

교통사고 통계자료 분석 기법연구

손 소 영*

최근 증가되고 있는 우리나라의 교통사고는 많은 문제를 야기하고 있다. 교통사고 감소를 위한 대책을 세우기 위해서는 증가된 교통량과 더불어 관련된 교통사고 발생원인 및 특성을 규명하는 일이 우선되어야 한다. 이를 위해 정확한 통계자료를 바탕으로 한 사고 모형 추정이 필수적인 점을 감안할 때 자료수집과정에 대한 연구와 통계분석의 중요성이 강조된다. 본 논문에서는 교통사고분석에 이용되어온 통계분석기법을 관련응용사례와 더불어 소개하고 우리나라에서 수집되고있는 교통통계자료의 효과적인 활용을 위해 앞으로 연구되어야 할 부분에 대한 방향성 제시를 하고자 한다.

< 차 례 >

1. 서론
2. 교통사고 통계자료
3. 교통사고통계분석문헌의 고찰
4. 향후 연구를 위한 제안
5. 참고문헌

1. 서론

우리 나라의 산업발전과 더불어 증가된 교통량은 환경문제와 함께 많은 문제를 야기하고 있다. 지난 92년부터 조금씩 줄었던 교통사고 사망자가 95년 이후 다시 크게 늘어나기 시작해 96년 한해 26만5천52건의 교통사고가 발생했고 그 중 1만2천6백53명은 사망자로 전년대비 22.6% 증가를 기록하고 있다. 이처럼 높은 교통사고 증가율은 강한 사회적 위감을 일으키고 있으며, 막대한 규모의 사회·경제적 손실을 초래하고 있다.

교통사고 감소를 위한 대책을 세우기 위해서는 증가된 교통량과 더불어 관련된 교통사고 발생원인 및 특성을 규명하는 일이 우선되어야 한다.

이를 위해 교통사고 발생과 관련된 도로시설구조, 교통이용실태, 기후조건, 운전자특성, 차량 특성등 다양한 자료의 과학적이며 종합적인 조사 분석이 필요하다. 교통사고의 원인별 규명이나, 사전방지를 위해서는 정확한 사고통계자료를 바탕으로한 사고모형 추정이 필수적인 점을 감안할 때 자료수집과정에 대한 연구와 통계분석의 중요성이 강조된다.

본 논문에서는 교통사고분석에 이용되어온 통계분석기법을 관련응용사례와 더불어 소개하고 우리 나라에서 수집되고 있는 교통통계자료의 효과적인 활용을 위해 앞으로 연구되어야 할 부문에 대한 방향성 제시를 하고자 한다. 제2장에서는 우리 나라 교통사고 통계분석의 근간이되는 통계원표에 대한 소개를 하고자한다. 제3장은 교통사고 통계분석 문헌 고찰에 할애하며, 제4장에는 효과적인 교통사고 분석을 위한 통계원표 활용면에 대해 논하였다.

2. 교통사고 통계자료

우리 나라에서 교통사고란 도로교통법 제2조에 규정하는 도로에서 차의 교통으로 인하여 발생한 인적, 물적 피해가 따르는 사고를 말한다. 교통사고가 발생하면 접수를 받은 담당 경찰관이 신속히 현장에 출동하여 사상

자의 구호 및 사고현장의 통제, 사고현장에 대한 기초조사, 교통소통을 한다. 이후 정확한 원인조사와 민·형사상의 책임을 규명하기 위하여 사고조사보고서를 작성한다. 이를 바탕으로 하여 장래의 교통사고 방지대책을 수립하기 위하여 사고관련 자료를 모으기 위한 통계원표를 작성한다. 통계원표는 본표와 보충표로 구성되어 있으며 본표에는 교통사고의 발생일시, 도로환경, 교통사고형태등 사고내용과 1, 2 당사자에 관련된 사항을 기재하고, 보충표에는 본표에 기재하지 못한 교통사고관련 제3당사자 이하의 사망하거나 부상한 사람에 관한 사항을 1인당 1매씩 작성토록 하고 있다. 본표는 교통사고 발생시 한 건당 1매씩 작성하고 작성범위는 물피사고시 조사항목의 1~50까지, 부상사고시 1~54까지, 사망사고시 1~69까지 각각 작성하여야 하며 고속도로사고시 조사항목의 70~79까지 추가로 작성하여야 한다. 우리 나라 통계원표의 총 79항목은 대부분 범주형 자료로 일본의 통계원표와 유사한 형식을 갖추고 있다.(<표 1> 참고). 수작업된 통계원표는 경찰서 관리시스템으로 온라인 입력이 되어 중앙전산처리를 위해 각 지방경찰청 관할 면허시험장으로 전송된다. 면허시험장에서는 이를 에러 체크 프로그램을 이용해 에러를 수정하고 본청 면허전산실로 보낸다.

