

■ 論 文 ■

# Empirical Bayes Method를 이용한 교통사고 예측모형

## A Study on the Traffic Accident Estimation Model using Empirical Bayes Method

강 현 건

(도로교통공단 연구원)

강 승 규

(계명대학교 교통공학과 부교수)

장 용 호

(계명대학교 교통공학과 석사과정)

### 목 차

- |   |  |
|---|--|
| <ul style="list-style-type: none"> <li>I. 서론                     <ul style="list-style-type: none"> <li>1. 연구의 배경 및 목적</li> <li>2. 연구의 범위와 방법</li> </ul> </li> <li>II. 문헌고찰                     <ul style="list-style-type: none"> <li>1. 국내기준</li> <li>2. 국외기준</li> <li>3. 경험적 베이즈 기법 적용사례</li> </ul> </li> <li>III. 자료수집 및 분석                     <ul style="list-style-type: none"> <li>1. 자료수집</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>2. 분석</li> <li>IV. 교통사고 예측모형                     <ul style="list-style-type: none"> <li>1. 안전성능함수</li> <li>2. 과분산 파라메타(<math>\phi</math>)</li> <li>3. 가중치(<math>\alpha</math>)</li> <li>4. 예상 사고건수(<math>\pi</math>)</li> </ul> </li> <li>V. 결론 및 향후과제</li> <li>참고문헌</li> </ul> |
|---|--|

Key Words : 경험적 베이즈 방법, 사고 잦은지점, 대물피해환산법, 안전성능함수, 교통사고 예측  
 Empirical Bayes Method, Hazardous Road Locations, Equivalent Property Damage Only,  
 Safety Performance Function, Traffic Accident Estimation

### 요 약

본 연구는 경북도내에서 발생한 4년간의 교통사고 자료를 대상으로 Empirical Bayes (EB) 방법을 이용하여 예상 사고건수를 예측하였다. 경북도내 각 군과 시 지역의 교통사고는 대물피해환산법을 적용하여 심각도를 반영하였으며, EB 방법을 적용하기 위해 군집분석을 통해 유사한 지역을 선정하였고, 선정된 유사지역을 대상으로 각 지역별 안전성능함수(SPF)를 도출하였다. 실제 사고건수와의 근원적인 확률분포를 일치시키기 위해 과분산 파라메타를 산출하였으며, 지역별 교통특성을 반영하기 위해 가중치를 적용하여 예상 사고건수를 예측하였다. 분석 결과 김천시, 영천시, 칠곡군 순으로 가장 높은 사고건수가 예상되는 반면, 군위군이 가장 낮은 사고건수가 발생할 것으로 예측되었다.

This study estimates the expected number of accidents in Kyungbuk Province to capitalize on experience gained from four years of accident history using the Empirical Bayes (EB) Method. The number of accidents of each site in Kyungbuk Province is recalculated using the Equivalent Property Damage Only (EPDO) method to reflect the severities of the accidents. A cluster analysis is performed to determine similar sites and a unique Safety Performance Function (SPF) is established for each site. The overdispersion parameter is built to correct the difference between the actual number of accidents and the underlying probability distribution. To adjust for varying traffic characteristics of each site, a relative weight is applied and eventually estimates the expected number of accidents. The results show that the highest accident sites are Kimcheon, Youngcheon, and Chilgok, but on the other hand the lowest is Gunwi.

본 연구는 계명대학교 비사연구비의 지원으로 수행되었음.

## I. 서론

### 1. 연구의 배경 및 목적

우리나라에서 2007년에 경찰에서 접수·처리한 교통사고는 211,662건으로 6,166명이 사망하고 335,906명이 부상을 당하여, 하루 평균 16.9명이 사망하고 920명이 부상을 당한 것으로 나타났다. 국민의 안전의식 제고와 더불어 교통사고 감소를 위한 국가적 안전정책의 수립 및 시행을 통한 국가 경쟁력 제고가 절실히 요구되는 실정이다.

정부에서는 범국가적인 차원에서 교통사고 줄이기 운동을 지속적으로 추진해오고 있는데 무엇보다도 교통사고에 대한 심층적인 원인분석을 근거로 한 교통안전대책이 수립되고 수행되어야 한다. 이러한 이유로 도로교통공단에서는 1987년 이후 『교통사고 잦은 곳 개선사업』을 시행하게 되었으며, 2008년 ‘선진교통문화 정착’을 디딤돌로 삼아 국가발전과 국민의 ‘삶의 질’향상을 이루고자, 교통사고 사상자를 향후 5년내 반으로 줄이는 『교통사고 사상자 절반 줄이기 프로젝트』를 추진하고 있다. 교통사고를 지속적으로 감소시키기 위해서는 지도, 단속, 계몽 및 홍보뿐만 아니라 더 근본적이며 기본적인 개선대안인 도로환경의 안전성확보가 중요하다. 도로의 안전은 선형 자체의 안전성 확보가 기본적인 요소이며, 이와 함께 운전자와 보행자의 안전성을 제고시키기 위하여 도로안전시설물 설치가 필요하다. 우리나라의 경우 『교통사고 잦은 곳 개선사업』에 많은 재원과 인력이 투입되고 있지만, ‘단순사고건수비교방법’에 의하여 사고 잦은지점을 선정하는데 그쳐 실제사고가 교통안전시설물로 인한 직접적인 영향에 의한 것인지 우연한 원인에 의하여 교통사고가 증가하거나 감소하는 것인지 명확하게 분석할 수 없다. 보다 적절한 통계적 분석방법으로 장래 교통사고 건수를 예측하여 한정된 예산으로 개선이 필요한 지점에 대하여 우선순위를 부여하여 효율적인 교통사고 잦은지점에 대한 개선사업이 시행되어야 할 것으로 판단된다.

