근 방법들을 검토해볼 필요가 있으며, 독립변수들의 자료가 4가지로 제약되어 있어 추가적인 자료의 확보가 필요하다. 또한 이런 연구들이 지속적으로 이루어지기 위해서는 관련 변수들에 대한 다양하고 지속적인 데이터 수집 및 관리가 필요할 것으로 판단된다. 또한 사고자료의 시계열적 분석과 사고지점의 적용을 교차점 램프별로 적용하여 검토해 볼 필요가 있다.

참고문헌

1. 강근석, 김충락(1999) 회귀분석, 교우사.
2. 김경석(1997) 도로 설계요소와 교통사고의 상호관계에 관한 연구, 국토연구, 제26권, pp. 148~160.
3. 김태완(1996) 도로기하구조가 교통사고에 미치는 영향-경부고속도로 중심으로, 석사학위논문, 서울대학교 환경대학원.
4. 류승욱, 박병호(2006) 고속도로 교차점의 기하구조와 교통사고의 다중회귀분석, 건설기술논문집 25권 제1호, 충북대학교 건설기술연구소.
5. 박병호, 류승욱(2002) 트럼펫 인터체인지의 형식과 교통사고율의 관계, 한국지역개발학회지 제14권 제1호, pp. 107~124.
6. 성낙문(2002) 고속도로 인터체인지에서 교통사고 예측모형의 개발, 대한토목학회논문집 제22권 제4-D호, pp.617~625.
7. 오주택, 성낙문, 하오근(2005) 국도변 신호교차로 안전성 향상을 위한 사고예측모형개발, 대한토목학회논문집 제25권 제1-D호, pp.9~15.
8. 이기영, 이용택(2004) 확률회귀모형을 이용한 고속도로의 사고요인 분석, 도로교통 제94호, pp.51~64.
9. 이인성(2004) 교통사고자료에 대한 ZIP 회귀분석 적용, 석사학위논문, 연세대학교.
10. Cameron, A. and Trevedi, P.(1990) Regression based tests for overdispersion in the Poisson model, Journal of Econometrics 46, pp. 347~364.
11. Berhanu, Girma(2004) Models relating traffic safety with road environment and traffic flows on arterial roads in Addis Ababa, Accident Analysis and Prevention, Vol. 36, pp.697~704.
12. Dufour, Jean-Marie, et al.(2004) Simulation-based finite-sample tests for heteroskedasticity and ARCH effects, Journal of Econometrics.
13. Greene, William H.(2000) Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function, Stern School of Business New York University.
14. Greibe, Poul(2003), Accident prediction models for urban roads, Accident Analysis and Prevention 839, pp. 1~13.
15. Jovanis, Paul P. and Chang, Hsin-li(1986) Modelling the Relationship of Accident to Miles Traveled, TRR 1068.
16. Karlaftis, Matthew G., and Ioannis Golias (2002) Effects of road geometry and traffic volumes on rural roadway accident rates, Accident Analysis and Prevention, Vol. 34, pp. 357~365.
17. Lee, J. and Mannering, F.(2002) Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis, Accident Analysis and Prevention 34, pp. 149~161.
18. Miaou, S.P. and Lum, H.(1993) statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements, Transportation Research Record 1407, Transportation Research Board, National Research Council, Washington, D.C., pp. 11~23.
19. Mountain, Linda, et al.(1996) Accident Prediction Models for Road with Minor Junctions, Accident Analysis & Prevention, Vol. 28, No. 6, pp.695~707.
20. Poch, Mark and Mannering, Fred(2004) Negative Binomial Analysis of Intersection-Accident Frequencies, Journal of Transportation Engineering, MARC/APRIL, pp.105~113.
21. Pendleton, O.(1996) Evaluation of accident analysis methodology, Report FHWA-RD-96-039, FHWA.
22. Vuong, Q.(1989) Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica 57, pp. 307~334.
23. Washington, Simon P., et al.(2003) Statistical and Econometric Methods for Transportation Data Analysis, Florida : Chapman & Hall/CRC.
24. Zegeer, C.V., et al.(1988) Accident Effects of Side-slope and Other Roadside Features on Two-Lane Roads, Transportation Research Record 1195, Transportation Research Board, National Research Council, Washington, D.C., pp. 33~47.