통계원표에 나타나 있듯이 사고의 종류도 그 상해정도에 따라 사망, 부상, 경상, 부상신고로 나뉘며, 사고의 원인도, 인적요인, 차량원인, 도로 환경적 원인 등으로 구분되어 있어 분석의 목적에 따라 다양한 사고의 구분을 지을 수 있다. 주의할 점은 사망에 대한 정의로 이것은 국가별로 차이가 있어 국가별 비교시 보정을 필요로 한다. 예를 들어, 우리 나라의 경우, 교통사고 이후 3일 이내 사망만 교통사고로 인한 사망으로 취급하고 있는데 반해, 미국의 경우 30일 기준이 적용된다.

교통사고는 차량, 운전자, 보행자, 도로시설 및 환경조건 등들의 상호 작용에 의해 발생된다고 볼 수 있으며 통계원표상에 나타나 이 요소들과 사고와의 관계를 논리적으로 추적할 수 있는 종합적이고 체계적인 분석이 뒤따라야 한다. 우리 나라의 경우 교통사고를 각종 관련 인자별로 구분 집계한 [교통사고통계],[교통사고통계분석],[도로교통안전백서]등이 있으며, 미국에서는 사망사고자료만을 모은 FARS (Fatality Analysis Reporting System)와

일반사고에서 표본추출한 GES(General Estimates System)의 사고자료를 바탕으로 집계 분석된 [Traffic Safety Facts]가 매년 출간되고 있다. 이들 책자에 분석된 평면적인 집계 이외에도 모형을 근거로 한 교통사고분석이 활발하며 다음 장에는 그 현황에 대한 고찰을 하고자 한다.

3. 교통사고통계분석문헌의 고찰

교통사고 통계분석 문헌의 대주제는 거시적인 사고량 추이 예측 및 미시적인 다발지점에 대한 사고원인 분석에 있다고 볼 수 있다. 분석목적에 따라 교통사고에 대한 분류를 하여 사고종류에 대한 다양한 변수들이 정의 될 수 있으며 사고 유발 환경 및 원인과 관련된 인자들과 더불어 다변량을 이루게 된다. 그러므로 교통사고 자료 분석에 있어 대부분의 경우 한꺼번에 여러 개의 요인을 다루어 다변량 분석이 많이 쓰이고 있다. 다변량분석의 다양한 용도와 분석기법들을 간단히 정리하면 다음과 같다(김규형,1993)

<표 1> 한국과 일본의 통계 원표 비교

	일본 통계 원표의 항목	우리 나라 통계 원표 항목
1. 서 식	자료 구분 도도촌현 경찰서 등급 코드 본표번호	1. 자료 구분 2. 발생지 경찰서 코드 3. 본표 번호
2. 사고내용 에 관한 내용	사고내용 사망자수 중상자수 경상자수 승차인원	4. 사고내용 5. 사망자수 6. 경상자수 7. 경상자수 8. 부상 신고자 수

	일본 통계 원표의 항목	우리 나라 통계 원표 항목
3. 발생장소 및 시간 에 관한 내용	노선코드 지점코드 시구정촌코드 발생일시 주야 요일 천후 지형 노면상태 도로형상 교차로형상	17. 노선 코드 15. 발생 위치 좌표 13. 발생지 시군 코드 10. 발생일시 12. 주야 11. 요일 22. 기상 26. 지형 29. 노면상태 24. 도로형태 해당 사항 없음
4. 사고지점 의 사세한 설명에 관한 내용	신호등 도로선형 차도폭 충돌지점 중앙분리대시설등급 보차도구분	27. 신호기 28. 도로 선형 25. 차도폭 해당 사항 없음 55. 중앙 분리대 시설 등 56. 보차도 구분
5. 특수사고 의 내용	특수사고	23. 특수사고
6. 사고유형 및 사고자에 대한정보	사고유형 성별 연령 국적코드 주거지코드 산업코드 직업코드	35. 사고 유형 20. 성별 19. 연령 해당 사항 없음 14. 주거지 코드 해당 사항 없음 18. 직업 코드

	일본 통계원표의 항목	우리 나라 통계 원표 항목
7. 사고차량 에 관한 정보	차량번호 면허증번호 운전자격 사고 차량의 운전 면허 경과 년수 당사자종별 차량형상 트레일러유무 용도별 적재상황 오토매틱차 핸들위치	해당 사항 없음 해당 사항 없음 45. 운전 자격 48. 사고 차종의 면허 경과 년수 31. 당사자 종별 32. 차체 형상 해당 사항 없음 33. 용도별 30. 화물 적재 상황 58. 자동 변속기 차 해당 사항 없음
8. 사고시 차와보행 자의 행동 유형	행동유형 통행목적 통행형태	65. 행동 유형 37. 통행 목적 66. 통행 형태
9. 관리 및 규제	선입사업소 초과금지규제 주정차금지규제 속도지정규제	38. 특별 관리자 유무 해당 사항 없음 해당 사항 없음 57. 속도 규제
10. 충돌당시의 상황	위험인지 속도 당사자의 진행 방향 차량의 충돌부위	36. 사고직전속도 해당 사항 없음 해당 사항 없음
11. 차량 훼손 정도및 음주 운전	차량의 훼손정도 음주운전	50. 차량의 훼손 정도 40. 음주 운전
12. 법령위반 및 원인 코드	법령위반코드 원인코드	41. 법규 위반 42. 원인 구분