본 연구는 현재 ‘단순사고건수비교방법’으로 실시하고 있는 교통사고 잦은지점 개선사업에 대한 문제점을 지적하고 이에 대한 적절한 기준과 경험적 베이지안 (EB, Empirical Bayes) Method를 이용하여 교통사고 예측모형을 제시하고자 한다.

이를 위해 최근 4년(2004년~2007년)간 경상북도에서 발생한 교통사고를 대상으로 인구, 자동차 등록대



<그림 1> 연구대상지역

수, 도로연장과 교통사고와의 상관분석을 통하여 실제 교통사고에 영향을 미치는 인자를 도출하고 지역별 안전성함수(SPF, Safety Performance Function)의 개발을 통하여 최종적으로 지역별로 장래에 발생할 교통사고를 예측하고자 한다.

### 2. 연구의 범위와 방법

#### 1) 연구 범위

본 연구의 공간적 범위는 <그림 1>과 같이 경상북도 내 행정 관할구역 중 지역특성상 울릉군을 제외한 22개 지역(시: 10개, 군: 12개)을 대상으로 한다.

시간적 범위는 2004년부터 2007년까지 4년간 경상북도에서 발생한 사고 자료를 이용하였으며, 관할지역내의 사고 유형별 교통사고건수와 사회지표인 인구, 자동차 등록대수, 도로연장을 대상으로 분석하였다.

#### 2) 연구 방법

본 연구에 앞서 선행연구 사례와 연구결과를 고찰하였다. 기존의 사고 잦은지점 선정방법의 문제점을 파악하고 EB Method를 적용하기 위한 타당성을 검토하였고, 대상 지역을 선정 후 군집분석을 통하여 유사지역을 선정하였다. 지역별 SPF모형을 산정하기 위해 유사지역으로 선정된 지역간의 변수(인구, 자동차 등록대수, 도로연장길이)를 대상으로 상관분석을 실시하여 교통사고에 영향을 미치는 변수들을 파악하였으며 이들 변수를 이용하여 다중회귀분석을 실시하였다. 최종적으로 산정된 SPF모형을 EB Method에 적용하여 각 지역별 예상 사고건수를 예측하였다. <그림 2>는 본 연구의 진행과정을 나타낸 것이다.



<그림 2> 연구 진행도

## II. 문헌고찰

### 1. 국내기준

#### 1) 교통사고 잦은지점 선정기준

『교통사고 잦은 곳』이란 교통사고 발생건수가 동일지점에서 1년간 특별시 및 광역시는 7건, 일반시는 5건, 기타지역은 3건 이상 발생한 지점(사망사고가 2건 이상 발생하는 자료는 현행 선정기준 보다 1건씩 적게 적용)이며, 공간적 지점 범위는 교차로 및 횡단보도는 차량 정지선에서 후방으로 30m이내이고, 단일로는 시가지의 경우 반경 100m 이내, 기타 단일로와 고속도로의 경우 반경 200m이내를 기준으로 한다. <표 1>은 사고 잦은지점 선정기준을 나타낸 것이다.

#### 2) 국토해양부 사고 잦은지점 선정방법

국토해양부의 사고 잦은지점은 급격한 곡선반경, 중

<표 1> 교통사고 잦은지점 선정기준

구분		선정기준	
지역	특별시	7건 이상	
	광역시	7건 이상	
	일반시	5건 이상	
	기타	3건 이상	
도로 형태	교차로 및 횡단보도		차량 정지선에서 후방으로 30m이내
	기타 단일로	시가지	반경 100m이내
		기타, 고속도로	반경 200m이내
	대상 사고		인피+물피사고

단경사 및 시거 등 도로의 기하학적 특징을 토대로 점수화하고 그 지역의 교통량을 반영하였다. 과거에 발생한 사고건수를 감안하고 기하구조 및 추가사고에 대하여 그 현장의 누적된 점수에 10%를 추가하고 있으나 최악의 현상이 낮은 우선순위로 주어질 가능성이 있다. 이러한 문제점을 개선하기 위하여 사고지점을 지적하는 우선순위 등급점수 체계를 도입하였고 최악의 50개 사고다발 지역을 선정하여 더 상세한 비율을 파악할 때까지 대물피해환산법과 동일한 가중비율을 적용하고 있다.

## 2. 국외기준

### 1) Hazardous Road Locations의 정의

외국의 경우 “위험도로” 또는 “교통사고 잦은 곳”과 유사한 개념으로 Hazardous Roads 또는 Black Spots를 이용하고 있다. Hazardous Road Locations에 관한 여러 정의가 제시되고 있으나 일반적으로 함유점이나 곡선부, 도로의 교차로와 같은 기하구조 형태에 의해 규정되는 특정 위치에서 발생하는 사고군과 관련된 “Black Spots”의 하나로 정의되고 있다.

### 2) Hazardous Road Locations의 선정방법

단일 지점에 대한 사고다발지역의 선정방법은 가장 단순한 사고건수만을 집계하는 **Number of Accidents Method**를 하는데, 사고의 건수가 많은 곳을 사고다발지역으로 선정하게 된다.

위 방법보다 교통사고 노출을 고려한, 즉 교통량을 고려한 방법으로는 **Rate of Accidents Method**가 있다. 사고율을 구간길이 또는 지점 단위로 비교해 사고율이 높은 곳이 사고다발지역으로 선정된다.

