

■ 論 文 ■

과분산된 교통사고자료에 대한 한계사고율법의 적용에 관한 연구

A study on Application of the Rate Quality Control
Method of Over-dispersed Traffic Crash Data

성 낙 문

(교통개발연구원 광역도시교통연구실 책임연구원)

목 차

- I. 서론
- II. 교통사고자료의 과분산 현상과 한계사고율의 문제점 분석
 - 1. 과분산 현상
 - 2. 과분산 현상에 따른 한계사고율법의 문제점
- III. 과분산을 반영한 한계사고율법의 개발 및 적용
 - 1. 과분산을 반영한 한계사고율법의 개발
 - 2. 과분산을 반영한 한계사고율법의 적용
 - 3. 적용결과
- IV. 개발된 방법의 타당성 평가 및 논점
 - 1. 개발된 방법의 타당성 평가
 - 2. 위험지점선정상의 논점
- V. 결론
- 참고문헌

Key Words : 교통안전, 교통사고자료, 과분산, 한계사고율법, 위험지점

요 약

교통안전프로그램을 수행하는데 있어서 위험지점을 적합하게 선정한다는 것은 매우 중요하다. 위험지점을 선정하는데 있어서 한계사고율법은 불확실성을 가지는 교통사고를 통계학적으로 해석 할 수 있다는 장점 때문에 자주 사용되어지는 방법이다. 한계사고율법은 교통사고의 발생이 포아송분포를 따른다는 가정에 기반을 두고 있으며 포아송분포에서는 교통사고의 기대값과 분산이 산술적으로 동일하다. 그러나 실제 교통사고자료에는 분산이 기대값보다 큰 현상이 자주 발생하는데 이것을 통계학적으로 과분산이라 정의하며 도로 및 교통여건이 변화하는 한 교통사고자료의 분석시 흔히 발생하는 현상이다.

본 연구에서는 과분산이 발생했을 때 포아송분포에 기반한 한계사고율법의 문제점을 분석하고 극복할 수 있는 방법을 제시하였다. 또한 실제 관측자료를 토대로 본 연구에서 정립된 방법이 기존의 방법보다 통계학적인 측면에서 타당한 결과를 왔다.

1. 서론

교통안전 프로그램을 수행하는데 있어서 위험지점(혹은 사고잠은지점)을 올바르게 선정하는 것은 제한된 예산을 효율적으로 이용한다는 측면에서 매우 중요하다. 교통여건이나 도로여건에 대한 고려 없이 발생한 교통사고건수만을 근거로 하는 교통사고건수에 의한 방법, 특정지점이나 구간의 교통사고율을 계산하여 사업의 우선순위를 결정하는 교통사고율법 등이 많은 나라에서 자주 사용되어지는 방법이다. 이외에 사고의 심각도(Severity)를 고려하여 하나의 피해단위로 산정하는 방법인 대물피해환산법도 하나의 대안으로서 제기된다. 이러한 방법들은 교통사고에 내재되어 있는 불확실성(Uncertainty)을 반영하지 못한다는 평가를 받고 있지만 매우 단순하다는 이유 때문에 자주 이용되어진다. 우리나라의 경우 주로 교통사고건수에 의해서 위험지점을 선정하고 있다.

앞에서 언급한 것처럼 교통사고에는 불확실성이 내재되어 있다. 예를 들면 지난해에 어떤 특정지점에서 교통사고가 30건이 발생하였다는 치다라도 내년에 같은 건 수의 교통사고가 발생할 것이라고는 누구도 장담할 수가 없다. 이러한 불확실성을 반영코자 교통사고를 확률적으로 해석하게 되었으며 한가지 예가 한계사고율법이다. 한계사고율법은 유사한 지점(Reference site)의 선정에서 발생하는 문제점 등 몇 가지 한계가 있지만 불확실성을 내포하고있는 교통사고를 확률적으로 해석한다는 측면에서 위의 두 방법 못지않게 자주 이용되어진다. 한계사고율법은 어떤 특정지점의 교통사고율과 비슷한 특성(도로특성, 교통특성 등)을 지닌 유사한 지점들의 교통사고율을 비교하여 그 특정지점의 교통사고율에 대하여 확률적인 해석을 내리는 방법이다. 한계사고율법은 교통사고의 발생이 포아송(Poisson)분포를 따른다는 가정에 기반을 두고 있으며 포아송분포에서는 교통여건이나 도로여건이 항상 동일하며 교통사고의 기대치와 교통사고의 분산이 산술적으로 일치한다는 의미가 내포되어있다.

그렇다면 교통여건이나 도로여건이 언제나 동일할까?

실제적으로 교통여건이나 도로여건의 변화를 초래하는 많은 요인들이 도처에 널려있다. 예를 들면 날씨, 경찰의 단속, 도로의 보수유지 상태 등 교통사고에 영향을 주는 많은 요인들이 매년 바뀐다. 이러한 요인의 변화는 어떤 지점에서 발생되리라 예측되는

교통사고건수의 변동(Fluctuation)폭을 넓혀주어 교통사고의 분산이 교통사고의 기대값보다 커지는 현상을 발생 시킨다.