	일본통계원표의 항목	우리 나라 통계 원표 항목
13. 사고자의 상해에 관한 내용	신체방호 신체손상정도 신체손상주부위 손상주부위의형태 인체가해, 부위 자택으로부터의거리	49. 보호 장구 51. 신체 상해 정도 52. 신체 상해 주부위 53. 신체 상해 부위 상태 54. 신체 가해 부위 69. 자택으로부터의 거리

우리나라의 통계 원표에만 있는 항목	일본의 통계 원표에만 있는 항목
8. 부상 신고자수 9. 물적 피해액 15. 발생 위치 16. 도로 구분 21. 학력 34. 총 배기량 코드 39. 난폭 운전 43. 명령요인 44. 면허의 종류 46. 신체 장애인 면허 조건(차) 47. 신체 장애인 면허 조건(사람) 59. 제작 회사 코드 60. 연식 61. 보험 가입 상황 62. 차량 검사 63. 최대 적재량 64. 화물차 적재 상황 67. 직업 운전자별	10 지점코드 22 교차로 형상 26 충돌지점 33 국적코드 35 산업코드 37 차량 번호 38 면허증 번호 43 트레일러 유무 47 핸들의 위치 52 초과 금지 규제 53 주정차 금지 규제 56 당사자 진행 방향 57 차량의 충돌 부위

- 다변량 모집단의 모수에 대한 가설 검증(다변량 분산 분석, 다변량 공분산 분석)
- 변수들 사이의 종속성 분석(회귀 분석, 상관 분석, 정준 상관 분석(canonical correlation analysis))

- 변수 사이의 상호 관련성에 기초한 자료의 축소 또는 구조의 단순화(주성분 분석(principal component analysis), 요인 분석(factor analysis))
- 집단 사이의 상호 관련성에 기초한 분류 또는 집단화(판별 분석(discriminant analysis), 군집 분석(cluster analysis), 다차원 척도 분석(multi-dimensional scaling))

이러한 여러 가지 분석 기법들이 교통사고문헌에 적용된 예를 고찰하고자하며 먼저 수많은 교통사고 관련자료의 단순화를 위한 주성분분석과 요인분석 적용예를 소개한다.

▶ 주성분분석(Principal Component Analysis)

주성분 분석이란 자료를 축소 또는 해석하기 위한 분석기법으로써 원래의 변수의 선형 결합으로 자료의 공분산 구조를 설명하는 기법이다. 주성분 분석의 교통사고분석 응용 예는 다음과 같다.

Oppe(1992)는 층화추출법(stratification scheme)에 따라 선정한 네덜란드의 32개 교차로에서 일정기간동안 발생한 312건의 교통사고자료를 주성분분석에 응용하였다. 분석에 이용한 변수는 교차로의 기하학적인 구조(교차로 크기, 각도, 안과 밖의 교차로 폭, 교차로 안쪽 지름 등)와 평균 24시간 동안의 차량 통행량(이륜차 유입량과 이출량, 차량 유입량과 이출량, 총 유입량과 이출량 등)에 따른 20개의 설명 변수들과 사고의 종류(진입/회전 사고, 접근 사고, 단독 사고, 기타사고, 보행자 사고)들이다. 분석결과 이들 변수들을 가장 잘 대표하는 선형결합의 제1성분과 제2성분을 구하여 이들 두 주성분을 축으로 원래 변수들의 관계 구조를 밝혔다. Oppe(1992) 분석의 특징은 주성분분석을 함에 있어 사고의 종류등 명의 척도(nominal scale)자료를 scale해서 쓰는 정성적인 주성분분석(Princa-1s)을 한다는 점이다.

이광희 (1994)는 부산시내에 있는 사고 다발 지점의 45개 도로환경별 특성을 바탕으로 7개의 주성분을 도출했다. 각사고 다발지점별로 7개 주성분별 값들을 구

하여 사고지역을 유형화하였다.

▶ 요인분석(Factor Analysis)

요인분석은 p 개의 관측 가능한 양적 변수들 사이에서 공분산 관계를 설명할 수 있는 $m(m < p)$ 개의 인자라고 불리는 가설적인 변수를 찾는 다변량 분석법이다. 즉, 변수가 여러 개 있는 경우에 비슷한 특성을 지닌 변수들끼리 모아 몇 개의 집단으로 나눈 후 각 집단으로 대표할 수 있는 새로운 대표군인 인자를 찾는 것이다. 원래의 변수에 의한 변동의 대부분을 훨씬 적은 수의 인자들을 이용해서 설명할 수 있다면 자료의 해석이 용이해지고 가능해 진다.

신용균(1992)은 사망사고를 줄이기 위한 연구의 일환으로 사망사고 원인의 우선 순위를 알아보기 위해 서울의 세 지역에서 발생한 136건의 사망사고 중 도로 사고자 특히 사고 가해 운전자의 문제 행동 요인을 다음과 같은 39개의 질적 분석 단위로 측정하였다.