하지만 사고율로 사고다발지역을 선정할 경우 교통량이 아주 적은 지역에서 발생하는 사고가 높은 사고율로 부각될 수 있어, 그런 점을 보완한 방법은 **Number-Rate Method**를 적용할 수 있다. 즉, 단순 사고건수에 의한 방법과 사고율에 의한 방법을 서로 보완한 방법으로서 평균 사고건수와 사고율의 2배를 초과하는 지역을 사고다발지역으로 선정하게 된다. 주로 많은 교차로와 도로구간이 있는 대도시에 적용하고 있다.

교통량의 변화에 따라 모든 크기의 도시에서 적용이 가능한 방법으로 한 지역 또는 도로구간이 이미 설정한 평균 사고율보다 비정상적으로 높은지 통계적 방법으로 분석하

는 **Rate Quality Control Method**를 적용하기도 한다. 통계적 분석은 일상적으로 사고의 발생이 포아송 분포와 일치한다는 것에 기초한다. 이 방법의 효과로 사고건수는 많으나 교통량이 적은 지역을 고려대상에서 제외하게 된다. 이 방법은 대상지역에 대한 임계 교통사고율을 계산하고, 한 지역의 사고율이 임계사고율보다 높으면 사고다발지역으로 선정한다. 임계값은 유사한 도로에서의 평균 교통사고율을 이용하여 아래식과 같이 계산한다.

$$R_c = R_a \pm K \cdot \sqrt{\frac{R_a}{M} \pm \frac{1}{2M}}$$

여기서,

$R_c$  : 대상지역의 임계교통사고율(Critical Rate)

$R_a$  : 유사한 도로에서의 평균교통사고율

$K$  : 확률계수

$M$  : 대상지역의 교통사고 노출량

$$M = \frac{(ADT) \cdot (365) \cdot (\text{구간길이}) \cdot (\text{년수})}{10^8}$$

사망사고, 부상사고, 대물사고의 각 피해 종류를 등가로 환산해서 하나의 피해단위로 나타내어 비교하는 방법으로 대물피해환산법(Equivalent Property Damage Only)을 적용하기도 한다. 부상사고 한 건당 전국의 평균 비용 및 사망사고 한 건당 평균비용을 대물사고만 발생한 사고 한 건당 평균피해액으로 나누어 부상사고 및 사망사고의 대물사고에 대한 가중치(Se : Severity Factor)를 구한다. 아래와 같이 지점이나 구간에 대하여 사망사고와 부상사고를 대물사고의 각각 12배와 3배로 가중하여 환산하는 방법이다.

□ 지점

$$S_e = \frac{[12 \cdot (Fatal) + 3 \cdot (Inj) + (PDO)] \cdot 10^8}{365 \cdot (\text{년수}) \cdot (ADT)}$$

□ 구간

$$S_e = \frac{[12 \cdot (Fatal) + 3 \cdot (Inj) + (PDO)] \cdot 10^8}{365 \cdot (\text{년수}) \cdot (ADT) \cdot (\text{구간길이})}$$

여기서,

$S_e$  : 사고의 가중치

$Fatal$  : 사망사고 발생건수

$Inj$  : 부상사고 발생건수

$PDO$  : 대물사고 발생건수

$ADT$  : 평균일교통량

### 3. 경험적 베이즈(Empirical Bayes) Method 적용사례

**Hauer**는 EB Method를 이용하여 교통사고 자료의 변동(Fluctuation)을 보정하기 위해서 어떤 지점에서 발생한 교통사고는 여러 가지 요인들로 인해서 변동적일 수 있으므로 참조그룹에서 발생한 교통사고를 토대로 보정되어야 한다는 개념을 제시하였다.

이 방법은 교통사고의 해석시 노출도인 교통량을 고려할 수 있고 교통사고의 불확실성을 반영할 수 있다는 측면에서 우수한 방법이다. 또한 유사지역의 교통사고 특성을 이용하여 다른 방법과 관련된 이론적 혹은 적용상의 문제점을 해결할 수 있다는 장점을 가지고 있다.

박민호 등(2006)은 EB Method를 이용하여 중앙분리대 설치에 따른 사고전환효과 분석을 실시하였다. 분석 결과, 중앙분리대 설치로 인해 정면충돌 사고는 감소하였으나 곡선부에서의 후미추돌사고 그리고 측면사고는 증가하는 것으로 나타났다. 또한, 사고심각도 측면에서 심각도가 높은 사망사고는 감소하지만 심각도가 다소 낮은 부상사고는 증가하는 것으로 제시하고 있다.

박규영 등(2006)은 도로안전시설물의 효과도를 비교하였는데 EB Method를 이용하여 시설물별 효과도 산정모형을 구축하였고 승산비(Odds Ratio)를 이용하여 상대효과도 산정기법을 제시하였다. 분석 결과, 중앙분리대와 가드레일은 상대효과가 크지 않게 나타났으나, 갈매기표지나 미끄럼방지포장 등은 구간사업에서 상대효과도가 매우 크게 나타나 향후 시설물 설치 시 구간측면에서 고려가 필요함을 제시하고 있다.

성낙문(2003)은 도로의 안전도를 평가하는데 있어서 교통사고에 내재되어 있는 특성(교통사고의 불확실성, 복잡한 교통사고 요인, 각 지점에 내재되어 있는 고유의 사고위험성)을 반영하는 방안으로 교통사고예측모델을 이용하여 도로의 안전도를 평가하는 방법을 제시하였다.

정성봉 등(2009)은 EB Method를 이용하여 사고곳은 곳 개선사업 우선순위 판정기법을 개발하였는데, 총 사고건수 추정 시 기존의 사고 예측방법인 비선형 회귀 모형과 EB Method의 결과 값 모두 예측력이 높은 것으로 나타났지만 지점별 사고건수 예측력을 함께 고려할 경우엔 EB Method가 우수하다고 발표하였다.