본 연구는 교통안전프로그램에서 교통사고위험지점의 선정 시, 교통사고의 과분산현상하에서 기존의 한계사고율법이 발생시킬 수 있는 개념적인 문제점을 분석하고 그 문제점을 극복 할 수 있는 방안을 제시함에 그 목적이 있다.

II. 교통사고자료의 과분산 현상과 한계사고율법의 문제점 분석

1. 과분산 현상

1) 과분산의 개념

기존의 한계사고율법은 교통사고의 발생은 포아송 분포를 따른다는 가정하에 정립된 방법으로서 포아송 분포에는 교통여건이나 도로여건이 항상 동일하여 교통사고의 기대값과 교통사고의 분산이 산술적으로 일치(기대값=분산) 한다는 의미가 내포되어 있다. 그러나 실제로는 날씨, 경찰의 단속, 도로의 보수유지 상태 등 교통사고에 영향을 주는 많은 요인들이 매년 바뀐다. 이러한 요인의 변화는 어떤 지점에서 발생되리라 예측되는 교통사고건수의 변동(Fluctuation)폭을 넓혀주어 교통사고의 분산이 교통사고의 기대값보다 커지는 현상(분산) 기대값이 발생한다. 이것을 통계학적으로 과분산(Over dispersion) 현상이라 정의하며 본 연구를 통하여 발견된 사실이 아니라 이미 많은 교통안전학자들이 인정하는 사실이다(Hauer 1996, Hauer and Persaud 1984, Hutchinson and Mayne 1977).

2) 과분산 현상을 극복하기위한 기존연구

위에서 논의한 과분산 현상을 극복하기 위하여 여러가지 방법이 제시되었다.

Higgle과 Witkowski는 과분산 현상과 관련된 문제점을 극복하고자 베이지언 이론을 적용하는 방법을 제시하였다(Higgle and Witkowski 1988). 이 방법은 사전분포(Prior distribution)의 불확실성을 사후분포(Posterior distribution)를 통하여 보정할 수 있다는 논리 하에 다음과 같은 두 가지 가정을 기반으로 하고 있다.

첫째는 사전 분포에 대한 가정이다. 즉, Reference sites의 교통사고율(λ)을 알고 있을 때 어떤 지점에서 발생한 교통사고건수(N_i)는 평균값(λN_i)을 가지는 포아송 분포를 따른다는 가정이다. 둘째는 사후분포에 대한 가정이다. 이것은 Reference sites의 교통사고율은 감마(Gamma)를 따른다는 가정이다.

이와 같은 두가지 가정을 기반으로 특정지점의 사전 분포는 Reference sites의 사후분포와 결합하여 새로운 형태의 확률밀도함수가 도출되며, 이 확률밀도함수를 이용하여 교통안전개선사업이 필요한 지점이 선정된다.

Higgle와 Witkowski의 방법보다 더욱 일반적으로 사용되는 방법은 Hauer이 제시한 방법이다. Hauer은 교통사고데이터의 변동(Fluctuation)을 보정하기 위하여 근사적 베이지언(Empirical Bayesian)의 개념을 적용하였는데 기본식은 다음과 같다.

$$E(m/x) = wE(m) + (1-w)x$$

여기서,

- $E(m/x)$: 어떤 지점의 보정된 교통사고건수
- $E(m)$: Reference sites 에서의 교통사고건수
- x : 어떤 지점에서 관측된 교통사고건수

$$w(\text{가중치}) = E(m) / [E(m) + \text{Var}(m)]$$

$$\text{Var}(m) = E(m)^2 / k$$

위 식에서 보는 바와 같이 어떤 지점의 보정된 교통사고건수, $E(m/x)$ 은 그 지점에서 실제로 발생한 교통사고건수와 Reference sites에서 발생한 교통사고건수 및 교통사고건수의 분산값을 이용하여 산정이 가능하다. 위와 같은 개념을 근거로 Hauer은 교통사고예측모델을 이용하여 $E(m)$ 과 $\text{Var}(m)$ 을 추정하는 방법을 정립하였다. Hauer이 제시한 방법은 Reference sites의 선정과 관련된 개념상 혹은 적용상의 한계를 극복할 수 있다는 측면에서 미국, 캐나다 등에서 최근에 자주 사용되어진다. Reference sites와 관련된 개념상 혹은 적용상의 문제점은 뒤에서 좀더 구체적으로 논해진다.

2. 과분산 현상에 따른 한계사고율법의 문제점

한계사고율법은 본래 공장에서 불량품을 골라내려

는 목적으로 개발된 방법이지만 어떤 특정 지점의 교통사고율을 유의수준(Significance level)의 개념 하에 해석할 수 있다는 장점 때문에 교통안전분야에서 폭넓게 이용되고 있다. 앞에서 논의한 것처럼 한계사고율법은 교통사고의 발생은 포아송분포를 따른다는 가정을 기반으로 하며, 시간이 지남에 따라 몇 차례의 변화가 있었지만 다음이 일반적으로 사용되어지는 기본식이다.

$$U_i = \lambda_{avg} + k \sqrt{\frac{\lambda_{avg}}{V_i} + \frac{1}{2V_i}} \tag{1}$$

여기서,

- U_i : 지점 i 에서의 한계교통사고율
- k : 유의수준에 따른 확률계수
- λ_{avg} : $\sum N_i / \sum V_i$
- N_i : 지점 i 에서의 교통사고건수
- V_i : 지점 i 에서의 교통량

식(1)에서 앞의 두 항은 포아송분포를 정규분포로 근사화하는데 필요한 항이며 셋째항은 불연속분포인 포아송분포를 연속분포인 정규분포로 변환시킴으로써 생기는 오차를 보정하기 위한 것이다(Stocks and Mutabazi 1996, Hauer 1996).