야간 교통 특성 무시, 과속, 심야, 좌우 탐색 불량, 전방 주시 태만, 좌우 후방만 주의, 발견 지연, 회피 조작 늦음, 보행자 예측 못함, 대향차 예측 못함, 갑자기 출현, 매우 당황, 회피 조작 오류, 기대 불량, 의사 소통 불일치, 운전 기술 과신, 운전 경험 과신, 과로, 장시간 운전, 신호 무시, 성급한 운전, 우선 위치 선택 불량, 동승자와 잡담, 잠 생각, 다른 행동, 이륜차 특성 이해 부족, 급제동, 혼잡, 일시 정지 위반, 위험인식 저조, 음주 운전, 중앙선 침범, 통행량 한산, 통행량 많음, 시설물 장애, 시인거리, 불량, 빗길 미끄러짐, 귀가중, 경쟁적 주행

이들 39개 변수들의 상호 상관성을 고려하여 몇 개의 인자로 사망사고의 특성을 규명하기 위하여 요인 분석한 결과 사망사고의 위험 요인의 순서를 심야의 과속 요인, 도로 등의 불량 요인, 관찰 요인 등으로 제시하였다.

윤문호(1989)는 인자요소와 환경적 요소에 관한 영향 요소를 찾기 위해 1988년 서울시 북부 지역 32개 사고 다발 지점의 자료와 현장 조사를 바탕으로 요인분석을 하였다.

Kanellaidis(1995)는 207명의 그리스의 운전자에게 14개 도로의 환경에 대한 설문조사를 통해 그들이 커브운전을 할 때 속도에 영향을 주는 변수들을 집계하여 요인분석을 한 결과 커브에서의 속도 선택은 4가지의 도로 환경요인(opposing traffic의 선징, cross-section 특성, alignment, signing)에 의해 설명되어질 수 있음을 보였다. 속도 규제를 위반한 운전자들과 위반하지 않는 운전자들의 반응 비교는, 후자는 주로 커브에서 신호등 요인에 의해 그들의 속도를 정함을 보여주고, 반면에 전자는 주로 도로 배치 요인에 의해 결정됨을 보여주었다.

주성분분석이나 요인분석을 통해 종속변수나 설명변수의 구분 없이 다량의 교통관계 변수들의 관계에 대한 파악을 한 후 교통사고량 또는 상해정도를 관련 변수로 모형화하기 위해 종종 쓰이는 회귀분석, 판별분석에 대해 알아보자.

▶ 회귀분석(Linear Regression)

회귀분석은 거시적인 교통사고량 추이 예측에 가장 많이 쓰인 기법으로 사고량과 설명변수 들간의 관계에 따라 크게 선형과 비선형으로 나뉠 수 있다. 교통사고량은 지역별(cross-sectional data), 혹은 시간별(time-series data)로 분석될 수 있으며 시간과 지역을 변화시킨 복합자료를 사용하기도 한다. 이러한 회귀분석의 교통사고자료분석 응용 예를 저자, 사용자료성격, 모형 등을 중심으로 정리하면 <표 2>와 같다.

일반적인 회귀분석을 통한 사고 모형자체가 갖는 한계성은 현 상태의 지속성이라는 가정에 근거하고 있다는 것이다. 따라서 이를 극복하기 위해서는 사고 모형에 정책변수를 포함시키는 노력이 필요하며, 아울러 교통사고 모형에 세분화된 설명변수들과 종속변수들을 포함한 개별모형(Disaggregate Models)의 개발도 매우 중요한 과제라고 판단된다.

▶ 시계열분석(Time-Series Model)

시계열 자료는 일반적으로 시간의 흐름에 따라 주기적으로 변하는 특성을 가지고 있으며, 동일한 계절(월, 분기, 년 등)의 시계열 관측값 사이에 존재하는 높은 상관관계는 계절요인을 가진 시계열의 특징이다. 시계열 자료가 계절성(seasonality)을 갖고 있으면, 이것은 non-stationary 계열을 의미한다. 이러한 형태의 시계열을 보정하기 위해서는 계절성을 제거하기 위한 차분방법을 사용하여 모형을 만든다. 이러한 차분과정을 모형화 시켜 만든 시계열 모형을 계절성을 가진 ARIMA(multiplicative seasonal autoregressive integrated moving average)모형이라고 한다.

이윤중(1991)은 교통사고 월별 자료를 Box-Jenkins시계열 모형에 적용하여 그 모형의 예측치를 분석하고 사고의 원인별로 파악함으로써 교통사고 발생의 모형을 제시했다.

▶ 판별분석(Discriminant Analysis, Logistic Regression)

이 분석은 여러 개의 모집단에서 추출한 표본들이 지니고 있는 관측값 벡터들을 이용하여 이 표본들이 어느 모집단에서 추출된 것인지를 판단하기 위하여 가장 잘 분류해 줄 수 있는 기준, 잘못 분류될 확률을 최소화하는 판별 함수를 찾는 방법이다. 판별분석은 관측값 벡터들이 다변량 정규분포를 따른다는 가정을 요하므로, 이러한 가정이 필요하지 않은 logistic regression이 같은 목적을 갖는 분석에 많이 적용되어 오고 있다.