<표 2> 교통사고 통계자료

구분	2004년			2005년			2006년			2007년		
	사 망	중 상	경상 및 물피	사 망	중 상	경상 및 물피	사 망	중 상	경상 및 물피	사 망	중 상	경상 및 물피
군위군	15	86	52	6	93	33	11	83	52	11	79	74
의성군	27	189	60	24	203	63	28	177	48	30	171	60
청송군	8	112	56	13	96	57	12	81	32	9	107	37
영양군	8	44	16	9	50	16	8	39	20	10	40	20
영덕군	27	157	96	17	170	94	15	172	62	10	145	70
청도군	17	136	62	19	121	64	19	189	61	13	172	72
고령군	20	118	58	15	144	49	18	157	50	11	158	47
성주군	22	197	108	25	204	68	21	193	99	18	208	90
칠곡군	45	452	359	34	437	276	51	379	273	35	349	280
예천군	14	139	48	17	136	29	14	144	38	21	148	35
봉화군	20	156	62	11	147	48	7	134	29	15	129	81
울진군	22	155	117	23	165	60	12	139	81	17	151	82
포항시	68	1,300	1,824	57	1,211	1,876	80	1,250	1,913	79	1,104	1,776
경주시	94	1,423	844	83	1,367	693	92	1,287	809	91	1,122	798
김천시	49	493	249	43	394	289	49	389	347	42	403	379
안동시	35	667	502	51	639	479	46	640	451	42	638	511
구미시	72	1,299	1,099	69	1,422	1,075	54	1,382	1,175	66	1,142	1,219
영주시	19	519	185	28	558	225	16	500	158	21	499	225
영천시	42	549	335	45	476	233	39	443	244	51	390	326
문경시	18	183	113	22	259	137	19	210	171	20	196	196
상주시	36	330	242	34	374	205	45	311	264	40	270	237
경산시	35	931	469	42	800	531	51	780	601	59	663	578

<표 3> 지역별 사회지표

구분	2004년			2005년			2006년			2007년		
	인구	자동차 등록대수	도로 연장	인구	자동차 등록대수	도로 연장	인구	자동차 등록대수	도로 연장	인구	자동차 등록대수	도로 연장
군위군	29,157	11,959	257	27,984	12,308	257	26,687	12,260	306	26,176	12,420	266
의성군	66,507	27,835	524	64,930	27,948	523	62,947	28,028	591	61,871	28,332	519
청송군	30,197	10,919	353	29,406	11,095	365	28,465	11,310	396	27,993	11,495	394
영양군	20,275	7,196	356	20,082	7,396	357	19,615	7,491	357	19,207	7,620	3,574
영덕군	45,959	15,978	384	45,182	16,309	398	46,460	16,927	463	43,726	17,070	396
청도군	48,043	19,049	363	47,099	19,681	363	46,144	20,316	402	45,984	21,045	380
고령군	35,389	17,171	210	35,143	17,649	212	34,777	18,224	236	36,047	18,625	212
성주군	48,408	26,257	331	47,682	26,764	331	46,358	27,402	340	46,634	27,896	330
칠곡군	109,416	45,802	333	110,388	48,152	335	111,390	49,622	370	117,234	52,409	331
예천군	52,311	19,996	397	51,200	20,508	383	50,218	21,098	398	49,213	21,598	384
봉화군	37,662	12,714	482	36,738	13,064	479	35,608	13,334	505	35,243	13,880	480
울진군	58,869	21,079	406	56,988	21,235	405	55,076	21,259	435	54,353	21,652	434
포항시	508,937	195,618	841	509,148	203,123	861	505,008	208,957	966	508,684	214,868	881
경주시	280,092	108,966	1,001	277,764	111,438	1,008	273,419	114,435	1,038	276,877	117,734	976
김천시	144,587	54,260	798	142,688	55,585	798	139,682	56,731	886	140,564	57,967	904
안동시	174,596	68,179	976	172,029	65,229	971	168,733	66,359	980	169,239	67,742	977
구미시	370,088	143,092	539	378,560	150,969	544	386,465	158,670	554	396,884	164,153	532
영주시	121,908	48,574	498	119,668	49,702	499	116,965	46,820	505	116,062	47,768	505
영천시	110,891	48,431	553	108,745	48,871	560	105,029	49,541	584	107,701	51,101	559
문경시	79,820	27,957	586	78,357	28,483	605	76,177	29,006	619	75,223	29,408	607
상주시	112,943	47,824	742	110,892	49,035	762	108,280	50,265	841	107,266	51,490	755
경산시	223,357	90,762	696	231,677	95,506	725	234,839	100,910	817	242,744	103,728	880

### III. 자료수집 및 분석

#### 1. 자료수집

##### 1) 교통사고 통계

경상북도내 행정관할구역 22개 지역(시: 10개, 군: 12개)에 대한 2004년에서 2007년까지 4년간 발생한 교통사고 건수에 대하여 사고 심각도별(사망, 중상, 경상 및 물피)로 조사한 결과는 <표 2>와 같다.

##### 2) 사회지표

교통사고에 영향을 미치는 여러 요인들 중 기초자치단체의 사회지표인 인구, 자동차 등록대수, 도로연장거리를 조사한 결과 <표 3>과 같다.