또한 교통량이 많은 경우, 셋째항은 매우 작은 값이므로 무시 할 수 있는 수치이다.

여기에서 k 는 유의수준에 따라 결정되는 확률계수로 다음과 같은 값을 사용한다

<표 1> 유의 수준에 따른 확률계수

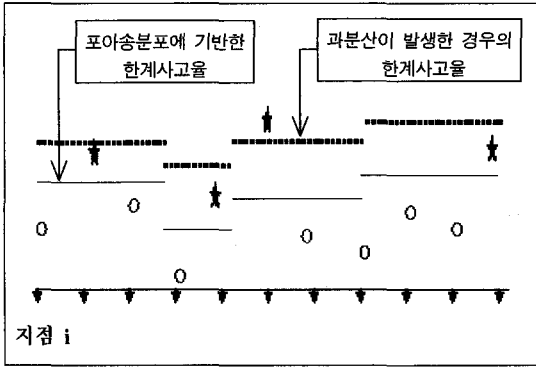
유의수준(%)	50	80	90	95	99	99.5
k값	0.00	0.842	1.282	1.645	2.325	2.576

주 : 도철용, 교통공학원론(하), pp.728~731, 1991.

Zeger, C.V., and Deen, R.C. "Identification of Hazardous Locations on City Streets". Traffic Quarterly, pp.601~610, 1977.

만약 주어진 기간동안에 어떤 특정지점에서 관측된 교통사고율이 식(1)에서 계산된 한계교통사고율보다 크다면 이것은 일시적 혹은 우연히 발생한 현상이 아니라 통계학적으로 충분히 설득력 있는 현상이므로 그 지점을 위험지점으로 간주하여 교통안전프로그램의 수행시 우선적으로 개선이 이루어져야 한다는 것이다.

그러나 앞에서 논의한 것처럼 과분산이 발생한다면



〈그림 1〉 포아송분포에 기반한 한계사고율과 과분산이 발생한 경우의 한계사고율 비교

교통사고자료에서 발생하는 관측된 분산이 포아송분포에서 가정하는 이론적인 분산보다 크므로 포아송분포에 기반한 한계사고율법은 위험지점의 선정 시 부적절한 결과를 발생시킬 수 있다. 〈그림 1〉은 이러한 현상을 개념적으로 나타내어준다. 식(1)에서 보는 것처럼 한계사고율은 유사한 지점의 평균사고율과 분산의 제곱근의 어떤 배수(유의수준, k)를 더한 값이다. 그러나 과분산이 발생한 경우, 분산은 포아송분포에서 가정한 값보다 커지므로 실제적인 한계사고율 또한 커져야 한다. 〈그림 1〉에서 보는 바와 같이 포아송분포에 기반한 한계교통사고율을 실선이라고 가정했을 때 과분산이 발생한 경우, 실제 한계 교통사고율은 그 실선보다 위쪽에 위치한다. 따라서 포아송분포에 기반한 한계사고율법은 교통사고의 과분산을 고려하지 못하므로 투자가 불필요한 지점까지 교통사고위험지점으로 인식하게 되므로, 기존의 포아송분포에 기반한 한계사고율법은 교통안전프로그램의 수행 시 제한된 예산의 효과적인 운용을 방해하게 된다.

III. 과분산을 반영한 한계사고율법의 개발 및 적용

1. 과분산을 반영한 한계사고율법의 개발

확률이론에서 평균값이 증가함에 따라 포아송분포의 확률분포함수(Probability Mass Function)는 평균을 중심으로 대칭적이며 종모양(Symmetric and bell shaped)으로 변화하여 정규분포화 한다(Rice 1997). 이 개념을 교통사고에 적용하여 식으로 표현하면 다음과 같다.

$$N_i \sim N(\mu_i, \mu_i) \tag{2}$$

여기서,

$$\mu_i = \lambda V_i$$

$$\lambda = \frac{\sum N_i}{\sum V_i}$$

N_i : 지점 i 에서의 교통 사고건수

V_i : 지점 i 에서의 차량통과대수

λ : 전체지점의 평균 교통사고율

μ_i : 지점 i 에서의 교통사고건수의 기대값

여기서 N_i 는 교통사고건수로서 확률변수이며 포아송분포이므로 $E(N_i)$, $Var(N_i)$ 는 같은 값을 갖게 된다. 따라서 확률변수 N_i 는 다음과 같이 표준화가 가능하다. 식(3)에 의해서 도출된 확률변수 X_i 는 정규

$$X_i = \frac{N_i - \mu_i}{\sqrt{\mu_i}} \tag{3}$$

여기서,

$$E(X_i) = 0$$

$$Var(X_i) = 1$$

즉, $X_i \sim N(0, 1)$

분포를 보일 것이며 이에 대한 검증이 필요하다. 교통사고의 분산이 평균보다 크다면 과분산을 고려하기 위해서 포아송분포에서와 같은 방식으로 그 교통사고자료를 다음과 같이 근사화 시킬 수 있다.