오윤석 & 고양선(1992)는 대형 교통사고에 있어 교통사고의 요인들 중 인적 요인에 속하는 운저자의 법규 위반 유형별과 가해 운전자의 사망 여부를 인적, 차량적, 그리고 도로 환경적 요인들에 대한 판별 함수를 이용해 분석하였다.

< 표 2> 회귀분석을 이용한 교통사고 건수 추정모형

저 자	자 료	모 형
[Smeed, 1972]	유럽 20 개국의 1938년 교통사고 통계 자료	$\text{교통사고 사망자수} = 0.000111204$ $(\text{인구수})^{0.33797} (\text{자동차 대수})^{0.72387}$ $\text{교통사고 사망자수} = -22.132 + 0.067864$ $(\text{인구수}) + 0.798548(\text{자동차 대수})$
[임헌연 · 이일병, 1990]	1962-1989년까지의 부산 시내 교통 사 고 자료	$\text{교통사고 사망자수} = 17.415 \cdot \text{자동차 등록}$ $\text{대수}^{-0.438}$
[Trichopoulos (1975)]	유럽 17개국의 1970년 교통사고 통계 자료	$\text{사망자수} = 58.770 + 30.322 * \text{자동차 보유율}$ $(\text{대/인}) + 4.278 * \text{도로밀도}(\text{km/km}^2) - 0.107 * \text{대도시 거주 인구 비율}(\%) - 0.776 * \text{19세}$ $\text{이하의 청소년 인구 비율}(\%) - 2.874 * \text{65세}$ $\text{이상의 노년층 인구 비율} + 0.147 * \text{자동차용}$ 과 택시의 비율
[김홍상,1987]	공업국가 중심의 19개국의 교통사고 자료	$\text{교통사고 사망자수} = f(\text{인구수, 자동차 보}$ $\text{유수, 도로 총 연장 거리, 국민 총생산, 연}$ $\text{주행거리, 국토 면적, 이륜차를 제외한 차}$ $\text{량 보유 대수, 정부 예산 중 도로 교통 부}$ $\text{분이 차지하는 비율})$
[윤문호,1989]	서울시의 사고 다발 지점 32개소 의 사고자료	$\text{사고율} = f(\text{사고형태, 사고유형, 사고원}$ $\text{인, 재산피해, 경력, 학력, 도로노면, 지역}$ $\text{여건, 도로형태, 도로폭, 차선수, 안전시}$ $\text{설수})$
[Mak,Chiva-Cha vala, and Higler 1986]	도시고속도로의 사고 자료	$\text{사고건수} = f(\text{유형, 사고시간, 경사정도,$ $\text{차량종류, 피해정도, 날씨노면상태, 운전자}$ $\text{나이, 마약복용여부, 면허발급지})$
[박병호,1995]	1947-1965년까지의 통계자료에 근거해 서 시계열 분석, 양 변자연대수(Double log)를 이용	$\text{사망자수} = f(\text{가격, 소득, 추세경향, 알}$ $\text{콜, 속도, 청년인구, 자동차주행거리})$
[Zlatoper(1984) 모형]	1947년-1980년까지 의 통계자료에 근거 해서 시계열 분석, 양변자연대수(Doubl e log)를 이용	$\text{사망자수} = f(\text{가격, 소득, 추세경향, 속도,$ $\text{청년인구, 자동차 주행거리, 규정, 화물차})$

저 자	자 료	모 형
[Peltzman(1976) 모형]	1947년-1965년까지의 통계자료에 근거해서 시계열 분석, 양변자연대수(Double log)를 이용	사망자수 = f(가격, 소득, 추세경향, 알콜, 속도, 청년인구 그리고 자동차주행거리)
[Partyka(1984) 모형]	1947년-1982년까지의 통계자료에 근거해서 시계열 분석, 선형모형을 이용	사망지수 = f(실업자수, 취업자수, 비경제 활동인구, 석유금수, 최고속도제한)
[Loeb(1985)모형]	1979년의 통계자료에 근거해서 크로스 섹션분석, 선형모형을 이용	사망지수 = f(소득, 알콜, 교육수준, 자동차 검사, 자동차연료소비량, 인구밀도)
[Naoum and AL-Jawahiry(1989)]	이라크의 사고 자료	사망자수 = $-366 + 0.0052 \cdot \text{차량대수}$
[임헌연·이일병, 1990]	부산시의 교통사고 자료의 Logistic 모형	교통사고 발생 건수 = $1 + \text{EXP}(1.46 - 0.16t)$
[제승화부, 1979]	1948년부터 1970년까지의 자료중 1959-1961년 자료를 제외한 20년간의 자료로 LSE(Least Square Estimate Method)방법을 이용	$D_v = 300 \left(\frac{1000V}{P} \right)^{-0.6686}$ $D_F = 3.0 \left(\frac{1000V}{P} \right)^{0.3314}$ $D_T = 1550 \left(\frac{T_V}{P} \right)^{0.6889}$ <p>(D_V = 차량사고율(인/1만대) D_F = 인구사고율(인/10만인) D_T = 주행사고율(인/1억대 • km) V = 자동차등록대수(대) T_V = 통행량(대 • km/년))</p>

Rosman et al. (1996)의 연구에서는 1988년 서부 오스트레일리아의 병원의 3609가지 도로사고의 진단 코드를 바탕으로 각 신체부위의 상해 정도를 Abbreviated Injury Scaling (AIS)의 점수로 바꾸고 다시 minor와 major로 분류한 후 상해자의 나이, 성별, road user type (보행자, 탑승자, 오토바이, 자전거)의 판별함수로 logistic 회귀분석을 이용 상해정도를 판별하였다.