#### 2. 분석

##### 1) 대물피해환산법(EPDO)

본 연구에서는 사고유형별(사망사고, 부상사고, 물피 사고) 각 피해 종류에 대해 교통사고 건당 동일한 피해 정도를 적용하기 위해 사고심각도에 따른 가중치를 적용하여 하나의 피해단위로 나타내는 방법인 대물피해환산법(EPDO, Equivalent Property Damage Only)을 사용하였다. 식(1)을 이용하여 교통사고 심각도별로 사고건수를 보정하였으며 각 자치단체별 사고심각도는 <표 4>와 같다.

$$EPDO = 12 \times (Fatal) + 3 \times (Inj) + PDO \quad (1)$$

여기서,

EPDO : 대물피해환산법에 의한 가중치

Fatal : 사망사고 발생건수

Inj : 부상사고 발생건수

PDO : 대물사고 발생건수

##### 2) 교통사고 건수와 사회지표와의 상관분석

사회지표(인구, 자동차등록대수, 도로연장)와 교통사고 건수와의 상관관계를 규명하기 위해 2004년부터 2007년까지의 자료들을 대상으로 상관 분석한 결과 <표 5>와 같이 인구, 자동차 등록대수, 도로연장거리가 모두

<표 4> 대물피해환산법(EPDO)에 의한 사고건수

구분	EPDO(건)			
	2004	2005	2006	2007
군위군	490	384	433	443
의성군	951	960	915	933
청송군	488	501	419	466
영양군	244	274	233	260
영덕군	891	808	758	625
청도군	674	655	856	744
고령군	652	661	737	653
성주군	963	980	930	930
칠곡군	2255	1995	2022	1747
예천군	633	641	638	731
봉화군	770	621	515	648
울진군	846	831	642	739
포항시	6540	6193	6623	6036
경주시	6241	5790	5774	5256
김천시	2316	1987	2102	2092
안동시	2923	3008	2923	2929
구미시	5860	6169	5969	5437
영주시	1970	2235	1850	1974
영천시	2486	2201	2041	2108
문경시	878	1178	1029	1024
상주시	1664	1735	1737	1527
경산시	3682	3435	3553	3275

<표 5> 교통사고 건수와 변수들간의 상관관계

		전체사고(EPDO)
		인구 (명)
	유의확률 (양쪽)	.000
	N	88
자동차 (대)	Pearson 상관계수	.962**
	유의확률 (양쪽)	.000
	N	88
도로연장 (km)	Pearson 상관계수	.319**
	유의확률 (양쪽)	.002
	N	88

\*\* 상관계수는 0.01 수준(양쪽)에서 유의.

유의수준 0.01(양측)에서 교통사고 건수와 상관관계가 있는 것으로 분석되었다.

### IV. 교통사고 예측모형

본 연구에서는 각 지역별 교통사고 예상 사고건수를 추정하기 위해 EB Method를 적용하였고 다음과 같이 4가지의 단계로 수행하였다.

- ① 안전성능함수(SPF)
- ② 과분산 파라메타( $\phi$ )
- ③ 가중치( $\alpha$ )
- ④ 예상 사고건수( $\pi$ )

1. 안전성능함수 (SPF - Safety Performance Function)

각 지역별 안전성능함수 모형 구축에 앞서 유사성이 높은 집단인 유사지역을 선정한 후, 선정된 유사지역을 대상으로 각 지역별 SPF모형을 산정함으로서 모형의 신뢰성을 높이기 위해 군집분석을 실시하였다.

군집분석은 대상들간의 유사성을 측정하여 유사성이 높은 대상들을 한 집단으로 묶어줌으로서 집단별 특성 등을 반영할 수 있게 해주는 분석방법이다. 즉, 특정자료가 얼마나 비슷한 값을 가지는 지를 거리로 환산하여 거리가 가까운 대상들을 동일한 집단으로 묶는다.

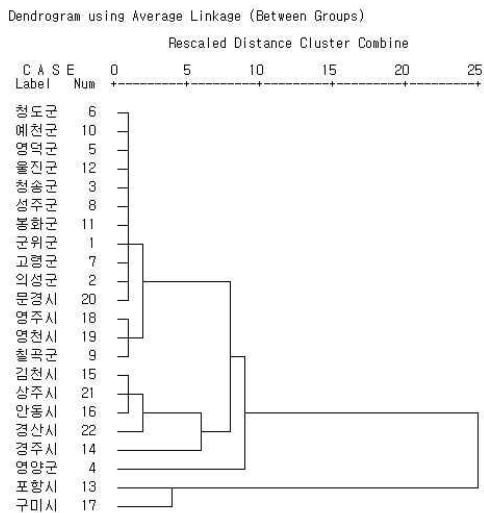
또한, 설명변수들이 각기 다른 측정단위를 가지고 있으므로 동일한 중요도를 갖게 하기 위하여 각 변수들의 표준화(Z 점수)를 통해 유사지역을 선정하였다.

분석결과 <표 6>과 같이 대상지역에 대해 제곱 유클리디안 거리가 1미만인 지역을 유사지역으로 선정하였다.

영양군, 포항시, 경주시, 안동시, 구미시, 경산시는 0~1개의 유사지역이 존재하는 것으로 분석되어 연구대상에서 제외하였다.

<그림 3>은 군집분석 결과를 덴드로그램(Dendrogram)으로 나타낸 것이다.