$$N_i \sim N(\mu_i, d\mu_i) \tag{4}$$

여기서,

$$\mu_i = \lambda V_i$$

$$\lambda = \frac{\sum N_i}{\sum V_i}$$

N_i : 지점 i 에서의 교통 사고건수

V_i : 지점 i 에서의 차량통과대수

λ : 전체지점의 평균 교통사고율

μ_i : 지점 i 에서의 교통사고건수의 기대값

$d\mu_i$: 지점 i 에서의 교통사고건수의 분산 ($d \geq 1$)

식(4)에서 $E(N_i)$ 는 μ_i , $Var(N_i)$ 는 $d\mu_i$ 이므로 과분산이 발생한 교통사고자료는 다음과 같이 표준화가 가능하다.

$$Z_i = \frac{N_i - \mu_i}{\sqrt{d\mu_i}} \quad (5)$$

여기서,

$$E(Z_i) = 0$$

$$Var(Z_i) = 1$$

즉, $Z_i \sim N(0, 1)$

식(5)를 이용하여 한계사고율법을 적용하기 위해서는 평균사고율 " λ " 와 과분산과 관계된 파라미터 " d "의 산정이 필요하다. " λ "는 발생한 총 교통사고를 통과교통량으로 나눈 값이며 " d "는 다음과 같은 방법을 이용하여 산정할 수 있다.

$$Var(Z_i) = Var\left(\frac{N_i - \mu_i}{\sqrt{d\mu_i}}\right) = \frac{1}{d} Var\left(\frac{N_i - \mu_i}{\sqrt{\mu_i}}\right) = \frac{1}{d} Var(X_i) = 1 \quad (6)$$

$$\therefore d = Var(X_i)$$

식(5)에서 보는 것처럼 분산과 관련된 파라미터 " d "는 식(3)에 의해서 표준화된 확률변수 X_i 의 분산 값이다. 식(5)의 확률변수 Z_i 를 이용하여 과분산이 발생한 교통사고자료에 대한 한계사고율법 적용이 가능하다. 즉, Z_i 가 어떤 특정 통계값, 혹은 한계사고율(예 : 95% 유의수준에서의 통계값은 1.64)을 초과했을 때 그 지점의 교통사고율이 비슷한 특성을 가진 다른 지점보다 높은 것은 우연이라기보다는 교통사고를 일으키는 필연적인 요소가 그 지점에 내재되어 있다는 것이다. 따라서 그 지점은 교통사고위험지점으로 분류되어 교통안전프로그램에서 우선적으로 개선이 이루어져야 한다.

2. 과분산을 반영한 한계사고율법의 적용

1) 자료설명

위해서 개발된 방법에 대한 이해를 용이하게 하기 위해서 실제 교통사고자료를 이용하여 적용과정을 살

펴보았다. 자료는 앞에서 미국 미시간주의 교통사고 관리시스템에 구축된 데이터베이스를 이용하였다. 교통사고관리시스템에는 도로에서 발생한 각각의 교통사고를 사고지점, 연평균 일 교통량, 사고형태, 사고의 심각도, 사고원인, 도로의 구조 등의 정보가 데이터 베이스화 되어있다.

본 연구에서는 많은 종류의 인터체인지중 32개의 다이아몬드형 인터체인지(즉 Reference sites가 32개란 의미임)를 선택하여 분석대상지점으로 설정하였으며 분석대상기간은 최근 3년간(1998~2000)이다. 미시간주의 고속도로 인터체인지는 22개의 형태로 분류되며 각 인터체인지마다 그 형태가 지정되어 있다. 이중 다이아몬드형은 고속도로에 적용되는 일반적인 인터체인지로서 본선과 4개의 유·출입 램프로 구성되어 있다.

인터체인지의 범위는 최초램프와 최종램프로부터 500 피트(152M) 전·후방의 본선구간과 램프구간으로 미시간주의 교통사고관리시스템에 정의되어 있다.

최근 3년 동안 발생한 교통사고자료만을 사용한 이유는 위험지점의 선정 시 도로의 실상을 반영하기에는 최근 3년 정도의 자료가 유효하다는 선행연구에 기초를 두고 있다(Stokes and Mutabazi 1996). 본 연구에서 사용한 데이터인 교통사고와 교통량을 요약하면 <표 2>와 같다.

<표 2> 샘플지점의 교통사고와 교통량 요약

구분		교통사고 및 교통량		
		최대	최소	평균
교통 사고	교통사고건수	322	23	131.2
	사망사고건수 (사망자수)	12 (18)	1 (3)	4.3 (5.8)
	부상사고건수 (부상자수)	198 (228)	8 (10)	35.2 (40.4)
교통량	일평균 교통량 (차량대수/일)	240,912	50,221	153,830

2) 적용과정

실제 자료를 이용하여 앞에서 제시한 방법을 적용하였는데 그 과정은 다음과 같다(<표 3> 참조).