Shankar & Mannering(1996)은 단독 사고에서의 모터사이클 운전자 사고 정도 심각성을 5가지(property damage only, possible injury, evident injury, disabling injury, fatality)로 구분하여 이들의 변화를 환경 요소들, 도로 컨디션, 차조의 특성, 운전자 태도와 연관시키기 위해 multinomial-logit 모형을 이용하였다. 모형추정을 위해서 위싱턴 주의 5년간의 단독 오토바이 사고자료를 사용하였다.

▶ 대수선형분석(Log-Linear Analysis)

이 분석은 종속변수 또는 반응변수가 명의를척도(nominal scale)로 포아송분포를 따르는 범주형 자료로 contingency table로 표시된 변수들간의 관계를 특정 교호작용 항에 대한 모형의 검증을 실시하여 변수들 사이의 관계를 분석하는 방법이다. 교통 사고 통계 원표상의 일부 (일시, 연령 등)를 제외하고 대부분이 범주형 자료이므로 이경우의 분석에는 대수선형 분석의 응용분야가 넓다고 볼 수 있다.

김미영(1993)은 교통사고 통계 원표상의 여러 자료들 중 인적 요인으로 음주, 과속운전 여부, 보행자의 무단 횡단에 따른 사망정도가 통계적으로 유의한 차이가 있는지를 대수선형분석으로 비교하였다.

▶ Generalized Linear Model(GLM)

고전적인 선형모형이 포아송 분포화이나 이항 분포를 따르는 종속변수에 적합하지 않은데 반해, GLM은 명의 척도로 측정된 종속변수까지 포용하는 응용범위가 넓은 모형이다. 교통사고 수를 지역적으로 분석할 때 한 지역에서

발생한 사고수가 포아송 분포를 따른다는 가정하에 평균사고량의 변화를 교통사고가 발생한 지점의 여러 가지 환경요인으로 회귀 분석할 수 있다.

Maheer & Summersgill(1996)은 영국의 TRL(Transport Research Laboratory)에서 성공적으로 수행된 GLM 응용예를(교통사고를 교통흐름, 도로나 교차로의 기하학적인 구조와의 관계식으로 개발)소개하였다. 이 연구에서는 최소자승법에 기초한 전통적인 회귀분석보다 Poisson분포를 가정한 GLM이 보다 효과적임을 보였다. 저자들은 GLM을 교통사고율 모형에 적용시 발견될 수 있는 가능한 몇가지 문제점을 과잉산포, 불확실성에 대한 추정, 서로 다른 교통사고 형태의 상관관계, 그리고 개별적인 사고발생율의 집합 예측등으로 나누어 정리하고 각각의 문제점의 해결방안을 제시하였다.

▶ 생존분석(Survival Analysis)

생존분석은 관심 있는 분석단위의 수명분포를 규명하는데 그 목적을 둔다. 비교 분석 단위들의 수명분포간에 차이가 있을 때, 그들의 묘수를 분석단위와 관련된 주변 환경 변수의 함수로 가정하고 모형 추정을 하면, 관련 변수들의 수명 분포에 대한 영향력 및 수명 예측을 할 수 있어 교통사고 자료에 응용 분야가 넓다고 본다.

Jovanis & Cnang (1989)은 차량을 단위로 사고 나기까지 걸린 시간을 수명이라고 보고 생존분석을 이용 수명 분포의 묘수를 사고와 관련한 요인들(사고가 일어난 계절, 시간, 트럭운전자의 나이, 화물의 무게, 사고가 나기 전 휴식을 취한 시간)의 함수로 모형화하였다.

▶ 베이지안분석(Bayesian Analysis)

베이지안분석은 평균사고수에 대한 확률분포를 사고수가 실제로 주어진 상태에서 update 하는 분석으로 그 활용도가 넓다.

Mountain, Fawaz & Jarrett (1996)은 5,000개의 Minor Junction을 포함한 약 3,800km에 달하는 영국 고속도로의 교통사고자료를 분석하며 베이지안분

석을 하였다. 교통사고 수는 Poiss 분포를 따르며, 평균사고수를 사고환경 변수의 함수로 가정하였다. 사고환경 변수의 값이 동일한 조건으로 주어진다고 해도 평균 사고수가 반드시 일치하는 것은 아니므로, 평균사고수의 확률오차에 대한 분포 가정을 필요로 했다. 이때 확률오차 분포의 모수는 알려져 있다고 가정하지 않음으로, 모수 추정을 위해 Empirical Bayes절차를 밟았다. 이러한 Poisson 모형은 6개의 고속도로 범주에 대한 평균사고의 회귀추정을 위해 사용되었으며, Empirical Bayes절차는 교통사고건수의 추정을 향상하기 위해서 사용되었다. 고속도로상하에서의 사고는 노출(교통흐름)과 Minor Junction의 빈도에 비선형함수로 나타났다.