SPF모형은 지역별로 교통사고에 영향을 미치는 요인을 고려하여 사고건수를 예측하는 수학적 모형이다. 즉, 특별한 조치를 취하기 전에 일정기간 동안 발생한 교통사고자료를 이용하여 결정되며, 다른 조건이 일정하다면 샘플의 수가 많을수록 예측 모형의 정확도는 높아지게



<그림 3> 덴드로그램(Dendrogram)

<표 6> 지역별 유사지역

지역	유사 지역(제곱 유클리디안 거리)
군위군	청송군(.160), 영덕군(.337), 청도군(.233), 고령군(.076), 성주군(.241), 예천군(.281), 봉화군(.683), 울진군(.433),
의성군	청송군(.634), 영덕군(.322), 청도군(.436), 성주군(.634), 예천군(.374), 봉화군(.200), 울진군(.236), 영주시(.699), 영천시(.779), 문경시(.074)
청송군	군위군(.160), 의성군(.634), 영덕군(.071), 청도군(.071), 고령군(.400), 성주군(.215), 예천군(.079), 봉화군(.188), 울진군(.139)
영덕군	군위군(.337), 의성군(.322), 청송군(.071), 청도군(.021), 고령군(.548), 성주군(.139), 예천군(.018), 봉화군(.099), 울진군(.018), 문경시(.684)
청도군	군위군(.217), 의성군(.398), 청송군(.071), 영덕군(.020), 고령군(.342), 성주군(.059), 예천군(.005), 봉화군(.183), 울진군(.031), 문경시(.782)
고령군	군위군(.076), 청송군(.400), 영덕군(.548), 청도군(.380), 성주군(.256), 예천군(.451), 울진군(.628),
성주군	군위군(.241), 의성군(.634), 청송군(.215), 영덕군(.139), 청도군(.062), 고령군(.256), 칠곡군(.778), 예천군(.089), 봉화군(.452), 울진군(.138)
칠곡군	성주군(.778), 영주시(.371), 영천시(.724)
예천군	군위군(.281), 의성군(.374), 청송군(.079), 영덕군(.018), 청도군(.005), 고령군(.451), 성주군(.089), 봉화군(.169), 울진군(.018), 문경시(.767)
봉화군	군위군(.683), 의성군(.200), 청송군(.188), 영덕군(.099), 청도군(.200), 성주군(.452), 예천군(.169), 울진군(.119), 문경시(.443)
울진군	군위군(.433), 의성군(.236), 청송군(.139), 영덕군(.018), 청도군(.033), 고령군(.628), 성주군(.138), 예천군(.018), 봉화군(.119), 문경시(.559)
김천시	안동시(.538), 상주시(.217)
영주시	의성군(.699), 칠곡군(.371), 영천시(.075), 문경시(.686)
영천시	의성군(.779), 칠곡군(.724), 영주시(.075), 문경시(.657), 상주시(.728)
문경시	의성군(.074), 영덕군(.684), 청도군(.858), 예천군(.767), 봉화군(.443), 울진군(.559), 영주시(.686), 영천시(.657), 상주시(.777)
상주시	김천시(.217), 영천시(.728), 문경시(.777)

된다. 교통사고에 영향을 미치는 요인과 사고기록을 이용하여 식(2)와 같이 다중선형회귀 방정식으로 SPF를 설정하였다. 분석결과 다중공선성이 존재하는 지역에 대해서는 인구를 제외한 자동차대수와 도로연장길이를 통해 선형식을 도출하였다. 각 지역별 SPF를 분석한 결과 <표 7>과 같다.

<표 7> SPF모형 구축 (종속변수 =  $SPF_i$ )

지역	SPF 모형				결정계수 ( $R^2$ )	분산 ( $\Phi$ )
	상수	인구	자동차	도로연장		
군위군	73.375	0.001	0.025	0.293	0.681	0.40
의성군	-126.935	-	0.043	0.070	0.918	0.92
청송군	143.161	0.002	0.023	0.145	0.738	0.37
영덕군	134.360	0.002	0.023	0.173	0.781	0.34
청도군	134.303	0.002	0.023	0.173	0.781	0.34
고령군	105.211	0.003	0.025	0.014	0.728	0.39
성주군	21.803	-	0.038	-0.058	0.884	1.68
칠곡군	-487.201	-	0.044	0.798	0.864	1.40
예천군	134.360	0.002	0.023	0.173	0.781	0.34
봉화군	99.958	0.001	0.023	0.284	0.786	0.21
울진군	134.338	0.002	0.023	0.173	0.781	0.34
김천시	-1896.390	-	0.067	0.328	0.911	0.16
영주시	-815.483	-	0.054	0.494	0.893	2.72
영천시	187.433	-	0.045	-0.816	0.832	2.49
문경시	267.724	-	0.045	-0.912	0.906	0.87
상주시	586.485	0.013	0.031	-2.360	0.907	0.80

$$SPF_i = \beta_0 + \beta_1 \chi_{i1} + \beta_2 \chi_{i2} + \dots + \beta_{p-1} \chi_{i,p-1} + \epsilon_i \quad (2)$$

여기서,

$SPF_i$  :  $i$  지역의 SPF

$\chi_{i1} - \chi_{i,p-1}$  : 추정변수

$\beta_0 - \beta_{p-1}$  : 파라메타

$\epsilon_i$  : 무작위 에러

## 2. 과분산 파라메타( $\phi$ )

<표 7>의 결과물을 그대로 사용할 수 없다. 그 이유는 과분산 파라메타 때문이다. 즉, 지역별로 SPF를 결정하기 위해 주어진 기간 내에 발생한 교통사고건수의 근원적인 분포에 대한 가정이 필요하다. 사고건수에 대한 근원적인 확률분포를 가정하는 것이 필요한데 지금까지는 사고건수가 포아송분포를 따르는 것으로 가정되었다. 그러나, 포아송분포를 이용하여 예측된 사고건수와 실제 사고건수가 차이가 있고 사고에 있어서는 많은 경우 분산이 평균보다 크게 나타나는 과분산(Overdispersion) 현상이 발생하는 것으로 선행연구에서 지적하고 있다.<sup>1)</sup>