1단계 : 비슷한 특성을 가진 지점(Reference site)의 평균사고율(λ) 산정

$$\lambda = \frac{\sum N_i}{\sum V_i} = \frac{4,195 \text{ crashes}}{4,922,555 \text{ vehicles}}$$

$$= 0.00085 (\text{crashes/vehicle})$$

2단계 : 각 지점에서 발생한 교통사고의 기대값 (μ_i)
예측

$$\mu_i = \lambda \cdot V_i$$

$$\mu_1 = \lambda \cdot V_1 = 0.00085 \cdot 84395 = 72 \text{ crashes}$$

.....

$$\mu_{32} = \lambda \cdot V_{32} = 0.00085 \cdot 132410 = 113 \text{ crashes}$$

3단계 : 포아송분포라는 가정에서 확률변수의 표준화
값(X_i) 결정 :

$$X_i = \frac{N_i - \mu_i}{\sqrt{\mu_i}}$$

$$X_1 = \frac{N_1 - \mu_1}{\sqrt{\mu_1}} = \frac{76 - 72}{\sqrt{72}} = 0.48$$

.....

$$X_{32} = \frac{N_{32} - \mu_{32}}{\sqrt{\mu_{32}}} = \frac{90 - 113}{\sqrt{113}} = -2.15$$

4단계 : 분산과 관련된 파라미터(d) 산정 :

$$d = \text{Variance}(X_1, X_2, X_3 \dots \dots X_{16}) = 38.47$$

5단계 : 파라미터(d)를 사용하여 확률변수(N_i)의 표
준화 :

$$Z_i = \frac{N_i - \mu_i}{\sqrt{d\mu_i}}$$

$$Z_1 = \frac{N_1 - \mu_1}{\sqrt{d\mu_1}} = \frac{76 - 72}{\sqrt{38.47 \times 72}} = 0.08$$

$$Z_2 = \frac{N_2 - \mu_2}{\sqrt{d\mu_2}} = \frac{39 - 43}{\sqrt{38.47 \times 43}} = -0.09$$

.....

$$Z_{32} = \frac{N_{32} - \mu_{32}}{\sqrt{d\mu_{32}}} = \frac{90 - 113}{\sqrt{38.47 \times 113}} = -0.35$$

6단계 : 사고위험지점의 선정 :

5단계에서 결정된 각 지점의 한계사고율과 관측된
교통사고율을 비교하여 위험지점을 선정 할 수 있다.

예를 들면, 95% 유의수준에서 교통사고위험지점을
선정하기로 결정했을 때, 20번째 지점의 "Z" 값은
2.45로서 주어진 유의수준에서의 "Z"값(1.64)보다
크다. 따라서 20번째 지점은 위험지점으로 선정된다.

3. 적용결과

본 연구에서 제시된 방법을 관측된 교통사고자료에
적용시켰으며 기존의 방법인 포아송분포에 기반한 한
계사고율법과의 비교가 이루어졌다. <표 3>에서 보여
주는 것처럼 파라미터 λ 는 0.00085, d는 38.47 이
며 이중 λ 는 포아송분포에 기반한 한계사고율법에도
똑같이 적용되는 파라미터이다. 산정된 파라미터를
사용한 분석결과, 95%의 유의수준에서 포아송분포에
기반한 한계사고율법은 32개의 분석지점중에서 10개
의 지점을 사고위험지점으로 선정(지점번호: 6, 7,
8, 16, 17, 18, 19, 20, 21, 23)했으며 새로운 방
법은 1개의 지점(지점번호: 20)을 위험지점으로 선
정 하였다(<표 3>의 *).

N. 개발된 방법의 타당성 평가 및 논점

1. 개발된 방법의 타당성 평가

본 연구에서 제시한 방법의 타당성평가는 두 가지
방법으로 이루어졌다. 첫째는 식(3)의 표준화된 확률
변수 X_i 가 정규분포를 따른다는 것을 증명하는 것이
고, 둘째는 본 연구에서 제시된 방법을 이용한 분석
결과가 통계학적으로 유의한 가를 분석하는 것이다.

타당성평가의 첫째 방법인 확률변수 X_i 가 정규분포
를 따르는가를 검증하기 위해서 정규분포를 확률적으
로 8개의 동일한 구간으로 구분하여 카이제곱 검정
(Chi Squared test)이 이루어졌다. 본 연구의 적
용과정에서 이용한 자료를 토대로 검증해본결과, χ^2
값은 4.5로서 $\chi^2_{0.05,5} = 11.07$ 보다 작았다. 이것은
확률변수 X_i 가 정규분포를 따른다는 것을 의미하며 본
연구에서 제시한 방법의 개발과정상의 가정이 통계학
적으로 문제가 없음을 나타낸다(Neter et al. 1993).