▶ 수량화이론(Qualification Theory)

수량화이론(Qualification Theory)은 Guttman의 예측이론으로부터 출발하였으며, 명의척도(Nominal Scale)나 순위척도(Ordinal Scale)로 조사된 자료에 더미변수와 같은 기계적인 수치를 부여하기보다는, 최적의 데이터 해석이 가능하도록 수치를 부여하여 분석하는 것이다. 그 수치는 질적 자료를 회귀분석에 적용해서 추정된 각 더미변수의 회귀계수와 유사하다.

강동수(1994)는 수량화 이론을 이용하여 7개 사고 유형별 사고율에 영향을 미치는 요인들에 대해 분석했다. 분석에 사용된 종속변수는 1백만 진입차량당 사고건수이고 25개의 교통관련 변수들을 설명변수로 이용하였다.

수량화 이론의 여타활용도를 살펴보면, 김희생(1991)의 부산시 도심지역의 주차행동결정 수량화 모형에 관한 연구와 김종재(1992)의 지방도시지가형성 요인 분석, 이한우(1993)의 주말 여가 통행형태 및 요인분석등 최근 사회 각 분야에 다양하게 적용되고 있다.

4. 향후 연구를 위한 제안

이상 정리된 바와 같이 집계된 교통통계자료를 바탕으로 한 분석의 범위는 매우 넓다고 보겠다. 본 장에서는 몇가지 더 연구되어야 될 분야에 대한 제언을 하면 다음과 같다.

첫째, 대부분 범주형 자료로 집계되는 사고 통계의 특징을 감안 할 때, 범주형 자료를 위한 다변량분석의 응용이 기대되는 방이다. 질적 자료에 대한 주성분분석, 정준상관분석, Correspondence Analysis 등은 SAS와 같은 상용 통계 package를 이용하여 용이하게 시행할 수 있다.

둘째로, 더 연구되어야 할 분야는 메타분석을 교통사고량에 수많은 회귀분석 결과에 이용될 수 있다. <표 2>에 나타난바와 같이 다양한 교통사고에 대한 연구결과가 나왔을 때 그 결과들을 또 하나의 자료로 동일한 모수임에도 결과가 상이하다면 다른 이유가 될 만한 인자들의 함수로 모수를 재추정함으로써 그 원인을 설명할 수 있어 매우 유익한 분석이라 할 수 있다.

셋째로, 우리 나라 교통사고다발지점 선정 및 분석을 위해 통계 원표의 활용방안이다. 우리 나라는 시가지도에서는 반경 30m내에서 연간 인적피해사고가 5-10건 이상인 지점(특별시 10건, 광역시 7건, 일반시급 5건)을 사고다발지점으로 선정하며 시가지외 도로에서는 100내에서 연간 인건피해사고 3건 이상인 지점을 사고다발지점이라고 지칭한다(경찰청, 1996). 1995년 기준으로 우리 나라에는 6,103개의 교통사고 다발지점이 있는데 매년 개선을 하고 있는 실정이다.

다발지점선정은 통계원표의 정보가 이용되는 것이 아니고 교통사고 조사보고서를 토대로 발생장소의 동일지점별 사망자수, 사고유형 등을 정밀분석 후 선정하고 있어 막대한 노력을 들여 작성되며 전산처리가 가능한 통계원표를 바탕으로 다발지점을 선정하고 원인규명을 할 수 있는 안이 더 연구되어야겠다. 하나의 대안으로 이러한 작업을 위해서 군집분석과 포아송회귀분석을 적용할 수있다.

통계원표의 사고발생위치 정보중 하나인 (중.횡측좌표)를 바탕으로 우선 군집분석을 하여 각군집별로 발생한 연간 총 사고량을 구한다. 사고다발자점

에 대한 정의에 따라 다발지점을 선정한 후 해당 지역의 교통관련 인자들의 함수로 사고 수를 포아송 회귀분석하여 통계적으로 유의한 관련인자를 찾아낸다.

이 밖에도 교통사고통계정보를 활용하여 교통안전대책에 도움을 줄 수 있는 통계모형 활용방안들이 많아 이 분야에 활기찬 연구가 기대되는 바이다.