최근 연구에서 사고건수의 분포는 음이항(negative binomial) 분포가 더 적합한 것으로 조사되었는데 실제 사고건수와 근원적인 확률분포가 일치하려면 과분산

파라메타를 적용해야 한다.<sup>2)</sup> 과분산 파라메타는 식(3)을 이용하여 계산할 수 있다.

$$\phi_i = \Phi_i \cdot SPF_i^r \quad (3)$$

여기서,

$\phi_i$  : 과분산 파라메타

$\Phi_i$  : 분산

$SPF_i^r$  : 유사지역의 평균교통사고건수

식(3)에서  $\gamma$ 값이 0이면 음이항분포가 되고  $\gamma$ 값이 0보다 크면 SPF값이 증가할 때 감마분포의 변화량은 감소하게 된다. 여러 선행 연구에서 사전-사후 결과를 비교하기 위해 과분산 파라메타를 사용했고 보다 정확한 결과를 도출하였다. 지방부 고속도로 사고모형을 예측하면서  $\gamma$ 값을 1로 했을 때 비교적 정확했다고 보고했으며<sup>3)</sup>, 다른 연구에서도  $\gamma$ 값을 1로 하는 것을 추천하고 있어 본 연구에서도  $\gamma$ 값을 1로 가정하였다.

## 3. 가중치( $\alpha$ )

지역별 교통특성이 다르기 때문에 지역마다 과분산 파라메타는 다르게 적용되어야 하며, 식(4)와 같이 과분산 파라메타와 유사지역의 평균교통사고건수를 이용하여 가중치를 산정한다.

$$\alpha_i = \frac{1}{1 + SPF_i / \phi_i} \quad (4)$$

여기서,

$\alpha_i$  :  $i$  지역의 가중치

$SPF_i$  :  $i$  지역과 유사한 지역의 평균교통사고건수

$\phi_i$  :  $i$  지역의 과분산 파라메타

## 4. 예상 사고건수( $\pi$ )

$i$  지역의 예상 사고건수( $\pi$ )는 유사지역의 평균교통사고건수( $SPF_i$ )와  $i$  지역의 실제 사고건수( $\lambda_i$ )의 가중치에 의해서 산정된다.

1) Hauer (1996), Hauer and Persaud (1984), Hutchinson and Mayne (1997)

2) Power, M. and J. Carson (2004)

3) Miaou, S. and H. Lum (1993)



<표 8> 예상 사고건수

지역	SPF Result	과분산 파라메타( $\phi$ )	가중치 ( $\alpha$ )	예상 사고 건수( $\pi$ )	실제 사고 건수( $\lambda$ )
군위군	487.99	197.62	0.10	443	443
의성군	1127.67	1034.40	0.23	983	933
청송군	520.66	191.90	0.09	473	466
영덕군	682.93	232.83	0.09	763	625
청도군	776.05	264.54	0.09	736	744
고령군	637.16	250.57	0.10	672	653
성주군	1062.71	1784.99	0.42	998	930
칠곡군	2082.93	2919.88	0.35	2032	1747
예천군	795.97	271.36	0.09	672	731
봉화군	590.76	125.17	0.05	636	648
울진군	816.12	278.21	0.09	769	739
김천시	2283.91	369.40	0.04	2131	2092
영주시	2013.46	5479.66	0.68	2011	1974
영천시	2030.83	5063.42	0.62	2098	2108
문경시	1037.50	905.73	0.22	1029	1024
상주시	1795.33	1432.61	0.20	1692	1527

$i$  지역에 대하여 예상 사고건수는 식(5)를 이용하여 계산할 수 있으며 그 결과는 <표 8>과 같이 김천시(2131건), 영천시(2098건), 칠곡군(2032건)순으로 가장 높은 교통사고 건수가 예상되는 반면, 군위군(443건)이 가장 낮은 사고건수가 발생할 것으로 분석되었다.

$$\pi_i = (\alpha_i) \cdot (SPF_i) + (1 - \alpha_i)(\lambda_i) \quad (5)$$

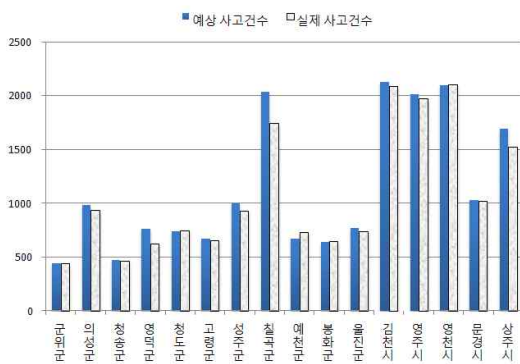
여기서,

$\pi$  :  $i$  지역의 예상 사고건수

$\alpha_i$  : 가중치

$SPF_i$  : 유사지역의 평균교통사고건수

$\lambda_i$  :  $i$  지역의 실제 사고건수



<그림 4> 실제 사고건수와 EB Method를 이용한 예상 사고건수 비교

각 지역별 실제 발생한 사고건수와 EB Method를 이용하여 예측된 사고건수와 비교하면 <그림 4>와 같다.

칠곡군의 경우 예상 사고건수 2032건, 실제 사고건수 1747건으로 다소 차이가 있는 반면, 이외의 지역에서는 상당히 유사한 것으로 분석되었다.