타당성 평가의 둘째 방법인 본 연구에서 제시된 과
분산을 고려한 한계사고율법을 관측된 자료에 적용했
을 때 적합한 결과를 도출하는가의 여부를 검토하기
위해서 기존의 방법인 포아송분포에 기반한 한계사고

〈표 3〉 분석결과 비교

지점 (i)	사고건수 (N _i)	교통량 (V _i)	교통 사고율 (N _i /V _i)	분석 I		μ_i	X _i	분석 II	
				한계 사고율	위험 지점			Z _i	위험 지점
1	76	84395	0.00090	0.00102		72	0.48	0.08	
2	39	50221	0.00078	0.00106		43	0.58	0.09	
3	48	60199	0.00080	0.00104		51	0.46	0.07	
4	60	78622	0.00076	0.00102		67	0.85	0.14	
5	57	107259	0.00053	0.00099		91	3.6	0.58	
6	137	123131	0.00111	0.00098	*	105	3.13	0.51	
7	196	128354	0.00153	0.00098	*	109	8.29	1.33	✓
8	178	108803	0.00164	0.00099	*	93	8.86	1.42	✓
9	95	121717	0.00078	0.00098		104	0.85	0.14	
10	201	184762	0.00109	0.00096		157	3.47	0.56	
11	80	206354	0.00039	0.00095		176	7.23	1.17	
12	36	109813	0.00033	0.00099		94	5.95	0.96	
13	193	240902	0.0008	0.00095		205	0.85	0.14	
14	194	227992	0.00085	0.00095		194	0.02	0.00	
15	49	92071	0.00053	0.00101		78	3.32	0.54	
16	154	105444	0.00146	0.00099	*	90	6.77	1.09	
17	247	171559	0.00144	0.00096	*	146	8.34	1.34	
18	213	150529	0.00142	0.00097	*	128	7.48	1.21	
19	207	161545	0.00128	0.00097	*	138	5.91	0.95	✓
20	322	166103	0.00194	0.00096	*	142	15.17	2.45	
21	222	197572	0.00112	0.00096	*	168	4.14	0.67	
22	160	183034	0.00087	0.00096		156	0.32	0.05	* ✓
23	242	179213	0.00135	0.00096	*	153	7.23	1.16	
24	102	207928	0.00049	0.00095		177	5.65	0.91	
25	111	203816	0.00054	0.00095		174	4.75	0.77	
26	158	207045	0.00076	0.00095		176	1.39	0.22	
27	161	217719	0.00074	0.00095		185	1.8	0.29	
28	94	196824	0.00048	0.00096		168	5.69	0.92	
29	27	175279	0.00015	0.00096		149		1.61	
30	23	179471	0.00013	0.00096		153		1.69	
31	23	162466	0.00014	0.00097		138		1.58	
32	90	132410	0.00068	0.00098		113		0.35	
				$\lambda=0.00085, \quad d=38.47$					

주 : 분석 I : 포아송분포에 기반한 한계사고율법
 분석 II : 과분산을 고려한 한계사고율법
 * : 95% 유의수준 적용, ✓ : 90% 유의수준 적용

율법에 의해서 도출된 결과와 비교하였다. 〈표 3〉에서 보는 것처럼 보는 것처럼 95%의 유의수준하에서 포아송분포에 기반한 한계사고율법은 32개의 분석지점중에서 10개의 지점을 사고위험지점으로 선정하였으며 새로운 방법은 1개의 지점을 위험지점으로 선정하였다. 통계학적 이론에 의하면 95%의 유의수준하에서는 평균적으로 5%(또는 임의의 20개지점 중에서 1지점)가 한계사고율을 벗어나야 한다(Rice 1997).

따라서 임의의 32개지점 중에서 10개의 지점이 한계사고율 밖에 있다는 것은 통계학적으로 기대하는 수치보다 훨씬 많은 지점이 한계사고율을 벗어났다는

것을 의미한다. 따라서 과분산이 발생한 교통사고자료의 경우, 포아송분포에 기반한 한계사고율법은 과분산을 반영할 수 없기 때문에 〈그림 1〉에서 보는 바와 같이 개념적으로 더 많은 지점을 위험지점으로 선정할 수 있다는 논리와 본 연구에서 제시한 방법에 의해서 과분산으로 발생한 문제점은 극복될 수 있다는 사실이 증명된 셈이다.

한편 95%의 유의수준을 적용한 것과 같은 개념으로 90%하에서는 평균적으로 10%인 3~4개의 지점이 한계사고율을 벗어나야 한다. 90% 유의수준을 적용한 경우, 4개의 지점(지점번호 : 7, 8, 17, 20)의

"Z" 값이 주어진 유의수준에서의 "Z" 값(1.282)보다 크므로 위험지점으로 선정된다(〈표 3〉의 \checkmark). 따라서 95% 유의수준을 적용했을 때와 마찬가지로 95% 유의수준에서도 통계적 개념과 부합된다고 할 수 있다.

2. 위험지점선정상의 논점

기존의 한계사고율법을 과분산이 발생한 교통사고 자료에 적용시, 투자가 필요치 않은 지점까지 위험지점으로 선정 할 수 있어 결국에는 교통안전재원의 효과적인 배분에 장애가 된다. 한계사고율법을 사용하여 위험지점을 선정할 때 과분산에 의해서 발생하는 개념적인 한계는 본 연구에서 제시하는 간단한 통계학적인 과정에 의해서 해결될 수 있다.