< 참 고 문 헌 >

- (1) 강동주(1994), “교통사고 유형별 사고발생특성 및 요인분석”, 한양대 석사논문.
- (2) 김규형(1993), 『다변량 분석의 이론과 실재』 중앙대학교 출판부, p9-11.
- (3) 김미영((1993), “교통사고 자료분석 기법 비교검토 및 우리나라 사례 분석”, 서울대 석사논문.
- (4) 김용석(1993), “Fuzzy 이론을 이용한 교통사고 많은 지점 사고원인 분석”, 한양대 석사논문.
- (5) 김홍상(1987), “교통사고의 기술방법에 관한 연구.” 서독 칼스루에 공과 대학박사학위 논문.
- (6) 박병호(1995), “충청권의 교통사고 예측모형 개발에 관한 연구”, 『대한 교통학회』 제13권 제1호, pp.68~82.
- (7) 박재언(1995), “교통사고 예측모형 개발과 적용연구”, 전북대 석사논문.
- (8) 오윤성, 고양선(1992), “대형교통사고 영향요인의 관별모델 구축에 관한 연구”, 『대한 교통학회』
- (9) 윤문호(1987), “교통사고 발생특성과 그에 따른 영향요소 분석에 관한 분석에 관한 연구”, 한양대 석사논문.
- (10) 이광희(1994), “교통사고 다발지점의 유형화와 그 대책에 관한 연구”, 부산대 석사논문.
- (11) 이윤중(1991), “시계열 모형을 이용한 자동차 사고 유형과 원인 분석”, 중앙대 석사논문.
- (12) 이일병, 임현연(1990), “한국의 교통사고 예측 모형 개발에 관한 연구”. 『대한 교통 학회지』 제 8권 1호. p73-88.
- (13) 임윤택(1992), “도로특성이 교통사고에 미치는 영향분석”, 한양대 석사.
- (14) 신용균(1992), “사망사고 요인분석”, 『대한교통 안전연구논집』, 제 11권.

- (15) 채남희(1996), 『교통사고사망자수 예측모델 구축』, 교통개발연구원.
- (16) Arizona D.O.T(1995), "Mother Vehicle Crash Facts for 1995".
- (17) Arizona D.O.T(1993), "Arizona Traffic Accident Report
<Instruction Manual & Glossary>".
- (18) Rosman, D. L., Knuiman, M. W. Ryan, G. (1996), "An Evaluation
of Road Crash Injury Severity Measuring", *Accid.Anal. and Prev.*
Vol.28, No.2, p.163-170.
- (19) Kanellaidis, G. (1995), "Factors Affecting Drivers' Choice of Speed
on Roadway Curves", *Journal of Safety Research*, Vol. 26, No.1 .
p.49-56.
- (20) Jovanis, P. and Chang, H.(1989), "Disaggregate Model of Highway
Accident Occurrence using Survival Theory", *Accid, Anal. and
Prev.* Vol.21, No.5, pp.45-458.
- (21) Mountain, L., Fawaz, B. and Jarrett, D. (1996), "Accident Pred-iction
Models for Road with Minor Junctions", *Accid.Anal. and Prev*, Vol.28,
No.6, pp.695-707.
- (22) Maher, M. J. and Summersgil, I.(1996), "A Comprehensive
Methodology for the Fitting of Predictive Accident Models", *Accid
Anal. and Prev*, Vol.28, No.3, pp.281-291.
- (23) Navin, F.(1993), "Fundamental Relationship for Road Read Safety",
Lectures to the First Highway Design Company, Xian, China.
- (24) Naoum, F. and Al-Jawahiry, J(1984), "The Factors Affecting
Growth of Accident Rates in Developing Countries : Iraq as a
Case Study", *Proceedings of 11th IRF World Meeting*, Seoul,
Korea.
- (25) Partyka. S. C. (1991), "Simple Models of Fatality Trends Using
Employment and Population Data" *Accid. Anal & Prev.*, Vol. 23,
No.5.
- (26) Oppe, S.(1992), "A Comparison of Some Statistical Technique for

- Road Accident Analysis", *Accid. Anal. & Prev.* Vol.24. No.4. p397-423.
- (27) Smeed, R. J.(1972), "The Usefulness of Formula in Traffic Engineering and Road Safety", *Accid Anal. & Prev.* Vol.4. p303-312.
- (28) Stinson, L. Fisher, L.& Kay, S. (1996), "Overview of Data Editing Procedures in Surveys Administered by The Bureau of Labor statistics : Procedure and Implication', *1st International CASIC Conference*, San Antonio.
- (29) Trinca, G.W & Johnston, I.R et al.(1990), "Reducing Traffic Injury : A Global Challenge," An International Traffic Safety Project of The Royal Australian College of Surgeons.
- (30) Shanker, V. and Mannering, F.(1996), "An Exploratory Multinomial Logit Analysis of Single- Vehicle Motorcycle Accident Severity". *Journal of Safety Research*, Vol. 27. No.3. p.183-194.
- (31) Zlatoper, T. J.(1989), " Models Explaining Motor Vehicle Death Rates in The United States." *Accid. Anal. & Prev*, Vol. 21. No.2.
- (32) 齊藤和夫(1979), わか國たおける 交通事故 死亡危険度の推移に 關するアクロ分析的 研究, 交通工學, Vol. 14, No.6, pp.3~12.

An Overview of Statistical Analysis of

Road Traffic Accident

So-Young Sohn

<abstract>

Recent increase in the road traffic accident in Korea draws special attention. In order to establish the safety policy to curve the increasing road traffic accident rate, causes of accident have to be identified. For this purpose, it is necessary to employ proper statistical models based on accurately collected traffic accident reports.

In this paper, we review the statistical literature which deals with road traffic accidents analysis and make some suggestions to improve the current practice of accident analysis.