### V. 결론 및 향후과제

본 연구는 지역별 예상 사고건수를 예측하기 위해 기존의 사고예측모형과는 달리 EB Method를 적용하였다. 사고유형별 사고 심각도를 반영하기 위해 대물피해 환산법을 적용하였고, 기존의 EB Method 경우 유사지역 선정시 명확한 통계적 분석방법이 제시되지 않았으나, 본 연구에서는 군집분석을 토대로 유사지역을 선정함으로써 기존의 EB Method의 문제점을 보완하여 다음과 같은 방법으로 적용하였다.

첫 번째, 연구 대상 지역에 대해 군집분석을 통하여 각 지역별 유사지역을 선정하였고, 다중회귀분석을 통하여 지역별 안전성능함수(SPF)모형을 산정하였다.

두 번째, 지역별 교통특성이 다르기 때문에 지역마다 과분산 파라메타는 다르게 적용되어야 하는데 각 지역별 과분산 파라메타를 보정하여야 하므로 가중치를 적용하였다.

마지막으로 유사지역의 평균교통사고건수( $SPF_i$ )와 실제 사고건수( $\lambda_i$ )에 가중치를 적용하여 예상 사고건수를 예측하였다.

분석 결과 김천시, 영천시, 칠곡군 순으로 가장 높은 사고건수가 예상되는 반면, 군위군이 가장 낮은 사고건수가 발생할 것으로 분석되었다.

칠곡군의 경우 예상 사고건수 2032건, 실제 사고건수 1747건으로 다소 차이가 있는 반면, 이외의 지역에서는 상당히 유사한 것으로 분석되었다.

본 연구에서는 사회지표(인구, 자동차 등록대수, 도로연장)를 변수로 하여 장래의 교통사고 건수를 예측하여 지역별로 전반적인 사고건수를 예측하였으나, 향후에는 국지적인 지역에 한정하여 교통사고에 영향을 미치는 다양하고 정확한 인자를 이용하여 미시적인 분석이 필요할 것으로 판단된다.

알림 : 본 논문은 대한교통학회 제59회 학술발표회(2009. 10.24)에서 발표된 내용을 수정·보완하여 작성된 것입니다.

## 참고문헌

1. 강승규 (2006), “교통안전공학”. 도서출판 대가.
2. 건설교통부-경찰청 (2002), “사고 잦은 곳 개선사업 업무편람”, 행정간행물.
3. 김장욱·남궁문·김정현·이수범 (2006), “과지 및 신경망 이론을 이용한 교통사고예측모형 개발에 관한 연구”, 대한교통학회지, 제24권 제7호, 대한교통학회, pp.81~90.
4. 도로교통안전관리공단 (2004~2007), “전국 시도 및 도로별 사고 잦은 곳 현황”.
5. 박규영·김태희·김성욱·이수범 (2006), “도로안전개선사업의 공간적 범위에 따른 도로안전시설물의 효과도 비교”, 대한교통학회지, 제24권 제4호, 대한교통학회, pp.31~42.
6. 박민호·박규영·장일준·이수범 (2006), “중양분리대 설치에 따른 사고전환효과 분석”, 대한교통학회지, 제24권 제2호, 대한교통학회, pp.113~124.
7. 성낙문 (2003), “교통사고예측모델을 이용한 도로의 안전도 평가방법 연구”, 교통개발연구원 연구보고서.
8. 이수범·박귀영 (2000), “도로안전시설의 교통사고 감소효과 측정에 관한 연구”, 대한토목학회 논문집 제20권 제2-D호 pp.139~147
9. 정성봉·황보희·성낙문·이선하 (2009), “EB기법을 이용한 사고잦은 곳 개선사업 우선순위 선정기법 개발”, 대한교통학회지, 제27권 제3호, 대한교통학회, pp.81~90.
10. Berry, D.A (1996), “Statistics : A Bayesian Method,” Cambridge University Press, Cambridge, U.K.
11. Cameron, A.C., and P.K. Trivedi (1998), “Regression Analysis of Count Data,” Cambridge University Press, Cambridge, U.K.
12. Carlin, B. and T. Louis. (2001), Bayes and Empirical Bayes Methods for Data Analysis. 2nd ed. Chapman and Hall.
13. Carriquiry, A.L. (2004), “From empirical bayes to full bayes : Methods for analyzing traffic safety data.”
14. Clayton and Kaldor (1987), “Empirical Bayes Estimates of Age-Standardized Relative Risk for use in Disease Mapping,” Biometrics, Vol. 43, No. 3, pp.671~681.
15. Congdon, C. (2001), Bayesian Statistical Modelling. John Wiley and Sons.
16. Hauer, E. (1997), “Observational before-after studies on road safety”, Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety.
17. Meza, J. (2003), “Empirical Bayes Estimation Smoothing of Relative Risk in Disease Mapping,” Journal of Statistical Planning and Inference, Vol. 112, pp.43~62.
18. Miaou, S. and H. Lum (1993), “Modeling Vehicle Accidents and Highway Geometric Design Relationships,” Accident Analysis and Prevention, Vol.25, pp.689~709.
19. Ogden, K.W. (1996), Safer Roads : A Guide to Road Safety Engineering
20. Power, M. and J. Carson (2004), “Before-After Crash Analysis: A Primer for Using the Empirical Bayes Method Tutorial,” Montana Department of Transportation.

✉ 주 작성자 : 강현건

✉ 교신저자 : 강승규

✉ 논문투고일 : 2008. 12. 1

✉ 논문심사일 : 2009. 1. 12 (1차)

2009. 5. 25 (2차)

2009. 7. 30 (3차)

2009. 8. 25 (4차)

2009. 9. 2 (5차)

✉ 심사판정일 : 2009. 9. 2

✉ 반론접수기한 : 2010. 2. 28

✉ 3인 익명 심사필

✉ 1인 abstract 교정필