본 연구와 관련하여 교통안전전문가들이 제기하기 쉬운 몇 가지 논점을 정리하면 다음과 같다. 첫째, 본 연구에서 제시된 한계사고율법을 적용하여 위험지점의 선정 시 적용 전에 항상 교통사고자료의 과분산 유무를 측정해야 하는가 라는 논점이다. 본 연구에서 제시한 방법은 앞의 식(4)에서 파라미터인 "d"가 과분산의 정도를 반영하므로 교통사고자료의 과분산의 유무에 대한 측정없이 사용 될 수 있다. 본 연구의 자료에서는 과분산의 정도가 크므로 "d"값이 38.47로 추정 되었지만 여기에서 제시한 방법을 적용시 과분산이 발생하지 않았거나 혹은 아주 미미한 과분산이 발생하는 경우 "d"는 1.0에 매우 근접한 값이 추정되어 오차정도는 포아송분포의 오차구조에 근접하게 되므로 본 연구에서 제시한 방법에 의한 결과와 포아송에 기반한 기존의 한계사고율법에 의해서 도출된 결과가 유사하게 된다. 따라서 본 연구에서 제시한 방법은 과분산 유무에 관계없이 사용될 수 있으며 인터체인지 외에도 교차로나 도로의 구간 등에도 적용 될 수 있다.

둘째, 본 연구에서 제시한 방법이 실제보다 적은 지점수를 선정하여 교통안전사업의 범위를 제한 하는 것이 아닌가? 하는 의문이다. 과분산이 발생한 경우, 본 연구에서 제시한 방법이 포아송분포에 기반한 기존의 한계사고율법에서 제시한 것보다 적은 수의 지점을 선정토록 하는 것은 사실이며 이론적으로 당연한 결과이다. 그렇다고 제시한 방법이 적용상의 장애를 가져오는 것은 아니다. 만약 주어진 예산이 충분하여 선정된 지점보다 더 많은 지점의 교통안전개선

이 가능하다면, 한계사고율을 낮추어(예: 95% 유의수준이 아니라 90%의 유의수준을 적용하는 것 등) 개선지점의 범위를 확대시키는 것이 타당하지 이론적인 측면에서 한계를 갖는 방법을 이용하여 억지로 개선지점의 범위를 확대하는 것은 올바른 접근방법이 아니다.

한계사고율법은 교통사고 전문가들이 인식하는 한계가 있다(Hauer and Persaud 1986, Hauer 1992). 첫째, 비슷한 특성을 가진 지점(Reference site)의 선택이 분석자에 의해서 임의로 결정되어 똑 같은 지역이 분석자에 의해서 다르게 평가 될 수 있다는 것으로 개념상의 한계이다. 둘째, 어떤 경우에는 한계사고율법을 적용 할 수 있을 정도로 충분히 많은 비슷한 지점(Reference site)이 존재하지 않는 경우가 발생 할 수 있다는 것으로서 적용상의 한계이다. 즉, 5지 교차로는 3지 교차로나 4지 교차로에 비하여 비슷한 지점을 찾기가 매우 어려운 예로서 한계사고율법을 적용하기가 쉽지 않다. 이런 문제점을 해결코자 Hauer은 회귀분석모델을 이용한 경험적 베이저언 방법(Empirical Bayesian Method)을 개발하였다(1992). 그러나 경험적 베이저언 방법은 베이저언 이론에 대한 지식, 비선형 모델구조의 결정, 유의한 설명변수의 선정, Spss와 S-plus 등 기존의 통계프로그램으로는 불가능한 음이항분포에 기반한 모델 파라미터의 산정 등 매우 전문적인 통계학적 지식이 요구된다.

교통안전프로그램의 1단계인 교통사고위험지점을 선정하기 위한 많은 방법들이 있다. 많은 방법들 중 어떤 방법을 선택 할 것인가는 교통안전사업의 규모, 구축된 교통사고관련 데이터베이스의 내용, 자원 등 여러가지 요인에 달려있다. 우리나라의 경우, 교통사고관련 데이터베이스 구축이 미비하여 경험적 베이저언 방법을 사용하기는 매우 어려울 것이다. 또한 교통사고건수법을 이용하여 위험지점을 선정하는 것은 간단한 방법이지만 교통사고에 내재된 불확실성을 통계학적으로 해석할 수 없다는 단점 외에 교통량에 대한 고려가 없다는 단점이 있다. 예를 들면 A라는 지점의 일평균 교통량이 1000대, 연평균 교통사고가 10건이라 가정하고, B라는 지점은 일평균 교통량이 10,000대, 연평균 교통사고가 10건이라 가정했을 때, 교통사고건수에 의한 방법은 두 지점을 같은 수준의 위험을 지닌 지점으로 간주하게 된다. 그러나 두 지점에 대한 개선이 이루어 졌을 때 개선효과는 A 지점이 일반

적으로 크다. 그 이유는 B지점에서 일어난 교통사고 중 상당부분은 교통안전개선사업에서 치유가 어려운 운전자의 과실일 가능성이 크기 때문이다. 이러한 개념적인 문제점을 해결할 수 있고 교통사고의 불확실성을 통계학적으로 해석 할 수 있다는 측면에서 교통사고율법은 충분히 타당성을 지닌 방법이며 도로교통량 통계연보 등을 통하여 교통량이 매년 집계되어 발표되므로 우리나라에서도 충분히 적용 가능한 방법이다. 이외에도 어떤 특정지점이 위험하다고 운전자로부터 자주 불만이 제기될 때 쉬운 방법은 비슷한 지역(Reference site)을 선정하여 교통사고율을 계산한 후 그 특정지점의 교통사고율과 비교하여 통계학적인 해석을 내리는 방법인 한계사고율법이 적합할 것이다.

V. 결론

도로 및 교통여건이 변화하는 한 교통사고자료의 과분산 현상은 흔히 발생하는 현상이다. 본 논문은 과분산이 발생한 경우, 포아송분포에 기반한 한계사고율법에 내제되어있는 개념적인 문제점을 분석하고 극복할 수 있는 방법을 제시하였다.

32개의 도로특성이 비슷한 지점을 별도로 선정하여 포아송분포에 기반한 한계사고율법과 본 연구에서 제시한 방법을 적용하여 그 결과를 비교 분석하였다.

분석결과 95%의 유의수준에서 포아송분포에 기반한 방법은 32개의 지점중 10개의 위험지점을 인식하였고 본 연구에서 제시한 방법은 1개의 위험지점을 인식 하였다. 통계학적 이론에 의하면 95%의 유의수준하에서는 평균적으로 5%(또는 임의의 20개 지점중에서 1지점)가 한계사고율을 벗어나야 한다. 따라서 개념적인 측면에서 본 연구에서 제시한 방법이 기존의 방법보다 타당한 결과를 산출해 낸다고 할 수 있다. 기존의 방법을 적용 할 때 이론적인 수치보다 더 많은 위험지점을 인식하는 이유는 기존의 포아송분포에 기반한 한계사고율법이 도로 및 교통여건의 변화에 의해서 생기는 과분산을 충분히 반영하지 못하기 때문이다.

끝으로 본 연구에서 제시한 방법의 적용을 위해서 모든 교통사고자료에 대해서 과분산의 유무를 알아야 하는가 라는 논점이 남는다. 그러나 본 연구에서 제시한 방법은 과분산의 유무에 관계없이 적용될 수 있

다. 그 이유는 과분산이 발생치 않은 경우, 본 연구에서 제시한 방법을 적용시 과분산과 관련된 파라메터인 "d" 가 1.0에 근사하여 기존의 포아송분포에 기반한 한계사고율법과 유사한 결과를 도출할 것이기 때문이다.

참고문헌

1. 사고찾은곳 개선사업 업무편람(2002). 건설교통부, pp.19~26.
2. 도철웅(1991), 교통공학원론(하), pp.728~731.
3. Hauer, E.(1992), Empirical Bayes approach to the estimation of 'unsafety': the multi-variate regression method. Accident Analysis and Prevention, Vol.24, No.5, pp.457~477.
4. Hauer, E.(1996), Identification of sites with promise. Transportation Research Record, 1542, Transportation Research Board, National Research Council, Washington D.C., pp.54~60.
5. Hauer, E., and Persaud, B. N.(1984), Problem of identifying hazardous locations using accident data. Transportation Research Record, 975, Transportation Research Board, National Research Council, Washington D.C., pp.31~43.
6. Hauer, E., and Persaud, B. N.(1986), How to estimate the safety of rail highway grade crossings and the safety effects of warning devices. Transportation Research Record, 1114, Transportation Research Board, National Research Council, Washington D.C., pp.131~139.
7. Hauer, E., Persaud, B. N., Smiley, A., and Duncan, D.(1991), Estimating the accident potential of an Ontario driver. Accident Analysis and Prevention, Vol.23, No.2/3, pp.133~152.
8. Hagle, J. L., and Witkowski, J. M.(1988), Bayesian Identification of Hazardous Locations. Transportation Research Record, 1185. Transportation Research Board, National Research Council, Washington D.C., pp.24~36.

9. Hutchinson, T. P., and Mayne, A. J.(1977), The year to year variability in the numbers of road accidents. Traffic Engineering and Control, Vol.18, p.432.
10. Maher, M. J., and Summersgill, I.(1996), A comprehensive methodology for the fitting of predictive accident models. Accident Analysis and Prevention, Vol.28, No.3, pp. 281~296.
11. Neter, J., Wasserman, W., and Whitmore, G. A.(1993), Applied statistics: 4th edition. A Division of Somon & Schuster, Inc.
12. Norden, M., Orlansky, J., and Jacobs, H. (1956), Application of Statistical Quality-Control Technique to analysis of highway-accident data. Bulletin 117, HRB, National Research Council, Washington D.C., pp.17~32.
13. Rice, J. A.(1997), Mathematical statistics and data analysis: 2nd edition. Wadsworth, Inc.
14. Stokes, R. W., and Mutabazi, M. I.(1996), Rate Quality Control Method of identifying hazardous road locations. Transportation Research Record, 1542, Transportation Research Board, National Research Council, Washington D.C., pp.44~48.
15. Tarko, A. P., Sinha, K. C., and Farooq, O. (1996), Methodology for identifying highway safety problem areas. Transportation Research Record, 1542, Transportation Research Board, National Research Council, Washington D.C., pp.49~53.
16. Zegeer, C. V., and Deen, R. C.(1977), Identification of Hazardous Locations on City Streets. Traffic Quarterly, pp.601~610.

✉ 주 작성자 : 성낙문

✉ 논문투고일 : 2003. 11. 18

논문심사일 : 2003. 12. 17 (1차)

2004. 1. 30 (2차)

심사판정일 : 2003. 1. 30

✉ 반론접수기한 : 2004. 6. 30